Dissertation

Self-organizing Context Spaces to Support Information Management and Knowledge Work

Thesis approved by the
Department of Computer Science of the
University of Kaiserslautern-Landau (RPTU)
for the award of the Doctoral Degree

Doctor of Engineering (Dr.-Ing.)

to

Christian Jilek

Date of Viva:	March 21st, 2024
Dean:	Prof. Dr. Christoph Garth
Head of PhD Committee:	Prof. Dr. Stefan Deßloch
Reviewers:	Prof. Dr. Prof. h.c. Andreas Dengel
	Prof. Dr. Ingo J. Timm (Trier University)

DE-386

Für meine Eltern

Acknowledgements

Supervisors & PhD Committee

First, I would like to express my sincere gratitude to Prof. Andreas Dengel for the opportunity to conduct my research. His guidance actually started in the last semesters of my diploma studies and continued during the work on my diploma thesis and the time as a researcher and PhD student in his research department (SDS) at DFKI. Andreas gave me the freedom to investigate various ideas and always had an "open door" to approach him for feedback and discussions regardless of how full his calendar was. I remember these meetings being highly motivating to continue my research as well as exploring the one or the other additional facet.

Further, I would like to thank Prof. Ingo Timm for becoming my second supervisor. Ultimately thanks to the German Research Foundation (DFG), we met each other in their priority program (SPP) on *Intentional Forgetting in Organizations*. Before becoming my supervisor, we thus had already had several workshops and a joint publication together. As with Andreas, I drew great insights and motivation from our discussions and brainstormings. And although suffering a heavy stroke of fate towards the very end of my PhD project, Ingo did not withdraw from his commitment as a supervisor and reviewer, for which I cannot thank him enough.

Finally, I would like to thank Prof. Stefan Deßloch for becoming the head of my PhD committee and his support in getting the final version of my thesis ready for publication. I also thank Bernd Schürmann, Sabine Owens (Dean's Office) and Brigitte Selzer (DFKI SDS) for their help in organizing and preparing my PhD defense.

External Research Partners

I already mentioned the DFG priority program (SPP) on *Intentional Forgetting in Organizations*, which also included our *Managed Forgetting* project. Together with the preceding EU-funded project *ForgetIT*, this allowed me to participate in about a decade of interdisciplinary research. In addition to my fellow researchers in the SPP, I would especially like to thank Prof. Tobias Tempel, Prof. Christian Frings, Yannick Runge, Paula Gauselmann, Christoph Geissler, Claudia Niederée, Andrea Ceroni, Huyen Nguyen, Anna-Sophie Ulfert-Blank, Mark Greenwood, Francesco Gallo, Prof. Robert Logie, Maria Wolters and Stephen Rhodes for our joint work in these projects. Especially the cognitive science perspective broadened my horizons and taught me so much with regard to being a better researcher in general.

Team "Knowledge Work" and the SDS Research Department at DFKI

In the last semester of my diploma studies, Andreas introduced me to Heiko Maus, head of the "Knowledge Work" team in the SDS research department at DFKI. Heiko not only became the (co-)supervisor of my diploma thesis but also the mentor of my PhD project. He taught me most of the things I know about the everyday life as a researcher. One could say he is the perfect embodiment of DFKI's mission statement of application-oriented foundational research focusing not only on knowledge gain but also on transferring said knowledge to industry.

Throughout my whole PhD project, I thus contributed to research as well as industry projects making my work being grounded on and applied to real-world problems. In addition to Heiko, I would also like to thank Sven Schwarz for being an advisor for my work. Sven serves as a coordinator of the knowledge transfer (i.e. industry projects) in Heiko's team thus having a high focus on long-term usability and bringing maturity and longevity to research components trying to make them building blocks for steady progress instead of throwaway demonstrators. I am grateful for his attitude, which helped making several of my components being used 24/7 in industry for more than half a decade by the time of publishing this thesis. Since Sven and I were forming a car pool, we had plenty of time to discuss various topics at the beginning and end of each work day – many interesting ideas sparked from these conversations.

While Heiko and Sven were mentors for me, I also had several peers that helped me with my PhD project. First, I would like to thank Markus Schröder, my long-term roommate at DFKI. We were more or less "always" talking about new research or software ideas, being at work or in our free time on hobby projects. Whenever I came up with an idea, Markus was there to provide an instant peer review, feedback for improvement, etc. Unsurprisingly, he was a co-author of several of my papers and vice versa leading to more than 20 joint publications during our time as PhD students – four of them award-winning and another award runner-up.

Next, I would like to thank Sven Hertling, our roommate for two very successful years at DFKI (two of the mentioned awards) and an external research peer at the University of Mannheim afterwards. Sven provided a bit more of an external view on my work – similarly to colleagues like Prof. Jörn Hees, Joachim Folz, Sebastian Palacio, Christian Reuschling, Ludger van Elst, Ansgar Bernardi, Thomas Kieninger, Stephan Baumann, Martin Memmel, Nicolas Großmann or Saleh Mozaffari, who did not directly work in one of my projects but occasionally had helpful advice for me and/or interesting discussions with me – thank you all for that.

I would also like to thank my long-term research assistants Rudolf Koch (née Novik), Jessica Chwalek and Desiree Heim, who themselves have become successful engineers and researchers, respectively. Speaking of engineers: thanks to Andreas Lauer, Michael Kraus and Emil Baitemirov for their software engineering and coding wisdom whenever I approached them with (sometimes crazy) ideas on how to implement things.

Finally, I would like to thank the *SensAI* and *SensAI* follow-up research teams, especially Prof. Adrian Ulges, Federico Raue, Michael Schulze and Mahta Bakhshizadeh for many interesting discussions and collaborations.

I am glad to call all of you my colleagues
and some of you even my friends (or even close friends).

Family & Friends

Last but not least, I would like to thank my family and friends, who a bit too often had to do without me while I was trying to meet another deadline, was away on a research or business trip, did not allow myself to lose momentum on a new discovery or idea, etc.

Here, I especially would like to thank my parents, Renate and Fridolin, to whom I dedicate this work. Thank you for your continuous love and support and for believing in me, especially in bad times when I myself had doubts.

– Thank you very much, Christian.

Grants

This work was supported by the German Research Foundation (DFG) in their priority program (SPP) on *Intentional Forgetting in Organizations* and in particular the *Managed Forgetting* project (Grants DE 420/19-1 and DE 420/19-2), by the Federal Ministry of Education and Research (BMBF) in the projects *supSpaces* (Grant 01IS15013B) and *SensAI* (Grant 01IW20007) and by the European Commission in the project *ForgetIT* (Grant 600826).

Misc

Layout and style of this thesis have been inspired by the *ClassicThesis* package by André Miede and Ivo Pletikosić. The cSpaces logo uses graphical elements provided by Sven Schwarz and a font style inspired by the *Urban Tribe Font* by Jonathan Swinn.

Abstract

Knowledge workers face an ever increasing flood of information in their daily work. They live in a "multi-tasking craziness", involving activities like creating, finding, processing, assessing or organizing information while constantly switching from one context to another, each being associated with different tasks, documents, mails, etc. Hence, their personal information sphere consisting of file, mail and bookmark folders as well as their content, calendar entries, etc. is cluttered with information that has become irrelevant. Finding important information thus gets harder and much of previously gained knowledge is practically lost.

This thesis explores new ways of solving this problem by investigating the potential of self-(re)organizing and especially forgetting-enabled personal knowledge assistants in the given scenario. It utilizes so-called Managed Forgetting, which is an escalating set of measures to overcome the binary keep-or-delete paradigm, ranging from temporal hiding, to condensation, to adaptive reorganization, synchronization, archiving and deletion. Managed Forgetting is combined with two other major ideas: First, it uses the Semantic Desktop as an ecosystem, which brings Semantic Web and thus knowledge graph technologies to a user's desktop, making it possible to capture and represent major parts of a user's personal mental model in a machine-understandable way and exploit it in many different applications. Second, the system uses explicated context information – so-called Context Spaces: context is seen as an explicit interaction element users can work *with* (i.e. a "tangible" object similar to a folder) and *in* (immersion). The thesis is structured according to the basic interaction cycle with such a system, ranging from evidence collection to information extraction and context elicitation, followed by information value assessment and the actual support measures consisting of self-(re)organization decisions (back-end) and user interface updates (front-end). The system's data foundation are personal or group knowledge graphs as well as native data. This work makes contributions to all of these aspects, whereas several of them have been investigated and developed in interdisciplinary research with cognitive scientists. On a more general level, searching and trust in such highly autonomous assistants have also been investigated.

In summary, a self-(re)organizing and especially forgetting-enabled support system for information management and knowledge work has been realized. Its different features vary in maturity: the most mature ones are already in practical use (also in industry), while the latest are just well elaborated (position papers) or rough ideas. Different evaluation strategies have been applied ranging from mere data-driven experiments to various user studies. Some of them were rather short-term with controlled laboratory conditions, others less controlled but spanning several months. Different benefits of working with such a system could be quantified, e.g. cognitive offloading effects and reduced task switching/resumption time. Other benefits were gathered qualitatively, e.g. tidiness of the information sphere and its better alignment with the user's mental model. The presented approach has been shown to hold a lot of potential. In some aspects, however, only first steps have been taken towards tapping it, e.g. several support measures can be further refined and automation further increased.

Contents

List of Publications

Publications as Part of this Thesis

Parts of the research and material (including figures, tables and algorithms) in this thesis have already been published as follows. All listed publications are authored or substantially co-authored (e.g. contribution of own sections, ideas or code) by the author of this thesis.

Journal Papers

[1] Christoph Geissler*, Paula Gauselmann*, **Christian Jilek**, Heiko Maus, Christian Frings and Tobias Tempel. “A Functional Near-infrared Spectroscopy Study on the Prefrontal Correlates of Cognitive Offloading via a Personal Knowledge Assistant.” In: *Scientific Reports* 13.13938 (2023). DOI: 10.1038/s41598-023-39540-5.
*) shared first authorship.

Note: The experiment reported in this paper, an fNIRS-based user study (see last paragraph of Section 8.3.2), was still in progress at the time of this thesis' submission. It could be completed (including publication of the results) during its review process (after the submission but before the doctoral viva to be precise).

[2] Paula Gauselmann*, Yannick Runge*, **Christian Jilek**, Christian Frings, Heiko Maus and Tobias Tempel. “A Relief from Mental Overload in a Digitalized World: How Context-sensitive User Interfaces Can Enhance Cognitive Performance.” In: *International Journal of Human-Computer Interaction* 39.1 (2023), pp. 140–150. DOI: 10.1080/10447318.2022.2041882.
*) shared first authorship.

[3] Tobias Tempel, Claudia Niederée, **Christian Jilek**, Andrea Ceroni, Heiko Maus, Yannick Runge and Christian Frings. “Temporarily Unavailable: Memory Inhibition in Cognitive and Computer Science.” In: *Interacting with Computers* 31.3 (May 2019), pp. 231–249. DOI: 10.1093/iwc/iwz013. arXiv: 1912.00760. URL: https://arxiv.org/pdf/1912.00760.pdf.

[4] **Christian Jilek**, Jessica Chwalek, Sven Schwarz, Markus Schröder, Heiko Maus and Andreas Dengel. “Advanced Memory Buoyancy for Forgetful Information Systems.” In: *AIS Transactions on Enterprise Systems* 4.1 (May 2019). Special Issue accompanied by Conference on Intentional Forgetting in Organizations and Work Settings (CoInFo 2018), Potsdam, Germany, September 11-12, 2018. DOI: 10.30844/aistes.v4i1.11. arXiv: 1811.12177. URL: https://www.aes-journal.com/index.php/ais-tes/article/view/11/11.

[5] **Christian Jilek**, Yannick Runge, Claudia Niederée, Heiko Maus, Tobias Tempel, Andreas Dengel and Christian Frings. "Managed Forgetting to Support Information Management and Knowledge Work." In: *KI – Künstliche Intelligenz, German Journal of Artificial Intelligence - Organ des Fachbereichs "Künstliche Intelligenz" der Gesellschaft für Informatik e.V.* 33.1 (Mar. 2019), pp. 45–55. DOI: 10.1007/s13218-018-00568-9. arXiv: 1811.12155. URL: https://arxiv.org/pdf/1811.12155.pdf.

Book Chapters

[1] Heiko Maus, **Christian Jilek** and Sven Schwarz. "Remembering and Forgetting for Personal Preservation." In: *Personal Multimedia Preservation: Remembering or Forgetting Images and Video*. Ed. by Vasileios Mezaris, Claudia Niederée and Robert H. Logie. Springer, 2018, pp. 233–277. DOI: 10.1007/978-3-319-73465-1_7.

Conference Papers

[1] Desiree Heim, **Christian Jilek**, Heiko Maus and Andreas Dengel. "A Retrospective Context Mining Approach for Bootstrapping Personal Knowledge Assistants." In: *Proceedings of the "Lernen, Wissen, Daten, Analysen – Learning, Knowledge, Data, Analytics" 2022 Workshops: FGWM, FGKD and FGDB (LWDA 2022), Hildesheim, Germany, October 5-7, 2022*. Ed. by Pascal Reuss, Viktor Eisenstadt, Jakob Schönborn and Jero Schäfer. Vol. 3341. CEUR Workshop Proceedings. CEUR-WS.org, 2022. URL: https://ceur-ws.org/Vol-3341/WM-LWDA_2022_CRC_8910.pdf.

[2] Mahta Bakhshizadeh, **Christian Jilek**, Heiko Maus and Andreas Dengel. "Leveraging Context-aware Recommender Systems for Improving Personal Knowledge Assistants by Introducing Contextual States." In: *Proceedings of the "Lernen, Wissen, Daten, Analysen – Learning, Knowledge, Data, Analytics" 2021 Workshops (LWDA 2021): FGWM, KDML, FGWI-BIA, and FGIR, Online, September 1-3, 2021*. Ed. by Thomas Seidl, Michael Fromm and Sandra Obermeier. Vol. 2993. CEUR Workshop Proceedings. CEUR-WS.org, 2021. URL: http://ceur-ws.org/Vol-2993/paper-01.pdf.

[3] **Christian Jilek**, Paula Gauselmann, Jessica Chwalek, Tobias Tempel and Andreas Dengel. "Quantifying Benefits of a Personal Knowledge Assistant on Task Resumption." In: *Arbeit HumAIne Gestalten – 67. Frühjahrskongress der Gesellschaft für Arbeitswissenschaft e.V. [HumAIne Work Design – 67th Spring Congress of the Human Factors and Ergonomics Society] (GfA 2021), Ruhr-Universität Bochum, Online, March 3-5, 2021*. GfA-Press, 2021. ISBN: 978-3-936804-29-4.

[4] Uwe Riss, Heiko Maus, Sabrina Javaid and **Christian Jilek**. "Digital Twins of an Organization for Enterprise Modeling." In: *The Practice of Enterprise Modeling – 13th IFIP WG 8.1 Working Conference on the Practice of Enterprise Modelling (PoEM 2020), Riga, Latvia, November 25-27, 2020, Proceedings*. Ed. by Jānis Grabis and Dominik Bork. Lecture Notes in Business Information Processing (LNBIP). Springer, 2020, pp. 25–40. DOI: 10.1007/978-3-030-63479-7_3. URL: https://www.dfki.uni-kl.de/~maus/dok/RissMausJavaid+2020.pdf.

[5] **Christian Jilek**, Markus Schröder, Rudolf Novik, Sven Schwarz, Heiko Maus and Andreas Dengel. "Inflection-Tolerant Ontology-Based Named Entity Recognition for Real-Time Applications." In: *2nd Conference on Language, Data and Knowledge (LDK 2019), Leipzig, Germany, May 20-23, 2019*. Ed. by Maria Eskevich, Gerard de Melo, Christian Fäth, John P. McCrae, Paul Buitelaar, Christian Chiarcos, Bettina Klimek and Milan Dojchinovski. Vol. 70. OpenAccess Series in Informatics (OASIcs). Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2019, 11:1–11:14. DOI: 10.4230/OASIcs.LDK.2019.11. arXiv: 1812.02119. URL: https://drops.dagstuhl.de/opus/volltexte/2019/10375/pdf/OASIcs-LDK-2019-11.pdf.
Received LDK 2019 Best Research Paper Award.

[6] Ingo J. Timm, Steffen Staab, Michael Siebers, Claudia Schon, Ute Schmid, Kai Sauerwald, Lukas Reuter, Marco Ragni, Claudia Niederée, Heiko Maus, Gabriele Kern-Isberner, **Christian Jilek**, Paulina Friemann, Thomas Eiter, Andreas Dengel, Hannah Dames, Tanja Bock, Jan Ole Berndt and Christoph Beierle. "Intentional Forgetting in Artificial Intelligence Systems: Perspectives and Challenges." In: *KI 2018: Advances in Artificial Intelligence – 41st German Conference on AI, Berlin, Germany, September 24–28, 2018, Proceedings*. Springer, 2018, pp. 357–365. DOI: 10.1007/978-3-030-00111-7_30. URL: https://fis.uni-bamberg.de/bitstream/uniba/46646/1/fisba46646.pdf.

[7] **Christian Jilek**, Markus Schröder, Sven Schwarz, Heiko Maus and Andreas Dengel. "Context Spaces as the Cornerstone of a Near-Transparent and Self-Reorganizing Semantic Desktop." In: *The Semantic Web: ESWC 2018 Satellite Events – ESWC 2018 Satellite Events, Heraklion, Crete, Greece, June 3-7, 2018, Revised Selected Papers*. Ed. by Aldo Gangemi, Anna Lisa Gentile, Andrea Giovanni Nuzzolese, Sebastian Rudolph, Maria Maleshkova, Heiko Paulheim, Jeff Z. Pan and Mehwish Alam. Springer, 2018, pp. 89–94. DOI: 10.1007/978-3-319-98192-5_17. arXiv: 1805.02181. URL: https://arxiv.org/pdf/1805.02181.pdf.

[8] **Christian Jilek**, Sven Schwarz, Heiko Maus and Andreas Dengel. "Managed Forgetting, Data Condensation & Preservation in Application." In: *Third Workshop on Ubiquitous Technologies for Augmenting the Human Mind (WAHM 2016), Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct, Heidelberg, Germany, September 12-16, 2016*. UbiComp 2016. ACM, 2016, pp. 1046–1053. DOI: 10.1145/2968219.2968567. URL: https://www.dfki.uni-kl.de/~jilek/files/2016/JilekSchwarzMaus+16_Managed_Forgetting_WAHM16.pdf.

[9] **Christian Jilek**, Heiko Maus, Sven Schwarz and Andreas Dengel. "Diary Generation from Personal Information Models to Support Contextual Remembering and Reminiscence." In: *Workshop on Human Memory-Inspired Multimedia Organization and Preservation (HMMP 2015), Proceedings of the 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW 2015), Turin, Italy, June 29 - July 3, 2015*. IEEE, 2015, pp. 1–6. DOI: 10.1109/ICMEW.2015.7169753. URL: https://pimo.opendfki.de/wp9-pilot/diary/JilekMausSchwarz+15_PIMO_Diary_HMMP15.pdf.

Technical Reports

[1] Markus Schröder, **Christian Jilek** and Andreas Dengel. "Interactive Concept Mining on Personal Data – Bootstrapping Semantic Services." In: *Computing Research Repository (CoRR)* (2019). arXiv: 1903.05872. URL: https://arxiv.org/pdf/1903.05872.pdf.

[2] Andreas Dengel, Heiko Maus, Sven Schwarz, **Christian Jilek**, Markus Schröder, Michael Schulze, Andreas Lauer, Rudolf Koch and Emil Baitemirov. *CoMem: a group-wide corporate memory ecosystem*. Web documentation. Smart Data & Knowledge Services (SDS) Research Department, DFKI, 2018. URL: https://comem.ai/.

Public Project Deliverables

[1] Xiaofei Zhu, Claudia Niederée, Tuan Tran, Andrea Ceroni, Kaweh Djafari Naini, Nam Khanh Tran, Heiko Maus and **Christian Jilek**. *ForgetIT Deliverable D3.4: Strategies and Components for Managed Forgetting*. Deliverable. ForgetIT Consortium, Mar. 2016. URL: https://www.forgetit-project.eu/fileadmin/fm-dam/deliverables/ForgetIT_WP3_D3.4.pdf.

[2] Heiko Maus, Sven Schwarz, **Christian Jilek**, Maria Wolters, Stephen Rhodes, Andrea Ceroni and Gürkan Gür. *ForgetIT Deliverable D9.5: Personal Preservation Report*. Deliverable. ForgetIT Consortium, Jan. 2016. URL: https://www.forgetit-project.eu/fileadmin/fm-dam/deliverables/ForgetIT_WP9_D9.5New.pdf.

[3] Heiko Maus, Sven Schwarz, **Christian Jilek** and Francesco Gallo. *ForgetIT Deliverable D9.4: Personal Preservation Pilot II: Concise Preserving Mobile Information Assistant*. Deliverable. ForgetIT Consortium, Dec. 2015. URL: https://www.dfki.de/fileadmin/user_upload/import/8360_D9.4.pdf.

[4] Heiko Maus, Sven Schwarz, Bahaa Eldesouky, **Christian Jilek**, Maria Wolters and Berker Loğoğlu. *ForgetIT Deliverable D9.3: Use Cases & Mock-up Development*. Deliverable. ForgetIT Consortium, Feb. 2015. URL: https://www.forgetit-project.eu/fileadmin/fm-dam/deliverables/ForgetIT_WP9_D9.3.pdf.

[5] Heiko Maus, Sven Schwarz, **Christian Jilek** and Bahaa Eldesouky. *ForgetIT Personal Preservation Pilot*. Web documentation. 2015. URL: https://pimo.opendfki.de/wp9-pilot/.

In Preparation

[1] Anna-Sophie Ulfert-Blank, Julia Knabe and **Christian Jilek**. "Forgetting-enabled AI systems: Exploring the role of transparency perceptions and trust." Short paper accepted at 24th International Conference on Human-Computer Interaction (HCII 2022) but withdrawn due to unforeseen scheduling conflicts. Extended journal manuscript in preparation. 2022.

[2] Kai Sauerwald, Lukas Reuter, **Christian Jilek**, Christoph Beierle, Jan Ole Berndt, Hannah Dames, Andreas Dengel, Diana Howey, Gabriele Kern-Isberner, Teresa Kraemer, Heiko Maus, Durgesh Nandini, Claudia Niederée, Thi Huyen Nguyen, Nico Potyka, Marco Ragni, Ute Schmid, Steffen Staab and Ingo J. Timm. "How to Intentionally Forget in AI Systems: A Multidisciplinary Approach to Belief Dynamics." Journal manuscript in preparation (preliminary title and author list). 2022.

Other Publications

The following publications are co-authored by the author of this thesis. They deepen some of this thesis' side issues or related topics but are not considered an explicit part of it.

Conference Papers

[1] Michael Schulze, Michelle Pelzer, Markus Schröder, **Christian Jilek**, Heiko Maus and Andreas Dengel. "Towards Knowledge Graph Based Services in Accounting Use Cases." In: *Proceedings of Poster and Demo Track and Workshop Track of the 18th International Conference on Semantic Systems co-located with 18th International Conference on Semantic Systems (SEMANTiCS 2022)*. Ed. by Umutcan Şimşek, David Chaves-Fraga, Tassilo Pellegrini and Sahar Vahdat. Vol. 3235. CEUR Workshop Proceedings. CEUR-WS.org, 2022. URL: https://ceur-ws.org/Vol-3235/paper19.pdf.

[2] Markus Schröder, **Christian Jilek** and Andreas Dengel. "A Human-in-the-Loop Approach for Personal Knowledge Graph Construction from File Names." In: *Proceedings of the 3rd International Workshop on Knowledge Graph Construction (KGCW 2022) at 19th Extended Semantic Web Conference (ESWC 2022), Hersonissos, Crete, Greece, May 29 - June 2, 2022*. Ed. by David Chaves-Fraga, Anastasia Dimou, Pieter Heyvaert, Freddy Priyatna and Juan Sequeda. Vol. 3141. CEUR Workshop Proceedings. CEUR-WS.org, 2022. URL: http://ceur-ws.org/Vol-3141/paper2.pdf.

[3] Markus Schröder, **Christian Jilek** and Andreas Dengel. "Spread2RML: Constructing Knowledge Graphs by Predicting RML Mappings on Messy Spreadsheets." In: *Proceedings of the 11th International Conference on Knowledge Capture (K-CAP 2021), Virtual Event, USA, December 2-3, 2021*. ACM, 2021, pp. 145–152. DOI: 10.1145/3460210.3493544. arXiv: 2110.12829. URL: https://arxiv.org/pdf/2110.12829.pdf.
Received K-CAP 2021 Best Paper Award.

[4] Michael Schulze, Markus Schröder, **Christian Jilek** and Andreas Dengel. "ptpDG: A Purchase-To-Pay Dataset Generator for Evaluating Knowledge-Graph-Based Services." In: *Proceedings of the ISWC 2021 Posters & Demonstrations and Industry Tracks: From Novel Ideas to Industrial Practice co-located with 20th International Semantic Web Conference (ISWC 2021), Virtual Conference, October 24-28, 2021*. Ed. by Oshani Seneviratne, Catia Pesquita, Juan Sequeda and Lorena Etcheverry. Vol. 2980. CEUR Workshop Proceedings. CEUR-WS.org, 2021. URL: http://ceur-ws.org/Vol-2980/paper386.pdf.

[5] Michael Schulze, Markus Schröder, **Christian Jilek**, Torsten Albers, Heiko Maus and Andreas Dengel. "P2P-O: A Purchase-To-Pay Ontology for Enabling Semantic Invoices." In: *The Semantic Web – 18th International Conference, ESWC 2021, Virtual Event, June 6-10, 2021, Proceedings*. Ed. by Ruben Verborgh, Katja Hose, Heiko Paulheim, Pierre-Antoine Champin, Maria Maleshkova, Oscar Corcho, Petar Ristoski and Mehwish Alam. Springer, 2021, pp. 647–663. DOI: 10.1007/978-3-030-77385-4_39. URL: https://openreview.net/pdf?id=50N86aYZ3n.

[6] Markus Schröder, **Christian Jilek** and Andreas Dengel. "Dataset Generation Patterns for Evaluating Knowledge Graph Construction." In: *The Semantic Web: ESWC 2021 Satellite Events – ESWC 2021 Satellite Events, Virtual Event, June 6-10, 2021, Revised Selected Papers*. Ed. by Ruben Verborgh, Anastasia Dimou, Aidan Hogan, Claudia d'Amato, Ilaria Tiddi, Arne Bröring, Simon Mayer, Femke Ongenae, Riccardo Tommasini and Mehwish Alam. Springer, 2021, pp. 27–32. DOI: 10.1007/978-3-030-80418-3_5. arXiv: 2104.13576. URL: https://arxiv.org/pdf/2104.13576.pdf.

[7] Markus Schröder, **Christian Jilek** and Andreas Dengel. "Mapping Spreadsheets to RDF: Supporting Excel in RML." In: *Proceedings of the 2nd International Workshop on Knowledge Graph Construction (KGCW 2021) at 18th Extended Semantic Web Conference (ESWC 2021), Virtual Event, June 6-10, 2021*. Ed. by David Chaves-Fraga, Anastasia Dimou, Pieter Heyvaert, Freddy Priyatna and Juan Sequeda. Vol. 2873. CEUR Workshop Proceedings. CEUR-WS.org, 2021. arXiv: 2104.13600. URL: http://ceur-ws.org/Vol-2873/paper3.pdf.

[8] Markus Schröder, **Christian Jilek**, Michael Schulze and Andreas Dengel. "The Person Index Challenge: Extraction of Persons from Messy, Short Texts." In: *Proceedings of the 13th International Conference on Agents and Artificial Intelligence (ICAART 2021), Online Streaming, February 4-6, 2021*. Ed. by Ana Paula Rocha, Luc Steels and Jaap van den Herik. Vol. 2. SciTePress, 2021, pp. 531–537. DOI: 10.5220/0010188405310537. arXiv: 2011.07990. URL: https://www.scitepress.org/Papers/2021/101884/101884.pdf.
Nominated for ICAART 2021 Best Poster Award.

[9] Markus Schröder, Michael Schulze, **Christian Jilek** and Andreas Dengel. "Bridging the Technology Gap Between Industry and Semantic Web: Generating Databases and Server Code From RDF." In: *Proceedings of the 13th International Conference on Agents and Artificial Intelligence (ICAART 2021), Online Streaming, February 4-6, 2021*. Ed. by Ana Paula Rocha, Luc Steels and Jaap van den Herik. Vol. 2. SciTePress, 2021, pp. 507–514. DOI: 10.5220/0010186005070514. arXiv: 2011.07957. URL: https://www.scitepress.org/Papers/2021/101860/101860.pdf.

[10] Markus Schröder, **Christian Jilek** and Andreas Dengel. "Deep Linking Desktop Resources." In: *The Semantic Web: ESWC 2018 Satellite Events – ESWC 2018 Satellite Events, Heraklion, Crete, Greece, June 3-7, 2018, Revised Selected Papers*. Ed. by Aldo Gangemi, Anna Lisa Gentile, Andrea Giovanni Nuzzolese, Sebastian Rudolph, Maria Maleshkova, Heiko Paulheim, Jeff Z. Pan and Mehwish Alam. Springer, 2018, pp. 202–207. DOI: 10.1007/978-3-319-98192-5_38. arXiv: 1805.03491. URL: https://arxiv.org/pdf/1805.03491.pdf.

[11] Markus Schröder, **Christian Jilek**, Jörn Hees, Sven Hertling and Andreas Dengel. "RDF Spreadsheet Editor: Get (G)rid of Your RDF Data Entry Problems." In: *Proceedings of the ISWC 2017 Posters & Demonstrations and Industry Tracks co-located with 16th International Semantic Web Conference (ISWC 2017), Vienna, Austria, October 23-25, 2017*. Ed. by Nadeschda Nikitina, Dezhao Song, Achille Fokoue and Peter Haase. Vol. 1963. CEUR Workshop Proceedings. CEUR-WS.org, 2017. URL: http://ceur-ws.org/Vol-1963/paper635.pdf.

[12] Sven Hertling, Markus Schröder, **Christian Jilek** and Andreas Dengel. "Where is that Button again ?! – Towards a Universal GUI Search Engine." In: *Proceedings of the 9th International Conference on Agents and Artificial Intelligence (ICAART 2017), Porto, Portugal, February 24-26, 2017*. Vol. 2. SciTePress, 2017, pp. 217–227. DOI: 10.5220/0006201402170227. URL: https://www.scitepress.org/papers/2017/62014/62014.pdf.
Received ICAART 2017 Best Paper Award (Area: Artificial Intelligence).

[13] Sven Hertling, Markus Schröder, **Christian Jilek** and Andreas Dengel. "Top-k Shortest Paths in Directed Labeled Multigraphs." In: *Semantic Web Challenges – Third SemWebEval Challenge at ESWC 2016, Heraklion, Crete, Greece, May 29 - June 2, 2016, Revised Selected Papers*. Ed. by Harald Sack, Stefan Dietze, Anna Tordai and Christoph Lange. Springer, 2016, pp. 200–212. DOI: 10.1007/978-3-319-46565-4_16.
Won ESWC 2016 Top-k Shortest Paths in Large Typed RDF Graphs Challenge.

Technical Reports

[1] Markus Schröder, **Christian Jilek** and Andreas Dengel. "A Linked Data Application Framework to Enable Rapid Prototyping." In: *Computing Research Repository (CoRR)* (2021). arXiv: 2104.13605. URL: https://arxiv.org/pdf/2104.13605.pdf.

[2] Markus Schröder, **Christian Jilek**, Michael Schulze and Andreas Dengel. "Interactively Constructing Knowledge Graphs from Messy User-Generated Spreadsheets." In: *Computing Research Repository (CoRR)* (2021). arXiv: 2103.03537. URL: https://arxiv.org/pdf/2103.03537.pdf.

[3] Markus Schröder, **Christian Jilek**, Jörn Hees and Andreas Dengel. "Towards Semantically Enhanced Data Understanding." In: *Computing Research Repository (CoRR)* (2018). arXiv: 1806.04952. URL: https://arxiv.org/pdf/1806.04952.pdf.

[4] Markus Schröder, **Christian Jilek**, Jörn Hees, Sven Hertling and Andreas Dengel. "An Easy & Collaborative RDF Data Entry Method using the Spreadsheet Metaphor." In: *Computing Research Repository (CoRR)* (2018). arXiv: 1804.04175. URL: https://arxiv.org/pdf/1804.04175.pdf.

List of Figures

Lists of Tables, Listings and Algorithms

List of Tables

List of Listings

List of Algorithms

Lists of Abbreviations, Acronyms and Measurement Units

List of Abbreviations

e.g.	exempli gratia [for example]
esp.	especially
et al.	et alii [and others]
etc.	et cetera [and so on, and more]
i.e.	id est [that is, that means]
max.	maximum
mgmt.	management
min.	minimum
num.	number
org.	organization
temp.	temporal/temporary
vs.	versus

List of Acronyms

Note: For the sake of clarity, only the more frequently used acronyms are listed here. For all other ones, the written-out form should be given in "close proximity" (since those acronyms are typically only used in a single paragraph or section).

AEBI	Activity Event Bus Infrastructure (see Sec. 6.1.1)
AI	Artificial Intelligence
API	Application Programming Interface
ATIC	special kind of knowledge graph whose eight main characteristics start with these letters, see Sec. 5.2.1
BMBF	Bundesministerium für Bildung und Forschung[1] [Federal Ministry of Education and Research]
CalDAV	Calendaring Extensions to WebDAV [75] (also see WebDAV)
CardDAV	vCard Extensions to WebDAV [74] (also see WebDAV)
CoMem	Group-wide Corporate Memory Ecosystem [87] (current Semantic Desktop [80, 304] prototype by DFKI SDS; see Sec. 2.3.3)
CPU	Central Processing Unit

[1] https://www.bmbf.de/

cSpaces	Self-organizing Context Spaces (see Ch. 3 and 8 in particular)
CSS	Cascading Style Sheets [193]
DAG	Directed Acyclic Graph
DAV	see WebDAV
DFG	Deutsche Forschungsgemeinschaft[2] [German Research Foundation]
DFKI	Deutsches Forschungszentrum für Künstliche Intelligenz[3] [German Research Center for Artificial Intelligence]
DTO	Digital Twin of an Organization [277] (see Sec. 9.3.1)
EU	European Union[4]
FIS	Forgetting-enabled ("Forgetful") Information System (see Sec. 3.1)
FST	Finite-state Transducer [292] (see Sec. 6.2.2)
GDPR	General Data Protection Regulation [112] by the EU (see Sec. 4.2)
GIMO	Group Information Model [241] (see Sec. 2.3.2)
GUI	Graphical User Interface
HCI	Human-Computer Interaction [52]
HFCI	Human-*Forgetful Computer* Interaction (see Sec. 3.1)
HTML	HyperText Markup Language [275]
HTTP	HyperText Transfer Protocol [33, 117]
IFI	Intelligent Folder Injection (see Sec. 8.5.4)
IM	Information Management (see PIM and Sec. 2.2)
IMAP	Internet Message Access Protocol [71, 248]
IT	Information Technology
IT-RTC-OBNER	Inflection-tolerant and Real-time Capable Ontology-based Named Entity Recognition (see Sec. 6.2.2)
IVA	Information Value Assessment (see Ch. 7)
JS	JavaScript [101, 102]
JSON	JavaScript Object Notation [72, 45]
KG	Knowledge Graph [168]
KW	Knowledge Work (see Sec. 2.2)
LOD	Linked Open Data [32]
MB	Memory Buoyancy (see Sec. 7.2)
MVN	Maven Central Repository[5]
NER	Named Entity Recognition (see Sec. 6.2)
NLP	Natural Language Processing (see Sec. 6.2)
PC	Personal Computer

[2] https://www.dfg.de/
[3] https://www.dfki.de/
[4] https://european-union.europa.eu/
[5] https://mvnrepository.com/repos/central

PDF	Portable Document Format [37, 176]
PID	Proactive Information Delivery
PIM	Personal Information Management [189] (see Sec. 2.2)
PIMO	Personal Information Model [306] (see Sec. 2.3.2) and preliminary name of CoMem until 2017 (see Sec. 2.3.3)
RAM	Random Access Memory
RDB	Relational Database [67]
RDF	Resource Description Framework [54, 73] (see Sec. 2.3.1)
REST	Representational State Transfer [118]
RPC	Remote Procedure Call [266]
RQ	Research Question (see Sec. 1.3)
SD	Semantic Desktop [80, 304] (see Sec. 2.3)
SDS	Smart Data & Knowledge Services (formerly Knowledge Management) Research Department[6] at DFKI
SMB	Server Message Block (Protocol) [161, 254]
SPARQL	SPARQL Protocol and RDF Query Language [279, 344] (see Sec. 5.2)
SPP	Schwerpunktprogramm [priority program] (by the DFG) (see Sec. 4.2.2)
SQL	Structured Query Language [57, 175]
UI	User Interface
URI	Uniform Resource Identifier [34, 35]
URL	Uniform Resource Locator (special kind of URI, see [35, Sec. 1.1.3.])
W3C	World Wide Web Consortium[7]
WebDAV	Web Distributed Authoring and Versioning [134, 99]
WWW	World Wide Web [31]
YAML	YAML Ain't Markup Language [25, 404] (see Sec. 8.1.2)

List of Measurement Units

The following list contains all measurement units used in the thesis. For the sake of enabling direct lookup (for example by clicking a unit in the text), SI[8] or binary prefixes were not omitted. Thus, the same basic unit may occur multiple times in the list but with different prefixes.

B	byte (1 B = 8 bit)
bit	bit ("binary digit"): basic unit of information
GHz	gigahertz (1 GHz = 10^9 Hz)
GiB	gibibyte (1 GiB = 2^{30} B)

[6] https://sds.dfki.de/
[7] https://www.w3.org/
[8] International System of Units

Hz hertz: basic unit of frequency
min minute (1 min = 60 s)
ms millisecond (1 ms = 10^{-3} s)
s second: basic unit of time
ZB zettabyte (1 ZB = 10^{21} B)

Part I

Introduction, Background and Vision

1 Introduction

This chapter presents the motivation (Sec. 1.1), research questions (Sec. 1.3) and general outline (Sec. 1.5) of the thesis. Especially the structure of its main part is highly influenced by the basic interaction cycle of the envisioned system. Thus, intermediate sections introduce this cycle (Sec. 1.4) as well as a few prerequisite concepts (Sec. 1.2). Last, some general remarks by the author are given, which may help in reading the thesis (Sec. 1.6).

Note: Parts of this chapter have already been published [185].

1.1 Motivation

Knowledge workers face an ever increasing flood of information in their daily work. They live in a "multi-tasking craziness" [137], involving activities like creating, finding, processing, assessing or organizing information while constantly switching from one context to another, each being associated with different tasks, documents, mails, etc. Hence, their personal information sphere consisting of file, mail and bookmark folders as well as their content, calendar entries, etc. is cluttered with information that has become irrelevant. Finding important information thus gets harder and much of previously gained knowledge is practically lost.

For several decades, researchers have worked on improving the situation of knowledge workers, providing many different solutions that in general rank and filter items to reduce cognitive overload (e.g. see the often cited Liu [228] and Shneiderman [353]). However, with the exponential growth of knowledge (see excursus in Section 2.2), an end of this endeavor is not yet in sight. Trying to further improve existing solutions or explore new ways of solving the problem is advisable. This thesis mainly strives for the latter by investigating the potential of self-(re)organizing and especially forgetting-enabled ("forgetful") personal knowledge assistants in the given scenario. Prof. Robert Logie[1], a renowned researcher in the area of human memory and cognition, once said that forgetting was the only "super-human power" people possess [229] metaphorically emphasizing how incredibly well this ability is developed in human beings. It is thus quite surprising that forgetting, in particular also intentional forgetting, is still under-investigated and -utilized in computer science (at least at the time when working on this thesis began). There are at least two directions worth investigating: first, machines mimicking mechanisms of human forgetting, which belongs to the field of human-memory- and cognition-inspired computing, e.g. things slowly fading out of memory when not re-stimulated. The second direction are measures to complement

[1] Human Cognitive Neuroscience Group, Psychology, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Scotland (https://www.ed.ac.uk/profile/robert-logie)

human forgetting, i.e. machines interfere when human forgetting happens (or may happen) unintentionally, e.g. bringing things back into the user's sight by reminders, re-ranking, etc.

In this PhD project, investigations in both direction have been performed. But, as already indicated, forgetting was thereby not seen in isolation but rather as an integral component of self-(re)organization. Without the ability to also select items to be archived, condensed, possibly deleted[2], a system – as self-organizing and thus autonomous as it may be – can more or less only move items "back and forth" without ever "shedding ballast", even if all indicators would advise to do so. A "one-way street" of information emerges and imposes ever increasing challenges to filtering, ranking or visualization procedures, for example. In the author's opinion, forgetting therefore is the last missing component to realize *unhindered* self-organization in the given scenario.

With these first remarks on forgetting and self-(re)organization in mind and supporting knowledge workers as a goal, the next two sections introduce a few prerequisite concepts and state the research questions of this thesis.

1.2 Prerequisite Concepts

Before presenting the research questions and structural overview of this thesis, providing a brief introduction to three important concepts is necessary:

- **Semantic Desktop**: The *Semantic Desktop* [80, 304] is an approach to support information management and knowledge work. As the name suggests, it can be used to express semantic relations between different information items on a user's computer in a machine understandable way, e.g. that the sender of an email works in a certain project or a file is about a certain topic.

- **Managed Forgetting**: *Managed Forgetting* is an escalating set of measures to overcome the binary keep-or-delete paradigm. According to its original definition stated in the EU project *ForgetIT* [120], they range from temporal hiding, to condensation, to adaptive synchronization, archiving and deletion (see Sec. 3.1). However, since the author of this thesis joined the ForgetIT project at the beginning of its second year in 2014, there was the opportunity to actively shape and contribute to the idea, first as a diploma student [178] and student research assistant and from 2015 on as a researcher and PhD student. Apart from contributing to the further elaboration and prototypical realization of Managed Forgetting (e.g. see [244, 242, 184, 245, 246, 407, 188, 240, 373]), the author later extended the idea by emphasizing more on self-(re)organization aspects especially also taking the whole life cycle (i.e. additionally early phases) of an information item or context into account (e.g. see [185, 187, 182]).

- **Context Spaces**: *Context Spaces* are an extension to the Semantic Desktop developed in this thesis to better capture and utilize users' different working contexts. The main idea

[2] three examples of forgetting (removal) – archiving: item removed from the user's view, condensation: removal of (some) details, deletion: total removal of an item; also see later section on Managed Forgetting in this regard (Section 3.1)

is to make contexts an explicit interaction element allowing to work *with* (i.e. a "tangible" object similar to a folder) and *in* them (immersion). Users may thus make more of their personal mental model explicit for the machine in a subtle way resulting in an increased amount of contextual meta-information that can be exploited to better support them in return. One goal is to increase the Semantic Desktop's degree of automation and provide services not yet available in similar systems today. These services are in particular based on Managed Forgetting and especially self-(re)organization ideas.

Bringing Semantic Desktop, Context Spaces and Managed Forgetting together, the envisioned and prototypically realized idea of this thesis is called **Self-organizing Context Spaces**, or **cSpaces** for short. All of these terms will be introduced more thoroughly in Chapters 2 and 3. However, the level of detail already provided should be sufficient to comprehend the research questions, which will be stated in the following.

1.3 Research Questions

The main research question (RQ) this thesis aims to answer is:

RQ1: How to realize a self-(re)organizing and especially forgetting-enabled work environment that

(a) eases focus,

(b) eases keeping an overview and

(c) helps keeping the user's personal information sphere (more) tidied up?

This question raises further sub- and follow-up questions.

RQ2: What support measures in particular could such a system provide to achieve the goals stated in RQ1 and how to realize them?

RQ3: Are *Context Spaces* a suitable user interface for such a system?

RQ4: How to enable users to actually work *with* and *in Context Spaces*?

RQ5: Is the *Semantic Desktop* [80, 304] a suitable ecosystem for realizing *Self-organizing Context Spaces*?

(a) What needs to be adapted or added?

(b) What can be learned from failed attempts (like *Nepomuk-KDE* [268, 154]) to establish the *Semantic Desktop* as a widespread tool in practice?

(c) How to bootstrap such a system, i.e. how to overcome (or at least reduce) the cold start problem of such systems?

RQ6: How to context-sensitively assess the value of information items and thus perform the "right" self-(re)organization measures?

RQ7: How could search interfaces for such a self-(re)organizing and especially forgetting-enabled information system look like?

RQ8: Will users trust such a highly autonomous system and what can be done to facilitate trust in this scenario?

These research questions are addressed in Chapters 3 to 9 and finally concluded in Chapter 10.

A more detailed overview of this thesis' structure is presented in the section after next. First, the following section introduces the basic interaction cycle of cSpaces, according to which the main part of the thesis is structured.

1.4 The cSpaces Interaction Cycle

In the first sections of this chapter, the idea of **Self-organizing Context Spaces (cSpaces)**, a self-(re)organizing and especially forgetting-enabled personal knowledge assistant, has been brought up. cSpaces are, to the author's best knowledge, the first incarnation of such a system. Since they are based on Semantic Desktop [80, 304] technology, one could more generally also speak of a *Self-(re)organizing Semantic Desktop*. The interaction of users with such a system can be understood as a cycle of six steps, shortly summarized in the following and also depicted in Figure 1.1:

1. **Evidence Collection:** observing user activities, e.g. browsed websites, files, mails, etc.
2. **Information Extraction:** analysis of gathered evidence snippets, e.g. finding relevant concepts like person or organization names, locations, topics, etc.
3. **Context Elicitation:** assigning each observed user action to one or more plausible contexts
4. **Information Value Assessment:** estimating an information item's current (or future) value for the user
5. **Self-(re)organization Measures:** evaluation and exploitation of analysis results and taking appropriate actions to support the user
6. **User Interface Updates:** updating the (graphical) user interface accordingly

After step 6, the cycle repeats itself starting over with step 1. Apart from these steps, two further aspects need to be considered: First, **0. Data Storage and especially Knowledge Graphs** are relevant for all phases of the cycle (thus drawn in the center of the figure). This especially involves the so-called *Personal Information Model* [306], or *PIMO* for short, which is a personal *Knowledge Graph* (KG) [168] and one of the cornerstones of the Semantic Desktop. Second, there are also relevant **7. Socio-technical Issues and Cognitive Psychology Aspects** (see upper-left corner of the figure).

The next section provides a structural overview of the thesis.

1.5 Thesis Overview

Apart from the appendices, this thesis consists of three parts, which are briefly summarized in the following. Its middle part covers most contributions with realizing cSpaces being the most important one. This part is thus structured according to the steps and aspects of the cSpaces interaction cycle mentioned before.

Part I: Introduction, Background and Vision

Part I of the thesis introduces into the topic (Ch. 1), and necessary background information is given (Ch. 2). After laying these foundations, the general approach and vision is sketched (Ch. 3) followed by an overview of related work (Ch. 4). This part of the thesis thus consists of three more chapters apart from the current one:

Chapter 2. First, relevant background information and foundations are provided in Chapter 2. This comprises definitions of terms like *data*, *information* and *knowledge* (2.1), *Information Management (IM)* and *Knowledge Work (KW)* (2.2). An overview of challenges in IM and KW support complements the first part of the chapter. The aforementioned *Semantic Desktop* (2.3) is introduced in the chapter's second part. In particular, the last section (2.3.3) presents the current Semantic Desktop prototype developed by the *Smart Data & Knowledge Services (SDS)* research department at DFKI, which is called *CoMem*.

Chapter 3. Chapter 3 presents the general approach and vision of *Self-organizing Context Spaces* (3.3) after introducing *Managed Forgetting* (3.1) and *Context Spaces* (3.2) as its two other core pillars beside the Semantic Desktop.

Chapter 4. Having provided the necessary background and vision of the thesis, Chapter 4 addresses related work. While one monolithic and several smaller, distributed related work sections both have their pros and cons, this thesis strives for a hybrid solution: The first part of the chapter provides general related work and overviews for the three cornerstones of cSpaces: the *Semantic Desktop* (Sec. 4.1), *Managed Forgetting* (Sec. 4.2) and *Context Spaces* (Sec. 4.3), whereas the last section also includes the overall idea of *Self-organizing Context Spaces*. From a high-level perspective, related ideas, research projects and groups are introduced, to which later sections may refer to. The idea is to avoid readers being overwhelmed with lots of very specific details before getting a proper introduction to each subtopic first. These detailed aspects are then addressed in related work sections dedicated to a specific subtopic or more briefly as part of that subtopic's introduction. After a conclusion on general related work (Sec. 4.4), an overview of these specific related work sections is provided in Section 4.5.

Part II: Approach

The second part of the thesis covers the main approach (Chapters 5 to 8). As mentioned before, its chapters are aligned with the interaction cycle of cSpaces – see Figure 1.1 for an overview and Table 1.1 for an aspects-to-chapters mapping.

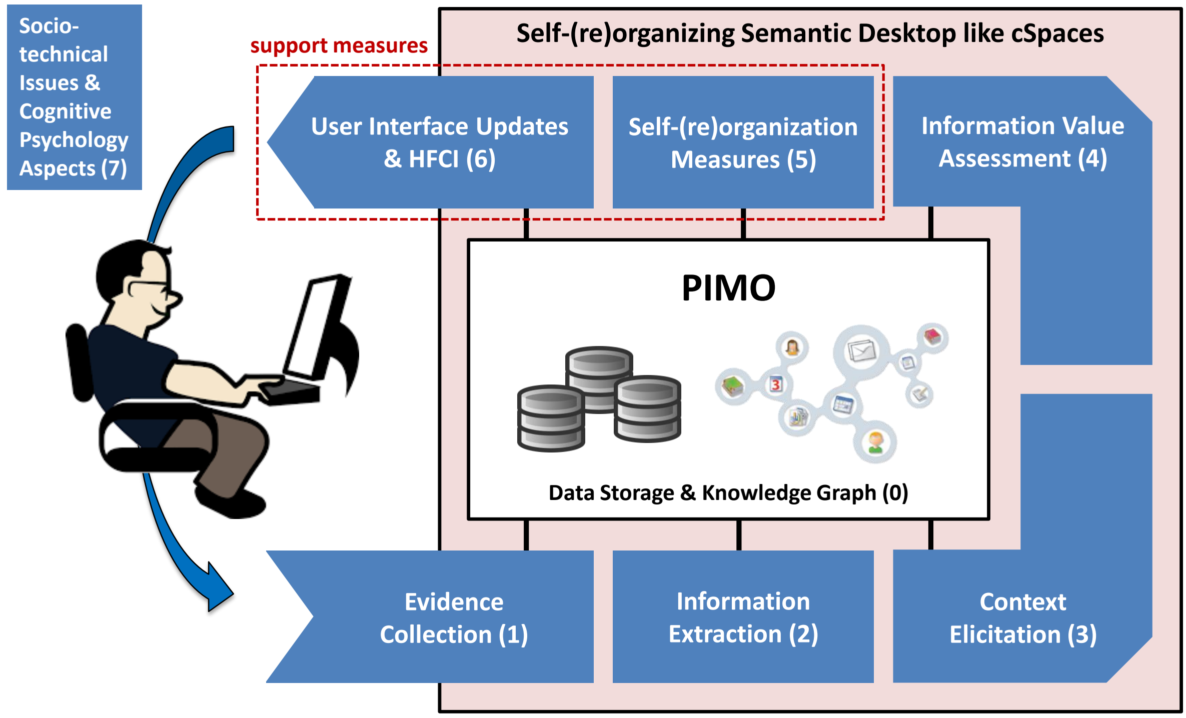


Figure 1.1: Interaction Cycle of a Self-(re)organizing Semantic Desktop like cSpaces.
(Note: An earlier version of this figure has been published in Jilek et al. [185].)

Chapter 5. Chapter 5 first presents context modeling (5.1) and storage aspects (5.2), i.e. step 0 of the interaction cycle. The latter in particular includes specialized, so-called *ATIC Knowledge Graphs* (5.2.1 and 5.2.2) and means for efficiently searching them (5.2.3).

These prerequisite aspects are complemented by a section on bootstrapping such systems to overcome – or at least reduce – their so-called *cold start problem* (5.3), i.e. a system being "empty" in the beginning since no information about the user, interaction data etc. are available to draw conclusions from and perform support measures accordingly. A three-step approach involving three different components is presented: First, the *PIM Crawler* (5.3.1) collects all relevant information items ("PIM data") available on a user's computer, i.e. data typically in focus of *Personal Information Management (PIM)* like files, mails, bookmarks, calendar entries, tasks, etc. (only the first four are supported by the current prototype). Second, a semantic leveraging is performed using the *Semantifier* (5.3.2). In the current implementation, this primarily means concept mining on the obtained PIM data. Last, the *Contextifier* (5.3.3) tries to identify user contexts spread across the different data systems, e.g. a file could have been the attachment of an email, things stored in a certain folder all belong to a specific meeting mentioned in the calendar, etc. Before its first start-up, the cSpaces system can then be pre-filled with the obtained contexts. Since the *Semantifier* and the *Contextifier* are both interactive tools, users may actively contribute to concept and context mining on their data making more of their mental model explicit for the machine, e.g. by adding concepts, giving feedback on suggestions, merging or splitting contexts, identifying preferred candidates upfront, etc. The outcome of this bootstrapping process is then a personal knowledge graph for each individual user (i.e.

Step/Aspect	Chapter	Priority
0. Data Storage & Knowledge Graph	5	third
1. Evidence Collection	6.1	third
2. Information Extraction	6.2	third
3. Context Elicitation	6.3	third
4. Information Value Assessment	7	second
5. Self-(re)organization Measures	8	first
6. User Interface Updates & HFCI	8	first
7. Socio-technical Issues & Cognitive Psychology Aspects	8, 7, 6, 3	second

Table 1.1: Mapping of cSpaces Aspects to Chapters and Research Priorities.

an initial version of their PIMO), in which these concepts and contexts are represented.

With regard to personal knowledge graph bootstrapping, it is worth mentioning that this PhD project's most prominent side issue was contributing to research on personal or corporate knowledge graph construction in general. Therefore, this chapter closes with an excursus section (5.3.4) briefly reporting on these activities. They were led by the author's colleague, Markus Schröder, as part of his PhD project [323].

Chapter 6. Steps 1 to 3 of the interaction cycle, i.e. evidence collection (6.1), information extraction (6.2) and context elicitation (6.3), are addressed in Chapter 6. Although they were not planned to be a core contribution, specifically tailored solutions for these steps still had to be found in order to allow for the main research to be conducted (steps/aspects 4 to 7). A challenge addressed in this and the previous chapter was to strike a balance between dealing with highly dynamic aspects and requirements that are easier met in static settings: On the one hand, there are continuously evolving knowledge graphs potentially being updated multiple times within a second. And on the other hand, real-time capable analysis and querying is needed, which typically involves heavy optimization, pre-calculations, etc. easier possible with static content. As a consequence, some of the found solutions became contributions of their own.

Chapter 7. Chapter 7 is about *Information Value Assessment (IVA)* – step 4 of the interaction cycle. It presents multiple variants of *Memory Buoyancy (MB)*, a measure for short- and medium-term information value (7.2). To the author's best knowledge, MB is the only IVA approach inspired by human memory and cognition. Additionally, an overview of existing IVA approaches is given (7.1), and the *Preservation Value* is introduced in Section 7.3 as a long-term counterpart to MB.

Chapter 8. As highlighted in Figure 1.1, steps 5 and 6 are the actual support measures, whereas step *5. Self-(re)organization Measures* focuses more on back-end aspects and step *6. User Interface Updates* more on the front-end. When discussing user interface-related issues, the thesis also addresses more general aspects of interacting with a self-(re)organizing and especially forgetting-enabled information system. In particular, forgetting marks such a distinct paradigm shift that this thesis also speaks of Human-*Forgetful Computer* Interaction (HFCI), emphasizing its quite novel focus within the overall area of Human-Computer Interaction (HCI) [52]. Step 6 is therefore actually called "User Interface Updates *and HFCI*". Since back- and front-end aspects are so highly interrelated, they are addressed together in Chapter 8.

The chapter first discusses evaluation aspects (8.1) like challenges in the research area of IM and KW support in general (8.1.1) and the chosen strategy, methods and tools in particular (8.1.2). Next, the prototypical implementation of cSpaces with its different variants and user interfaces is presented in Section 8.2. The main part of the chapter covers all experiments and studies conducted with variants of cSpaces. Section 8.3 discusses working *with* and *in* (Self-organizing) Context Spaces in general and measured benefits in particular. The rest of this part is basically sorted by the different types of support measures: condensation and summarization (8.4), temporal reorganization, fading out and resurfacing (8.5) and permanent reorganization (8.6). Although HFCI aspects are also part of other sections, the last part of the chapter particularly focuses on them: Searching (8.7) and trust (8.8) in such highly autonomous, *forgetting-enabled information systems* (FIS) like cSpaces are discussed and various prototypes are presented.

In general, evaluation sections are spread across the thesis for several aspects individually. However, by building upon all previous chapters and covering most studies and experiments, Chapter 8 contains most of the evaluation sections. It is therefore comparable to a dedicated evaluation chapter. In short, evaluation sections are thus organized in the same hybrid way as related work sections in this thesis (see summary of Chapter 4 above).

Part II Summary. Table 1.1 summarizes Part II by mapping aspects of the cSpaces interaction cycle to chapters of the thesis. The author's main research focus (first priority) was on support measures (steps 5 and 6), followed by IVA approaches (4) and socio-technical issues and cognitive psychology aspects (7) as a second priority. The latter (7) is a meta-topic and therefore present in various chapters as shown in the table. Finally, steps/aspects 0 to 3 were the third and last research priority.

Part III: Conclusion and Outlook

Part III discusses the transfer into practice of some of the ideas and prototypes developed in this PhD project as well as further potential and application scenarios (Ch. 9). A last chapter contains this thesis' overall summary, conclusion and outlook (Ch. 10).

Chapter 9. After presenting this thesis' main approach, Chapter 9 discusses its transfer into practice as well as further potential and application scenarios only briefly investigated so far. The chapter is ordered according to five descending levels of maturity:

1. the most elaborated ideas and stable prototypes of the main approach which have already found their way into practice, e.g. as software components, parts of feasibility studies or as a mere knowledge transfer in the form of consulting jobs (9.1),
2. elaborated ideas realized in early prototypes (9.2),
3. elaborated ideas with experimental code in pre-prototypical state,
4. elaborated ideas sketched in position papers (9.3) and
5. early ideas (9.4).

Although there is nothing else to report in category 3, the author still decided to mention it here since these levels of maturity will also be used when summarizing all contributions of this PhD project in the final chapter.

Chapter 10. Chapter 10 summarizes (10.1) and concludes (10.2) this thesis and gives an outlook on possible future work (10.3).

Part IV: Appendices

Appendix A. Mainly for the sake of completeness, Appendix A contains additional details on conducted experiments and studies.

Appendix B. There are a few general remarks by the author, of which the most important ones are stated in the next section (1.6) while others can be found in Appendix B.

Appendix C. Last, Appendix C contains software references and dependencies – also see general remarks on this matter in the following.

1.6 General Remarks

There are a few general remarks by the author that may help in reading the thesis. Some of them can be seen as "global footnotes" relevant for several passages of the thesis. Instead of stating them several times, they are written down in a single location and referred to multiple times later on. The most important ones are stated in the following, while others, for example about technical protocol and standard citations (B.1) or used box plot semantics (B.2), can be found in Appendix B.

Remark on Software References

Software references are listed in Appendix C. Since they typically point to a software artifact, i.e. source or compiled code to be executed or imported into one's project to gain a certain functionality, the author decided to separate these references from the ones pointing to primarily textual/prose content (scientific papers in most cases). Datasets or other assets that are rather meant to be processed by machines are also part of these references.

Software references are denoted with "Sw" as a prefix to their given number and have a different link color. If a scientific paper is associated with a software artifact relevant for the thesis, the entries in each reference list refer to each other reflecting this association. For example, this is the case with Zesch et al. [405] and the *Java-based Wiktionary Library (JWKTL)* [Sw115].

Remark on Questionnaires

Several surveys were conducted as part of this PhD project. All of them can be separated into multiple thematic clusters. Especially the final interview questionnaire of a multi-month user study on cSpaces covers a lot of different topics. Consequently, questionnaire clusters are mentioned in different passages of the thesis. In order to ease keeping and overview, the author decided against thematic splitting and renaming their items and instead introduced a prefix for each questionnaire followed by the individual item's number. For example, Q_{M1}, Q_{M2} and Q_{M3} refer to the first three items of the aforementioned interview questionnaire. As a drawback of this labeling method, the cluster an item belongs to is not reflected by its label and has to be given by the surrounding text passage, e.g. the three previously mentioned items all belong to the questionnaire's first cluster of *contextual note taking*. The advantage is, however, that each item's position in the overall questionnaire is obvious and thus easy to find. This is especially relevant for questionnaires that are also available as a whole (full document) in Appendix A. All questionnaires and their abbreviations are listed in Table 1.2.

Abbr.	**Questionnaire**
Q_C	**C**ontextifier
Q_M	**M**ulti-month User Study: Final Interview
Q_S	FIS **S**earch Prototype II

Table 1.2: Overview of Questionnaire Abbreviations.

2 Background and Foundations

Having presented the motivation, research questions and general overview in the previous chapter (Ch. 1), the next steps are providing necessary background information and foundations of this work. This is done across three chapters: First, the current chapter (Ch. 2) introduces very fundamental terms and concepts that existed long before this PhD project started. The introduction continues in the next chapter, but concepts presented there have been envisioned or significantly influenced and enhanced by the author making them part of this thesis' overall vision and approach (Ch. 3). Having provided all necessary background information, an overview of related work is given in Chapter 4, which concludes the introductory part of the thesis.

The first part of this chapter provides definitions of *data*, *information* and *knowledge* (Sec. 2.1) as well as *Information Management (IM)* and *Knowledge Work (KW)*. It is complemented by an overview of existing challenges in IM and KW support (Sec. 2.2). Its second part, Section 2.3, introduces one of the three core pillars of *Self-organizing Context Spaces (cSpaces)*, the *Semantic Desktop*, which may serve as an ecosystem for approaches supporting IM and KW. Apart from the Semantic Desktop itself, its major cornerstone, the *Personal Information Model (PIMO)*, as well as *CoMem*, the current prototype developed by DFKI SDS, are presented.

Note: Parts of this chapter have already been published [182, 240].

2.1 Data, Information and Knowledge

Before addressing *information management* and *knowledge work*, the more basic terms of *information* and *knowledge* need to be defined first. This thesis adopts the *"Wissenstreppe" [Knowledge Ladder]* by North et al. [272].

The Knowledge Ladder. Roughly speaking (details will follow), the knowledge ladder starts with *symbols* at the bottom. Adding *syntax* leads to *data*, adding *meaning* leads to *information*, and *connecting information* including (personal) context, experience and expectations leads to *knowledge* [272]. Phrasing the last part even shorter, one could say: knowledge is *information turned into action* [82]. More thorough definitions read as follows:

Definition: **Data, Information and Knowledge**

Zeichen (z.B. Buchstaben, Ziffern, rotes Licht einer Ampel) werden durch Ordnungsregeln (einen Code oder eine Syntax) zu Daten.	*Symbols (e.g. characters, figures, red traffic lights) become data by rules of order (a code or a syntax).*
***Daten** sind Fakten über ein Objekt, wobei ein Objekt eine Einheit, ein Gegenstand, etwas Wahrnehmbares oder Vorstellbares sein kann.*	***Data** are facts about an object, whereby an object can be an entity, an item, something perceptible or imaginable.*
***Informationen** sind Daten, die in einem Bedeutungskontext stehen. Informationen sind für Betrachter wertlos, wenn sie diese nicht mit anderen aktuellen oder in der Vergangenheit gespeicherten Informationen vernetzen können.*	***Information** is data in a context of meaning. Information is worthless to observers, if they cannot connect it with other current information or information stored in the past.*
*Aus dieser Sicht ist **Wissen** der Prozess der zweckdienlichen Vernetzung von Informationen. Wissen entsteht als Ergebnis der Verarbeitung von Informationen durch das menschliche Bewusstsein. Informationen sind sozusagen der Rohstoff, aus dem Wissen generiert wird und die Form, in der Wissen kommuniziert und gespeichert wird. Die Interpretation von Informationen kann insbesondere in unterschiedlichen kulturellen Kontexten sehr unterschiedlich ausfallen. Wissen ist daher geprägt von individuellen Erfahrungen, ist kontextspezifisch und an Personen gebunden.*	*From this point of view, **Knowledge** is the process of purposefully connecting information. Knowledge is created as a result of the processing of information by human consciousness. Information is, so to speak, the raw material from which knowledge is generated and the form in which knowledge is communicated and stored. The interpretation of information can vary greatly, especially in different cultural contexts. Knowledge is therefore shaped by individual experiences, it is context-specific and bound to individuals.*
K. North et al. [272]	(English translation by the author)

Excursus: The Exponential Growth of Knowledge. The motivation section of this thesis (Sec. 1.1) mentioned the exponential growth of knowledge. Having introduced the terms above, this paragraph can now provide a bit more background information on this statement. The *knowledge doubling curve* by Fuller [125] is well-known: He noticed that until 1900 human knowledge doubled approximately every century and by the end of World War II it was doubling every 25 years [56]. There is also an anecdotal prediction by IBM[1] [173] from 2006 according to which knowledge is already doubling on an hourly basis. This claim has been widely treated as authoritative on the web, e.g. Schilling [315], although any source or evidence had been given as pointed out in detail by Dorbolo [95]. However, even though an exact rate

[1] https://www.ibm.com/

of growth is hard to determine, Fuller's observations already indicate exponential growth, and, four decades later, Chamberlain [56] sees it as "generally acknowledged that human knowledge is increasing at an extraordinary rate". This statement is supported by statistics about the growing amount of Big Data [59]: For example, Bahri et al. [20] refer to a 2017 report of the International Data Corporation (IDC)[2] showing a growth rate of a few zettabytes (ZB) in 2010 to 163 ZB predicted for 2025. This data can be categorized into *social media data* (e.g. posts in social networks), *machine-to-machine data* (e.g. sensors, meters, other devices), *transaction data* (e.g. recovered from fingerprints, genetics, handwriting, medical images), *human-generated data* (e.g. prescriptions, email, messages, documents, electronic medical reports) and *web data* (e.g. captured click streams) [20]. In 2018, the IDC revised up their prediction to 175 ZB in 2025 growing from 33 ZB in 2018 [285]. Although data and information do not translate into knowledge one-to-one (see definitions above), there is nevertheless a notable portion of human knowledge emerging from (exponentially growing) Big Data alone – with many other knowledge sources not even considered. It is therefore reasonable to assume that the growth rates of data and knowledge are basically similar and exponential in both cases.

2.2 Information Management (IM) and Knowledge Work (KW)

Having provided definitions of *data*, *information* and *knowledge*, this section introduces the concepts of *Information Management (IM)* and *Knowledge Work (KW)* as well as challenges for approaches supporting them.

Information Management (IM). In analogy to knowledge in an organization having individual and shared parts [84], *Information Management* comprises personal and organizational aspects. The former is known as *Personal Information Management (PIM)*, while the latter is mainly addressed in this thesis by the concept of an *Organizational Memory*. Both terms will be explained in the following:

> Definition: **Personal Information Management (PIM)**
>
> *"Personal information management (PIM) refers to both the practice and the study of the activities a person performs in order to acquire or create, store, organize, maintain, retrieve, use and distribute the information needed to meet life's many goals (everyday and long-term, work-related and not) and to fulfill life's many roles and responsibilities (as parent, spouse, friend, employee, member of community, etc.).*
> *PIM places special emphasis on the organization and maintenance of personal information collections in which information items, such as paper documents, electronic documents, email messages, web references, handwritten notes, etc. are stored for later use and repeated re-use."*
>
> W. Jones [189]

[2] https://www.idc.com/

Definition: **Organizational Memory**

Ein Organizational Memory ist „ein Computersystem, das in der Organisation Wissen und Information fortlaufend sammelt, aktualisiert und strukturiert und für verschiedenste Aufgaben kontextabhängig, gezielt und aktiv zur Verfügung stellt."	*An Organizational Memory is a computer system that continuously collects, updates and structures knowledge and information in the organization and context-sensitively, specifically and actively provides them for various tasks.*
A. Dengel et al. [84]	(English translation by the author)

On the one hand (bottom-up), information and knowledge emerge with an individual while performing typical PIM activities. Parts of them subsequently spread into surrounding groups like a team, the department, the organization. On the other hand (top-down), an individual is influenced by shared information and knowledge of the organization: they adapt it, use it, extend it, benefit from it, etc. The interaction of individuals with the organizational memory can thus be seen as a bidirectional relation of adding/contributing and receiving.

Knowledge Work (KW). The aforementioned "individuals" performing PIM activities are actually *knowledge workers*. A definition of the term reads as follows:

Definition: **Knowledge Work (KW) / Knowledge Worker**

"Knowledge workers have high degrees of expertise, education, or experience, and the primary purpose of their jobs involves the creation, distribution, or application of knowledge. Knowledge workers think for a living. They live by their wits – any heavy lifting on the job is intellectual, not physical. They solve problems, they understand and meet needs of customers, they make decisions, and they collaborate and communicate with other people in the course of doing their own work." [77, p. 10 f.]

"I would put the following [U.S. worker] categories into the knowledge worker camp: Management, Business and financial operations, Computer and mathematical, Architecture and engineering, Life, physical, and social scientists, Legal, Healthcare practitioners, Community and social services, Education, training, and library, Arts, design, entertainment, sports, media." [77, p. 5 f.]

T. H. Davenport [77]

Challenges in IM and KW Support. Over several decades, many approaches have been presented to support IM and KW. Shortly before working on this PhD project started, Warren published a comprehensive overview article called "PIM: The Case for an Evolutionary Approach" [391], in which he still saw three "**pressing requirements for which current [PIM] systems [did] not fully cater**:

[P1]: To combat information overload, as the volume of information increases.

[P2]: To ease context switching, in particular, for users who face frequent interrupts in their work.

[P3]: To be supported in information integration, across a variety of applications" [391].

P1 and P2 were also addressed in this thesis' motivation (Sec. 1.1). In particular, with regard to P2, a study by González and Mark [137] was hinted at which showed that knowledge workers switch contexts about every twelve minutes, thus speaking of a "*constant, constant, multi-tasking craziness*". Not only is PIM fragmented over time, but also over storage systems: This is known as the *project fragmentation problem* [27], which is a part of P3. Although information items belong together in a person's mind (e.g. since they all belong to the same project – or the same *context*, more generally speaking), users store them in various classification schemes: documents in file folder hierarchies, mails in a mail folder hierarchy and bookmarks in the web browser's bookmark folder hierarchy.

Warren [391] further provided an overview of several studies with observations, strategies and behaviors regarding PIM. The following is a selection of addressed topics particularly interesting with regard to cSpaces:

- analyses regarding the overlap in file, mail and bookmark folder structures [(28) as cited in (391)],
- individuals preferring navigation to search [(29) as cited in (391)],
- users reluctant to give up folders, even if provided with fast search tools [(190) as cited in (391)],
- large proportions of an individual's personal files rarely or never accessed [(30) as cited in (391)].

Taking this into account, Warren [391] identified four **broad technological themes** that he found **promising with regard to solving these problems**. They have a high overlap with ideas of cSpaces and are as follows:

[T1]: "A unified file system for all personal information, and also for shared information.

[T2]: The introduction of tagging in an evolutionary way which combines the benefits of tags with the benefits of folders.

[T3]: The management of the user's contextual state to support information retrieval and reduce the overhead of context switching.

[T4]: The use of semantic technologies, e.g. machine-learning and natural language processing accompanied by standards such as RDF(S)[3] and OWL[3], to aid information retrieval and support information exchange between applications." [391]

cSpaces and other more recent approaches to support IM and KW involve the Semantic Desktop [80, 304] or similar technologies as an ecosystem, which inherently addresses P3,

[3] RDF [54, 73], RDFS [142, 143] and OWL [167] are introduced in the next section.

T2 and T4. As mentioned, most of the other problems and technological themes are also addressed by cSpaces. This is complemented by a close collaboration with cognitive scientists in its development, for which Warren [391] also argued.

There is also an earlier position paper by Voit et al. called "Why PIM Technologies Are Not Widespread (And What to Do About It)" [387], in which the authors stated **eight requirements for future PIM tools to improve the situation**. These requirements are denoted with V1 to V8 in the following. Like Warren, they also argue for supporting context switches **[V3]** and storing additional metadata **[V8]**. From the perspective of cSpaces, it is particularly interesting that they further recommend to "radically [improve] current hierarchy browsing mechanisms **[V4]** but to also "be compatible with current user habits" **[V1]** [387]. They further argue that a typical user is unwilling to install special tools like an alternative file system if it is not guaranteed to be compatible with their familiar operating system tools [387]. Additionally, they support the idea of introducing expiry dates **[V7]** to information items as proposed in Mayer-Schönberger [247], one of the first and nowadays much-cited works on digital forgetting. These last two ideas go into the direction of *Context Spaces* and *Managed Forgetting* and are addressed again in the next chapter when introducing cSpaces in more detail. (For the sake of completeness, the other requirements stated by Voit et al. are minimal interfaces to keep learning efforts low **[V2]**, no unnecessary limitations **[V5]** and transparency **[V6]** [387].)

A last major challenge to be mentioned here is subjectiveness in the actual IM and KW support as well as the evaluation of proposed methods and tools: users have subjective views on their data [88, 28]. Thus, different users would likely file the very same information item under different keywords and at very different positions in a folder hierarchy (assuming a shared one in this example). Dengel [88] and studies by Bergman et al. [28] address subjectiveness in PIM, for example. Section 8.1.1 further addresses this topic.

Before getting into further details of cSpaces in Chapter 3, the next section introduces the first and most fundamental of its three cornerstones, the *Semantic Desktop*. (The other cornerstones, *Managed Forgetting* and *Context Spaces*, follow in the next chapter when presenting the overall vision and approach of this thesis.)

2.3 The Semantic Desktop as an Ecosystem for IM and KW Support

This section is about the Semantic Desktop [80, 304], especially from the perspective of the SDS research department at DFKI that was heavily involved in various Semantic Desktop-related projects over the last two decades. By far the most prominent one was "NEPOMUK – The Social Semantic Desktop" (2006–2008) [267], to the author's best knowledge, the so far highest budgeted European project dedicated to Semantic Desktop research (11.5 million EUR). It was coordinated by DFKI, and there were several directly adjacent DFKI projects either before, in parallel or after it. Additionally, the experiences and spirit of NEPOMUK were carried over to many other, even much later projects, of which especially this PhD thesis also was a part of. Thus, this section focuses on DFKI SDS' ideas and realization of a Semantic Desktop. However, other projects, alternative developments as well as a summarizing high-level perspective on the topic are given in the related work chapter that follows, especially Section 4.1.

In the next section (2.3.1), the Semantic Desktop idea as well as a few prerequisites like the

Semantic Web [36] are presented. Section 2.3.2 then focuses on one of the Semantic Desktop's core components, the *Personal Information Model (PIMO)* which serves as the basis for its knowledge representation. Last, Section 2.3.3 presents *CoMem*, DFKI SDS' current Semantic Desktop prototype.

2.3.1 Semantic Desktop – Bringing the Semantic Web to Users' Desktops

The basic idea of the Semantic Desktop is to bring Semantic Web technology to users' desktops. Both concepts are presented more thoroughly in the following.

The Semantic Web. The idea of a Semantic Web was introduced in 2001 by Berners-Lee et al. [36] followed by activities of the *World Wide Web Consortium (W3C)*. Their activity statement [162] reads as follows:

> **Semantic Web Activity Statement**
>
> *"The goal of the Semantic Web initiative is as broad as that of the Web: to create a universal medium for the exchange of data. It is envisaged to smoothly interconnect personal information management, enterprise application integration, and the global sharing of commercial, scientific and cultural data. Facilities to put machine-understandable data on the Web are quickly becoming a high priority for many organizations, individuals and communities.*
> *The Web can reach its full potential only if it becomes a place where data can be shared and processed by automated tools as well as by people. For the Web to scale, tomorrow's programs must be able to share and process data even when these programs have been designed totally independently. [...]"*
>
> I. Herman [162], Semantic Web Activity Lead

In short, the Semantic Web is an extension of the traditional *World Wide Web (WWW)* making its data machine-understandable. One of the goals is to allow for higher automation.

Some of the principal technologies of the Semantic Web, also coordinated, standardized and published by the W3C, are as follows:

- **RDF:** In a nutshell, the *Resource Description Framework (RDF)* [54, 73] is meant to express statements in the form of *(subject, predicate, object)-triples*, or *(s, p, o)-triples* for short. Such a triple can be seen as an edge in a (knowledge) graph going from the subject node to the object node, whereas the edge itself is the predicate: $S \text{ –}P\rightarrow O$. Each part of the statement, can be addressed using a *Uniform Resource Identifier (URI)*. For readers not familiar with the Semantic Web, the most well-known URI is probably the URL, a special kind of URI to address resources on the WWW. The object of a statement may also be a *literal*, e.g. a textual label, a numerical value, a currency, etc. Especially in the context of CoMem, these literal statements are also referred to as *(s, p, l)-triples*.

- **SPARQL:** *SPARQL* [279, 344] is a standard query language for RDF graphs.

- **RDFS and OWL:** *RDF Schema (RDFS)* [142, 143] and the *Web Ontology Language (OWL)* [167] are ontology languages to describe RDF data (terminologies, vocabularies). The language defined by RDFS basically consists of *resources* – everything described in RDF is a resource. Resources can be further divided into *classes*, and classes can have concrete *instances* similar to object-oriented programming. Each of these instances then has a *hasType*-relation to its respective class, which is a very basic *property*. Such *properties* can be defined between resources/classes and have a *domain* and *range*, which both again point to a class. Sub-classes and sub-properties may also be defined. This short introduction to RDFS is a very brief excerpt of Guha and Brickley [142], intended to be just enough to comprehend topics addressed in this thesis. OWL has a higher expressiveness than RDFS, for example allowing to express cardinality restrictions or disjointness of classes. However, CoMem mainly uses RDFS except for stating *sameAs* (e.g. *München –sameAs→ Munich*[4]) and *inverse object property* relations (e.g. *isPartOf –inverseObjectPropertyOf→ hasPart*[4]). Since cSpaces is derived from CoMem (with the intention to stay compatible), OWL is not relevant for this thesis apart from the mentioned exceptions.

In this brief Semantic Web introduction, the term *ontology* was mentioned multiple times without providing a definition, yet. This thesis adopts the one by Studer et al. [367]:

> Definition: **Ontology**
>
> *"An ontology is a formal, explicit specification of a shared conceptualization.*
>
> *A 'conceptualization' refers to an abstract model of some phenomenon in the world by having identified the relevant concepts of that phenomenon. 'Explicit' means that the type of concepts used, and the constraints on their use are explicitly defined. For example, in medical domains, the concepts are diseases and symptoms, the relations between them are causal and a constraint is that a disease cannot cause itself. 'Formal' refers to the fact that the ontology should be machine readable, which excludes natural language. 'Shared' reflects the notion that an ontology captures consensual knowledge, that is, it is not private to some individual, but accepted by a group."*
>
> R. Studer et al. [367]

For the sake of completeness, it is worth mentioning that (triple) statements can be divided into *terminology box (TBox)* and *assertion box (ABox)* statements [(83) referring to (43)]. The TBox contains statements about the aforementioned classes and properties, which can be used as vocabulary to describe a domain of interest, while ABox statements are assertions about instances of these classes. This thesis does not explicitly distinguish between the two.

Having introduced all necessary terms and concepts, the next paragraph presents the Semantic Desktop in more detail.

The Semantic Desktop. The term *Semantic Desktop* was coined in 2003 [96] by Decker and Frank [80] and adopted by Sauermann et al. [304]. Based on various publications, Sauermann

[4] This is a simplified example that especially uses resource labels instead of URIs for the sake of readability.

et al. stated [304] and later refined [303] a definition that finally reads as follows:

> Definition: **Semantic Desktop (SD)**
>
> *"A Semantic Desktop is a device in which an individual stores all her digital information such as documents, multimedia and messages. These are interpreted as Semantic Web resources, each is identified by a URI and all data is accessible and queryable as RDF graph. Ontologies allow the user to express personal mental models and form the semantic glue interconnecting information and systems, and Semantic Web protocols are used for inter-application communication. The use of Semantic Web standards allows existing web resources to be incorporated into the personal knowledge space, and does also facilitate the sharing of knowledge with others, for example within a work-group. The Semantic Desktop is an enlarged supplement to the user's memory."*
>
> L. Sauermann [303]

It is noteworthy that this definition deliberately uses a similar wording as Vannevar Bush in his famous article "As We May Think" [50]. In this article, Bush introduced *Memex*, a theoretical personal computing device providing an inspiring vision to many IM-related approaches and the Semantic Desktop (SD) in particular.

Whether they explicitly call themselves an SD or not, *SD-like systems* typically are "aware" of (in the sense of crawling, registering, indexing) the user's different information items like files, mails, bookmarks, calendar events, etc. These items form the basic data foundation, which is complemented by metadata ranging from simple usage statistics (like file modification dates, mail sending/reception dates, simultaneous use of multiple items, etc.) to captured user activities (like keyboard inputs, mouse clicks, active application windows, etc.) (1). The latter are further processed (2) to infer higher-level tasks from individual low-level actions. This is known as the mining (synonyms: detection, prediction or elicitation) of tasks – or of *contexts* to be more general (3).

As hinted at by the numbers in brackets, these steps are actually the first three of the cSpaces interaction cycle (see Sec. 1.4). This should not come as a surprise since cSpaces itself is an SD approach. To accomplish these steps, the SD is typically connected to traditional desktop applications like browsing or communication tools or typical office applications like text or spreadsheet editors. This is either done by injecting plug-ins or monitoring the applications (see Sec. 6.1). There is also the possibility to create "native" SD applications that are "aware" of the SD and directly communicate with it making plug-in injection or monitoring obsolete.

In summary, one can say that the first half of the cSpaces interaction cycle (lower half of Figure 1.1) covers typical aspects of an SD justifying its use as an ecosystem. Other researchers came to the same conclusion and saw it as "natural to leverage on work in the Semantic Desktop field in order to simplify the development of context-aware applications" since "Semantic Desktop systems already capture a relevant portion of the user's context by interfacing with the conventional applications" [44]. Several projects and studies have shown that the SD can be successfully applied to support IM and KW (e.g. [308, 303, 121, 96, 241, 87]). Even its superiority compared to traditional systems in this regard has been demonstrated

[121]. However, SD approaches are still not very widespread – or not yet brought "into the public eye" as stated by Dragan and Decker [96]. Some attempts to bring the SD to a broader user base like *Nepomuk-KDE* [268] even failed [154]. These topics are also addressed in the related work and overview chapter on the Semantic Desktop (4.1).

The aforementioned (typically initial) capture or crawling of the user's information items can be seen as being part of the cSpaces interaction cycle's prerequisite step 0 (*Data Storage and especially Knowledge Graphs*). Knowledge is represented in the SD using the *Personal Information Model (PIMO)* which is a personal knowledge graph and the next section's topic. After introducing PIMO, the final section of this chapter presents CoMem as a recent example of an SD. Several screenshots and examples will be provided to illustrate how an SD may look like in practice. This also serves as a practical complement to the more theoretical character of the current section.

2.3.2 PIMO – Personal Knowledge Graphs as a Semantic Desktop Cornerstone

The Personal Information Model (PIMO). As explained in the previous section, the Semantic Desktop uses triple statements to express knowledge. These triples are stored in a personal knowledge graph, the *Personal Information Model*, or *PIMO* for short. The goal is to capture a person's mental model as accurate as possible in a machine understandable way. A detailed definition of PIMO reads as follows:

> Definition: **Personal Information Model (PIMO)**
>
> *"A PIMO is a Personal Information Model of one person. It is a formal representation of parts of the users Mental Model. Each concept in the Mental Model can be represented using a Thing or a subclass of this class in RDF. Native Resources found in the Personal Knowledge Workspace can be categorized, then they are occurrences of a Thing.*
> *The vision is that a Personal Information Model reflects and captures a user's personal knowledge, e.g., about people and their roles, about organizations, processes, things, and so forth, by providing the vocabulary (concepts and their relationships) for required expressing it as well as concrete instances. In other words, the domain of a PIMO is meant to be 'all things and native resources that are in the attention of the user when doing knowledge work'."*
>
> L. Sauermann et al. [306]

The definition mentions *native resources*, which are the information items (i.e. files, mails, bookmarks, etc.) mentioned earlier in the chapter. Whenever a statement about such an item should be made, e.g. a user tags a document with a certain topic, the item is "rebirthed", which means it gets a concrete representation – a node with a URI – in the user's PIMO. By definition, a PIMO belongs to exactly one person. However, users have the possibility to share (parts of) their PIMO with others (colleagues, friends, etc.). These shared parts of multiple PIMOs then result in a *Group Information Model (GIMO)* [241] that forms the basis for an organizational memory. Studies showed that such PIMOs do indeed reflect the subjective view of users and can be used to increase productivity [305].

A mini example of a PIMO is illustrated in Figure 2.1: In 2013, Peter Stainer was on a vacation in Costa Rica, where he took a photo. The example thus involves three "entities" of reality (see bottom of the figure): Peter, the vacation and the photo on Peter's computer. All three of them are represented in Peter's PIMO as "things" (mid section): "Peter Stainer", "Costa Rica 2013" and the image file "IMG_4120". In this case, each thing belongs to a different class (top part), which is expressed by the following triple statements:

- *Peter Stainer –is a→ Person*,
- *Costa Rica 2013 –is a→ Vacation* and
- *IMG_4120 –is an→ Image*.

Dependent on the used ontology, one could imagine further statements like:

- *Peter Stainer –was on vacation→ Costa Rica 2013*,
- *Peter Stainer –is creator of→ IMG_4120* or
- *IMG_4120 –is related to→ Costa Rica 2013*.

While this example focuses on private life, Figure 3.3 shows the scenario of a business meeting.

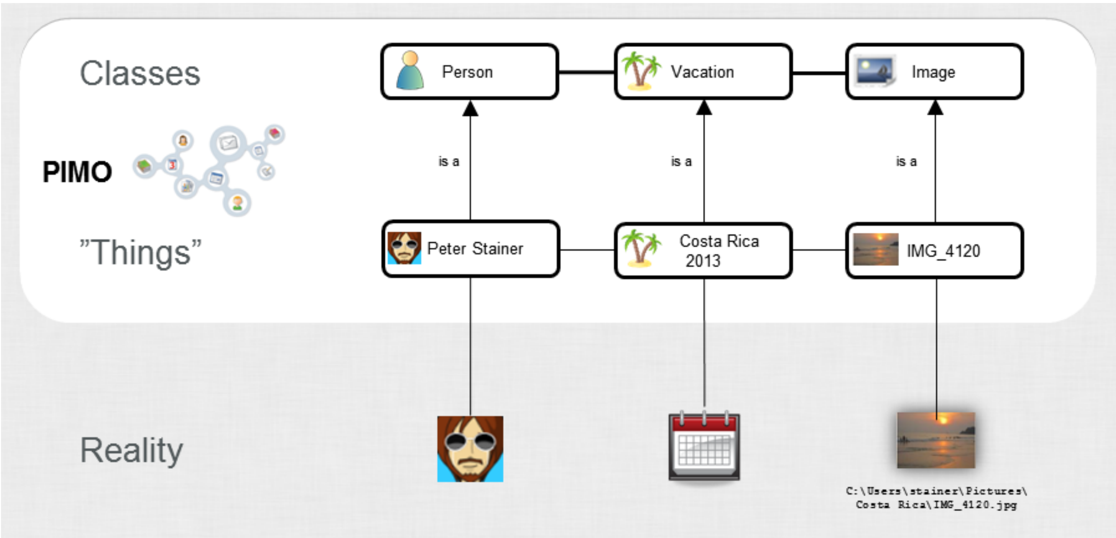


Figure 2.1: PIMO in a Nutshell.
(Note: This figure was created by Heiko Maus during the *ForgetIT* project (2013–2016) [120] and has been published multiple times since then, e.g. [240, 242].)

PIMO Ontologies. To express PIMOs, cSpaces uses a general PIM ontology, which basically is an exported excerpt [Sw110] from CoMem (see next section). Thus, it has proven itself in more than a decade of productive use in daily work. It emerged from the *PIMO ontology* [305–307] envisioned and realized in the EPOS [110] and NEPOMUK [267] projects. After the export, some cSpaces-related extensions were made by the author, especially with respect to the modeling of contexts. More details on the used ontology can be found in the Section A.1.

The next section concludes this chapter by presenting CoMem, DFKI SDS' most recent Semantic Desktop prototype.

2.3.3 CoMem – Current Semantic Desktop Prototype by DFKI SDS

The most recent Semantic Desktop prototype by DFKI SDS is called *CoMem*, which is short for *Corporate Memory* or *Group-wide Corporate Memory Ecosystem* [87] to be most specific. It is part of DFKI SDS' long-term vision of realizing an *information butler* [85, 86] serving knowledge workers. CoMem's implementation started in 2011 by Sven Schwarz et al. recapitulating experiences from the NEPOMUK [267] and adjacent projects (see Sec. 4.1) and transferring them to a new software foundation, i.e. a re-implementation from the ground up. While users' desktops were still the primary target, there was, however, a strong shift towards the intranet. In a nutshell, the core intention was to ensure that users' PIMOs as well as the GIMO – in other words: the organizational or *corporate* memory – would still be available, even if computers of individual users were currently turned off. More technical details as well as consequences of this decision are addressed in Section 5.2 when discussing how cSpaces differs from CoMem and what they have in common. The author of this thesis has contributed to CoMem as a researcher, developer, architect and consultant since 2014.

With CoMem, SDS follows a hybrid approach. On the one hand, it is a large software asset for research: In projects funded, for example, by the German *Bundesministerium für Bildung und Forschung [Federal Ministry of Education and Research] (BMBF)*, the *Deutsche Forschungsgemeinschaft [German Research Foundation] (DFG)* or the European Union (EU), research prototypes like the ones presented in this thesis can be developed to steadily move forward the state of the art. While such projects typically span two to three years, CoMem enables an additional long-term research lane: Following the *eat-your-own-dog-food* credo, DFKI SDS staff members have been productively using the system for more than a decade now yielding interesting results that can hardly be achieved in short-lived studies or laboratory experiments. But the exchange is twofold: Proof-of-concept implementations proven to have a benefit are typically incorporated into future versions of CoMem – for PhD students this can be a cycle of give and take. Incorporating new or enhanced components into CoMem is sometimes accompanied by industry partners that invest in maturing such components after their benefits have been demonstrated in the aforementioned research studies or in industry-commissioned feasibility studies or exploratory projects. This way, "CoMem researchers" are able to perform foundational research motivated by or grounded in real-world problems. Last not least, industry partners often provide valuable data from their daily work (real-world scenarios).

CoMem Architecture. A technical overview of CoMem is depicted in Figure 2.2. CoMem, in general, is a meta-model interconnecting an arbitrary amount of (legacy) systems (left) and providing knowledge services on top (right). Its architecture (blue box in the middle) consists of five layers. The first three of them (left), *connection*, *semantics* and *storage layer*, involve a "semanticfication" process (see Sec. 5.3) and hold the actual PIMOs and the GIMO as well as additional metadata, for example index data, access rights, etc. The *knowledge service layer* (middle) is the core and back-end of the system. IM and KW support measures are calculated here and finally presented to various front-end applications using the *interface layer* (right).

The remainder of this section presents various components of CoMem, so readers may get a better impression of how a Semantic Desktop may look like in practice. This serves as a complement to the more theoretical statements so far. The selection of what to show was,

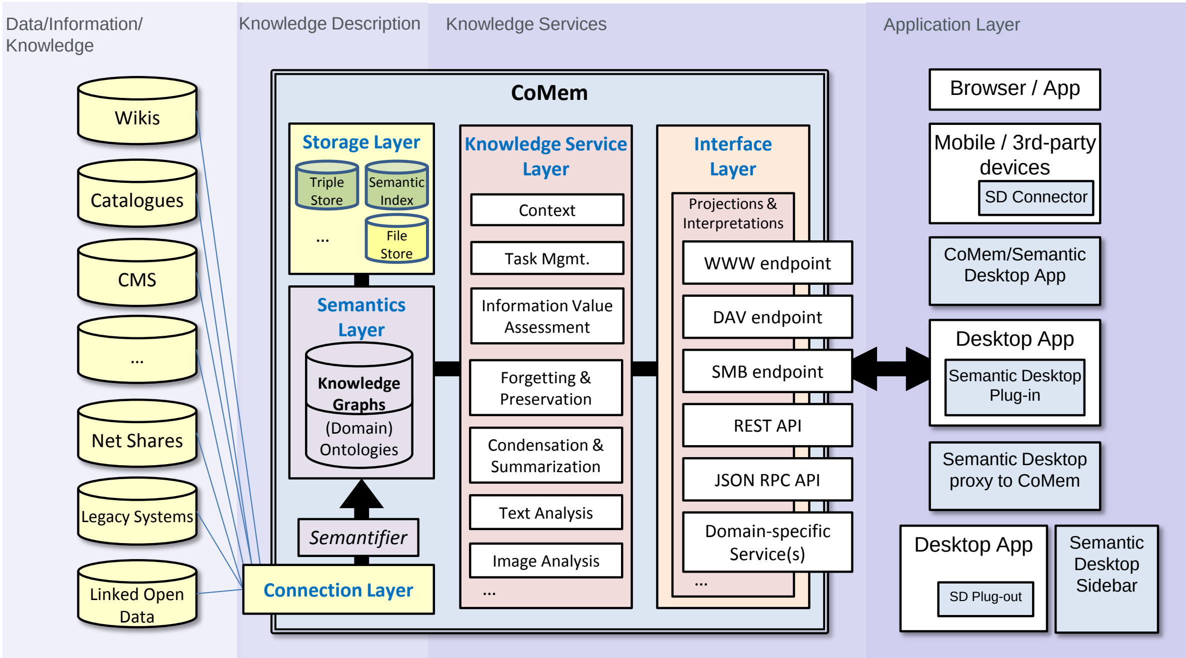


Figure 2.2: Architecture of CoMem.
Acronyms: *API: Application Programming Interface / CMS: Content Management System / DAV: (Web) Distributed Authoring and Versioning [134, 99] / JSON: JavaScript Object Notation [72, 45] / REST: Representational State Transfer [118] / RPC: Remote Procedure Call [266] / SD: Semantic Desktop [80, 304] / SMB: Server Message Block (Protocol) [161, 254] / WWW: World Wide Web [31]*. Abbreviations: *Mgmt.: Management*.
(Note: This figure was originally created by the author in 2016 for general use in the SDS research department. Multiple variants of it have thus been published since then, e.g. [179, 240, 244].)

however, not arbitrary: presented apps and features are referred to in later parts of the thesis in the context of related experiments.

Preliminary Remarks on CoMem Impressions. Before going into details, there are some preliminary remarks to be made:

- As its written-out name "Corporate Memory" suggests, CoMem is a comprehensive organizational memory approach with the Semantic Desktop as a major part connecting it to users' desktops. It also covers corporate intranet and internet resources and holds the GIMO, i.e. multiple users' PIMOs.

- This transition from *pure Semantic Desktop* to *Organizational Memory system involving the Semantic Desktop* is also reflected by a name change: For several years, CoMem went by the name of "PIMO", which was short for "PIMO server" [241, Fig. 2]. Readers looking for further papers and material should thus be aware of this renaming that took place in 2017. Several works from 2011 to 2017 mentioning PIMO are also about CoMem without using the term.

- Since this is a background and foundations chapter, **contributions by the author are deliberately not mentioned here**. They are addressed in later chapters allowing direct comparisons for cases in which features have been modified or enhanced. An overview showing which parts of the thesis mention which contributions is at the end of the impression section.

- Screenshots presented here therefore show CoMem's state as of 2015 to 2017 (dependent on the component).

- Since then, several engineers and researchers (including the author) have contributed to CoMem's development, especially with CoMem-focused industry collaborations experiencing a further boost since 2017 [358, 388], for example with *enviaM* [109], a German energy provider, or *Wacom* [389], a Japanese company specializing in graphics tablets and related products[5]. Such collaborations have brought various components closer to the state of an industry-ready product – also see Section 9.1 (*Transfer into Practice*) in this regard.

- For a recent overview of CoMem see its web documentation/showcase [87]. Additionally, the website's publication section lists further works and apps not shown here.

- Although there were major improvements (with respect to usability and look-and-feel of user interfaces, scaling, new components for specific domains and tasks, etc.), the basic principles have stayed the same making the presented impressions still a reasonable overview despite the time gap.

CoMem Impressions

Integration into Browsers Allowing for Tagging and Adding Items to the PIMO. One of the major features of the Semantic Desktop is allowing users to tag (associate) information items with things of their PIMO. Figures 2.4, 2.3 and 2.5 show this feature in three different applications: *FireTag* [Sw119] is a plug-in allowing for tagging websites in *Mozilla Firefox* [Sw87] and mails in *Mozilla Thunderbird* [Sw88]. Their counterpart for files in the *Windows Explorer* [Sw85] is called *SemanticFileExplorer* [313]. Each plug-in adds a sidebar to the left- or right-hand side of the respective app showing the results of an information extraction process, i.e. the system tries to identify things of the user's PIMO in the text of the currently browsed item. Matches are listed as so-called *suggested topics*. Users may confirm them individually by clicking the green plus next to each topic or all at once using the yellow one. Both makes the confirmed topics *annotations* (or *tags*) of the respective item and triggers the "rebirth" of the item in the user's PIMO. Such annotations may, for example, be used in a search process to ease re-finding the items. Search is a major component of CoMem, but since its current back-end was a contribution by the author, it is not listed here but instead in Section 5.2.3.

[5] The selection of companies to be named here was determined by two criteria: the author of the thesis was directly involved in the respective projects and the collaboration has been reported to a public audience by the industry partner (see [358] and [388]).

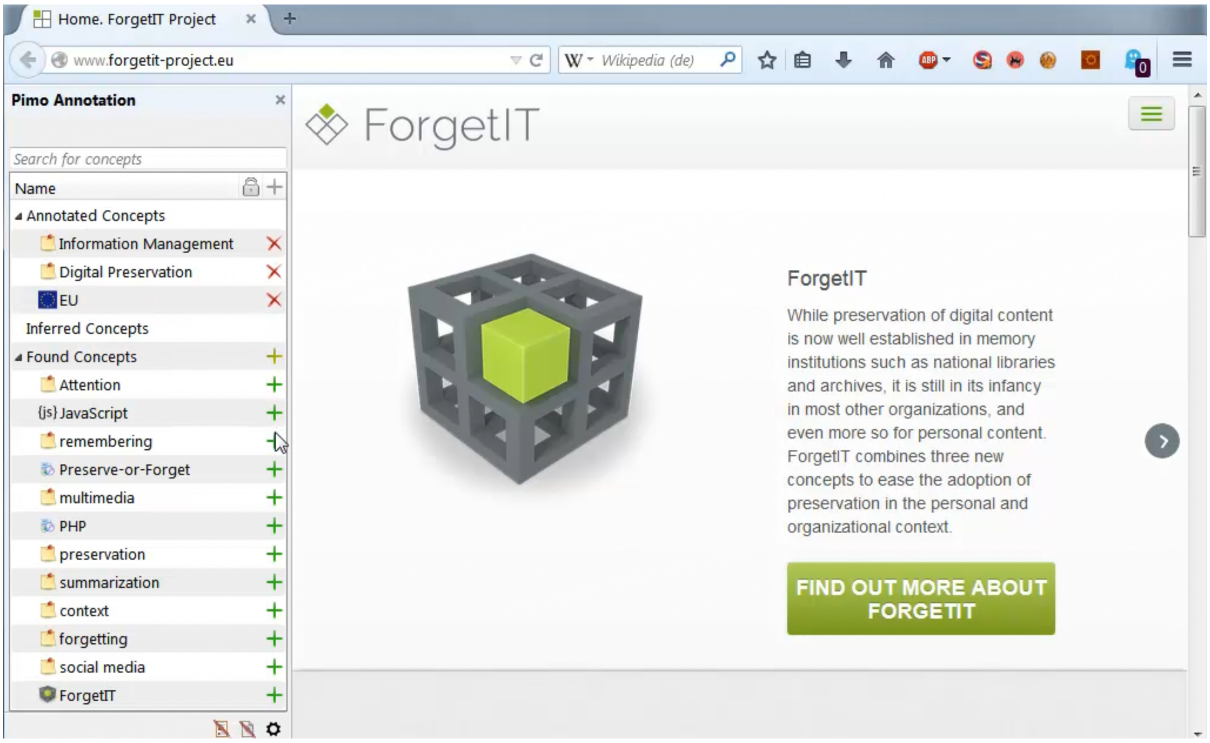


Figure 2.3: CoMem: Tagging Websites with *FireTag* [Sw119] in *Mozilla Firefox* [Sw87].
(Note: This is a screenshot by Heiko Maus published in Maus et al. [244].)

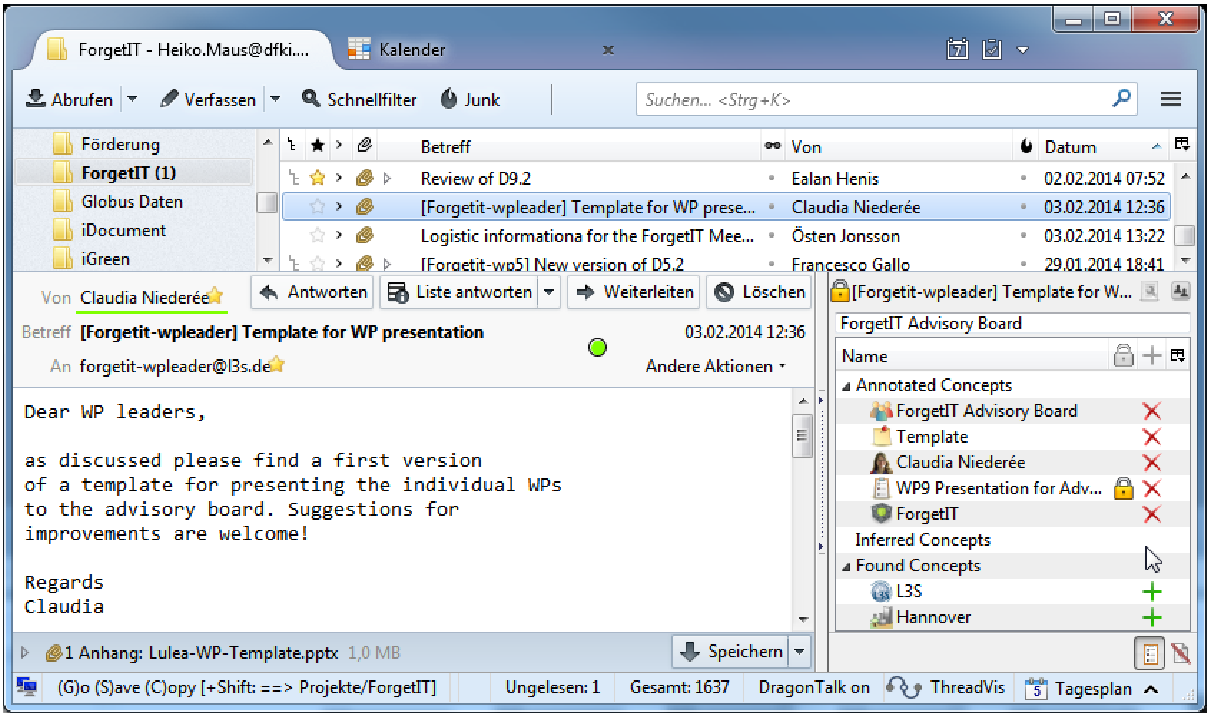


Figure 2.4: CoMem: Tagging Emails with *FireTag* [Sw119] in *Mozilla Thunderbird* [Sw88].
(Note: This is a screenshot by Heiko Maus published in Maus et al. [240].)

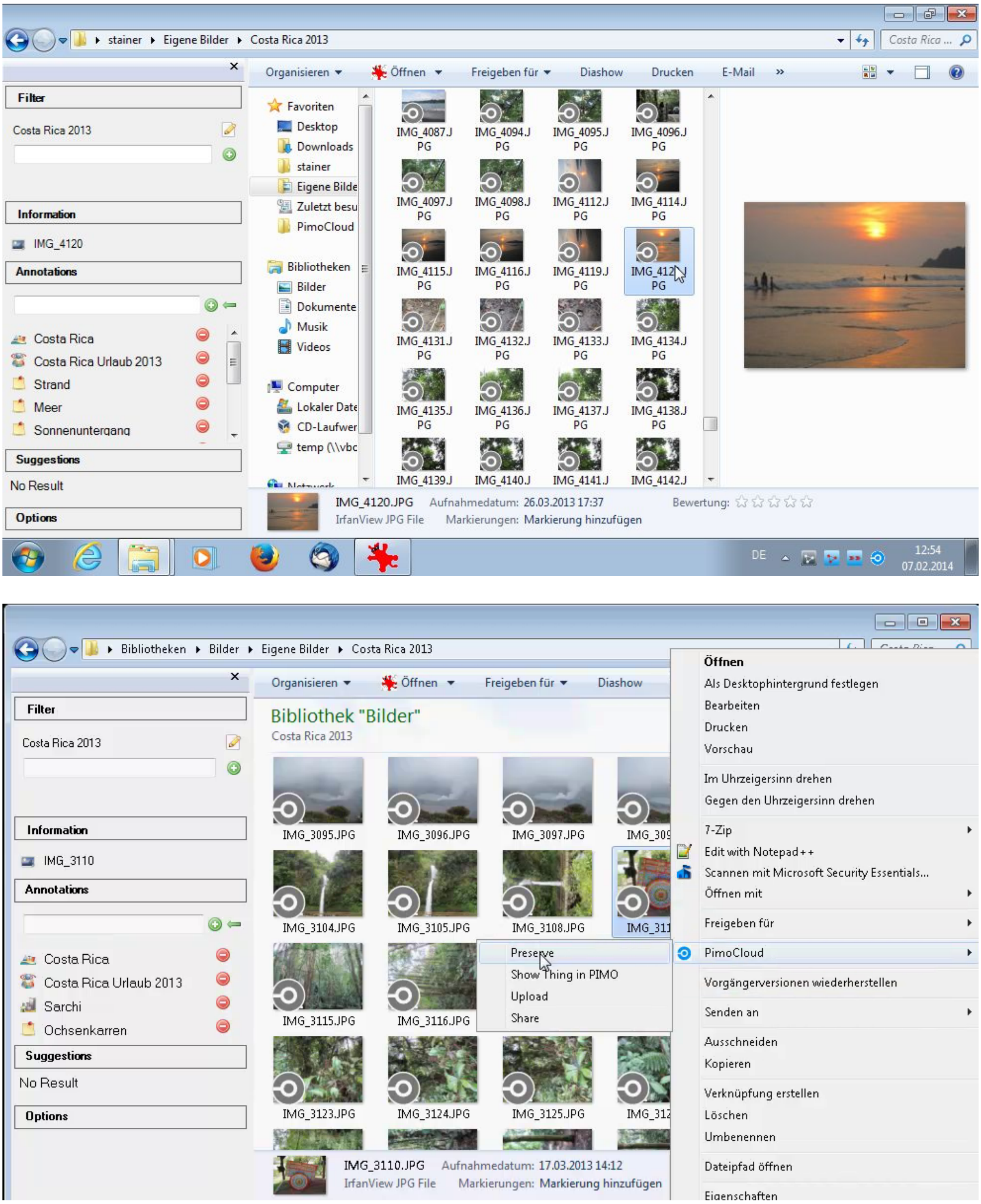


Figure 2.5: CoMem: *SemanticFileExplorer* [313] and *PimoCloud* [314].
(Note: These are screenshots by Heiko Maus published multiple times, e.g. [240, 242].)

Cloud Service. CoMem offers a cloud service for *Microsoft Windows* [Sw83] called *PimoCloud* [314]. Users may upload files either via the file explorer's context menu or by simply dropping them into a watched folder that automatically triggers the upload. Changes to the files are synchronized automatically. Automated up- and downloading became an advanced feature in combination with *Managed Forgetting* as discussed in later chapters (in particular 3.1 and 8.5.1). Utilizing *Git* [Sw43], CoMem also keeps a version history of each file allowing users to switch to arbitrary versions of the past (or back to most recent one). As illustrated in Figure 2.5 (middle part), overlay icons in the file explorer indicate different file states.

Semantic Editor. To write textual annotations like comments, notes, etc., the semantic editor *Seed* [106, 105, 104] can be used. *Seed* scans the text for known entities, either ones from the user's PIMO or from *Linked Open Data (LOD)* [32] sources like *DBpedia* [Sw34], *Freebase* [Sw75] (shut down in 2016) and *Wikidata* [Sw121]. As shown in Figure 2.6, found entities are highlighted in the text. (Note that later versions allowed for more subtle highlighting – the screenshot shows a research prototype.) All found entities are linked to the note in the user's PIMO easing later retrieval. When hovering over a highlighted entity, additional information is displayed as it is the case with PIMO in the screenshot. Users' notes are thus contextualized while typing.

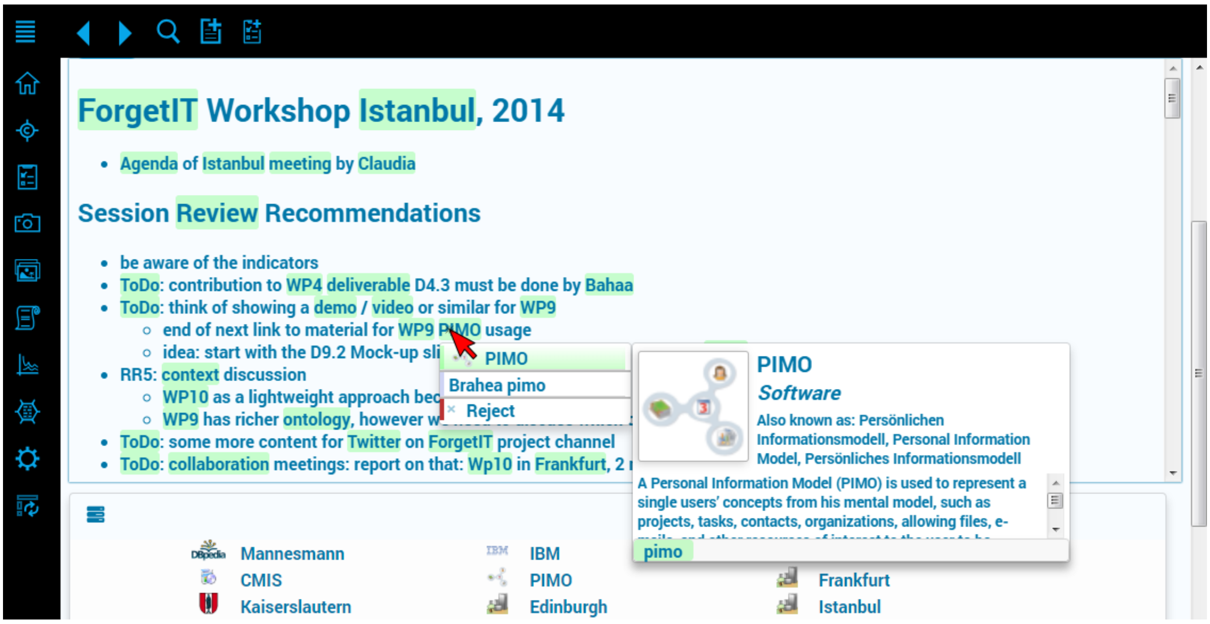


Figure 2.6: CoMem: Semantic Editor *Seed* [106, 105, 104].
(Note: This is a screenshot by the author already published in Jilek et al. [182].)

Calendar and Address Book Integration. *PimoDAV* [216] adds calendar and address book functionality to CoMem. As the name suggests, *PimoDAV* utilizes the *WebDAV* [134, 99] extensions *CalDAV* [75] and *CardDAV* [74] to provide calendars and address books via standard protocols. An example is depicted in Figure 2.7: Calendar entries (left) are automatically "rebirthed" in CoMem and further analyzed by *PimoDAV*. Finding entities of a user's PIMO like persons, organizations, locations, topics, etc. is one of the basic tasks in this regard. More complex tasks involve trying to infer person or organization names especially if they occur in a user's calendar for the first time and are not found in their PIMO, yet. *PimoDAV* subsumes these activities as *projections and interpretations*. Results can, for example, be viewed using CoMem's web graphical user interface (GUI) that is shown in the right part of the figure.

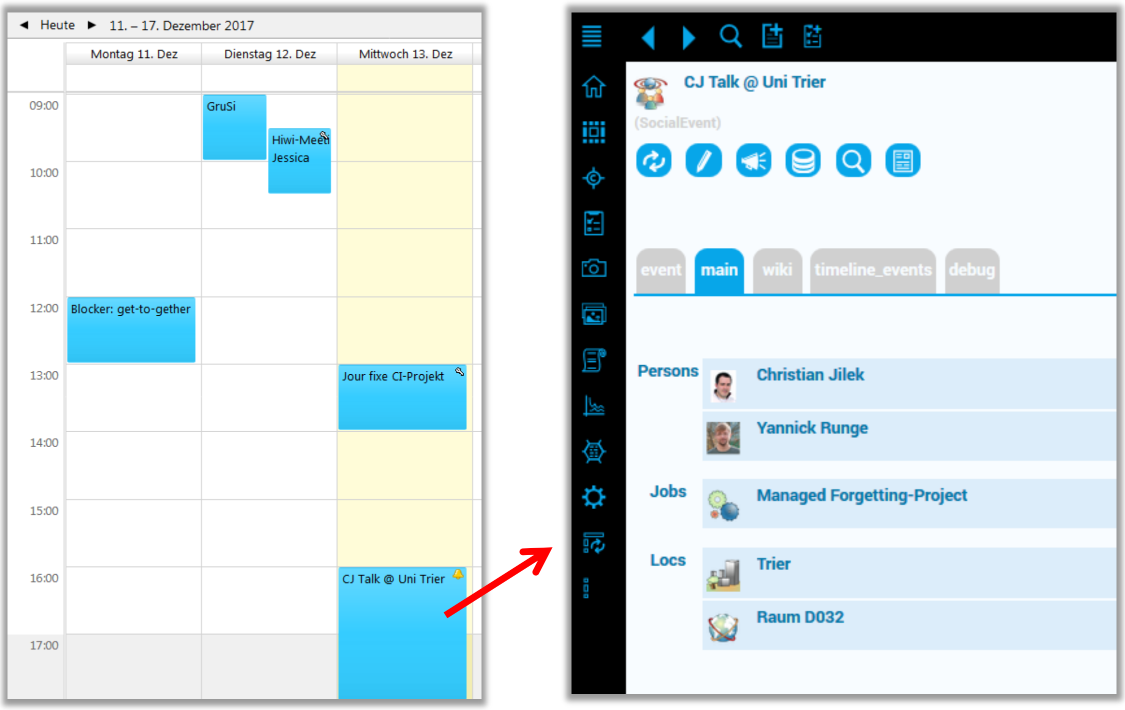
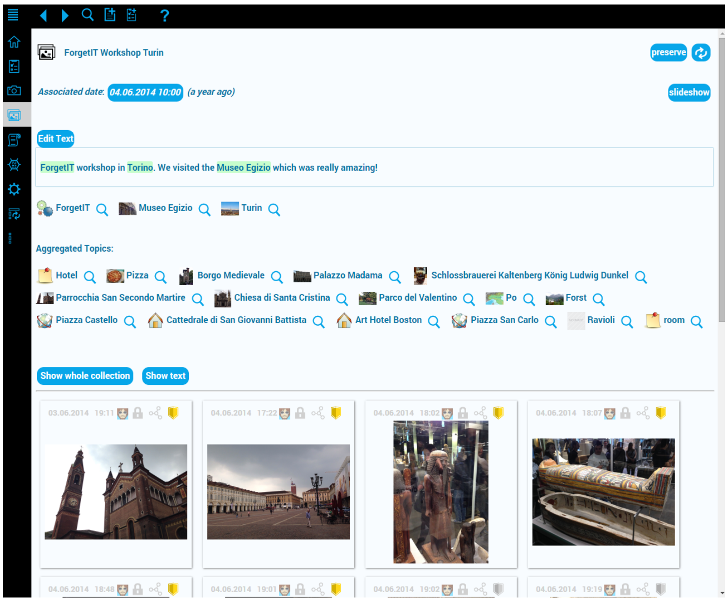


Figure 2.7: CoMem: Calender and Address Book Integration Using *PimoDAV* [216].

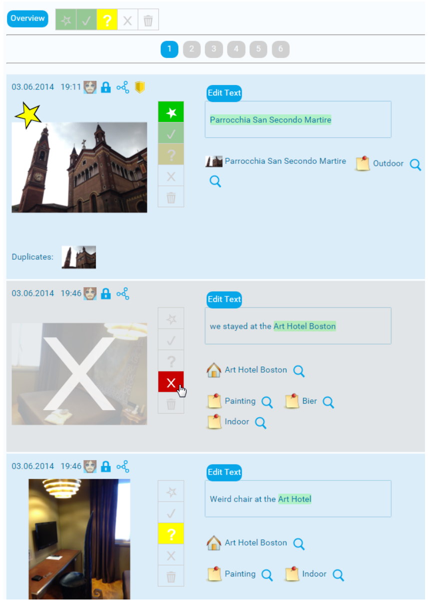
Photo Management. Users may also manage their photo collections using CoMem. Photos can be organized in collections optionally corresponding to explicitly modeled life situations like a vacation, wedding, the birth of a child, etc. The upper part of Figure 2.8 shows two browsing situations: one of a whole collection on the left and individual photos on the right. For each photo, a comment or note can be written using the previously introduced editor *Seed* (center of the figure). Entities found in the text are associated with the photo. Additionally, image analysis measures try to identify what is in the picture adding further tags like sky, people, furniture, etc. (lower part of the figure). Users can rate photos on a 5-point scale ranging from *favorite* to *can be deleted*. Photo collections can have their own comments and tags, but they also aggregate all tags given to any of their contained photos.

During the *ForgetIT* project (2013–2016) [120], CoMem's photo management especially utilized image content analysis and near-duplicate detection measures [14] by project partners of the *Centre for Research & Technology Hellas (CERTH)*[6].

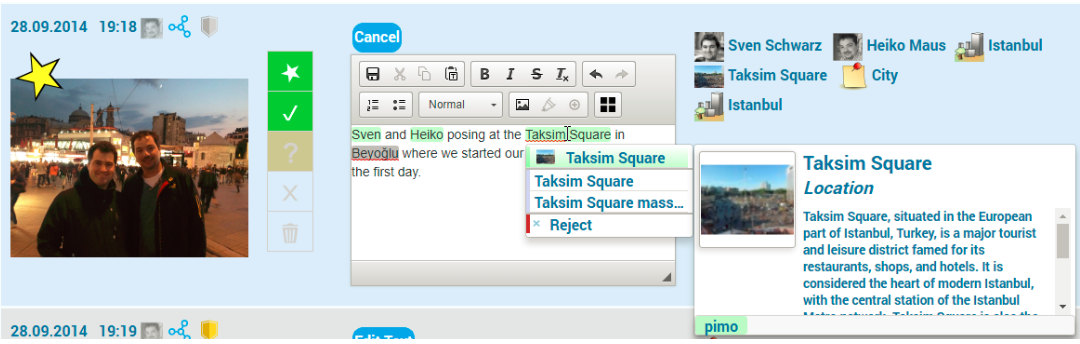
[6] https://www.certh.gr/root.en.aspx



(a) browsing a photo collection



(b) browsing individual photos



(c) writing a note or comment to a photo

(d) image analysis results: sky, people and furniture detected

Figure 2.8: CoMem: Photo Management.
(Note: These are screenshots by Heiko Maus published multiple times, e.g. [240, 245].)

Contributions of this PhD Project to CoMem. As mentioned in the preliminary remarks, there are several contributions to CoMem by the author, which were not presented in the impressions above. Since they are spread across the thesis, Table 2.1 provides an overview of relevant sections for each contribution.

Contribution	**Sections**
Revised KG Storage, Indexing, Access Rights, etc.	5.2.1, 5.2.2
CoMem Search Back-end	5.2.3
KG and especially PIMO Bootstrapping	5.3
Revised User Activity Tracking and Plug-in Architecture	6.1
Revised Ontology-based Information Extraction	6.2
Information Value Assessment in CoMem	7.2.1, 7.3
Dashboard and Sidebar Interfaces	8.2
Diary Generation from Users' PIMOs	8.4.1
Managed Forgetting in CoMem	8.5.1, 8.6.1, 8.7.1
Case-based Knowledge Work Support	9.2.1
Context-sensitive Chat Bots in CoMem	9.2.2

Table 2.1: Contributions of this PhD Project to CoMem.

3

General Approach and Vision of Self-organizing Context Spaces (cSpaces)

The last part of the previous chapter (Sec. 2.3) presented the *Semantic Desktop* as one of three cornerstones of *Self-organizing Context Spaces (cSpaces)*. This chapter continues with the other two, *Managed Forgetting* (Sec. 3.1) and *Context Spaces* (Sec. 3.2). In contrast to aforementioned concepts, they have been envisioned or significantly influenced and enhanced by the author making them part of the actual approach rather than being pre-existing background knowledge. Last, the chapter presents the overall vision and approach by explaining the idea of cSpaces in more detail (Sec. 3.3).

Note: Parts of this chapter have already been published [126, 185, 187, 373, 374].

3.1 Managed Forgetting – A Variant of Intentional Forgetting

Managed Forgetting. *Managed Forgetting* was envisioned in the EU project *ForgetIT* (2013–2016) [120] that brought together an interdisciplinary team of eleven partners from different universities, research institutes and companies across Europe. Their backgrounds covered various topics of computer science (especially PIM, information retrieval, multimedia analysis and cloud computing) as well as cognitive psychology, law and economics. The author of this thesis joined the project at the beginning of its second year in 2014, first as a diploma student [178] and student research assistant and from 2015 on as a researcher and PhD student. Apart from contributing to the further elaboration and prototypical realization of Managed Forgetting, the author later extended the idea by emphasizing more on self-(re)organization aspects especially also taking the whole life cycle (i.e. additionally early phases) of an information item or context into account.

Managed Forgetting is a variant of *intentional forgetting*, and there was a spiritual, two-phased successor project to ForgetIT that was explicitly called *Managed Forgetting* (Phase I: 2016–2020, Phase II: 2020–2023) [234]. It was part of a DFG-funded *Schwerpunktprogramm [priority program] (SPP)* on *Intentional Forgetting in Organizations* [359]. In this program, eight interdisciplinary research teams, mostly consisting of computer scientists, cognitive psychologists or ergonomists, investigated different variants and application scenarios of intentional forgetting as discussed in the respective related work and overview section (4.2).

Members of the ForgetIT consortium first published the idea of Managed Forgetting in Kanhabua et al. (2013) [198] and Niederée et al. (2015) [270]. Based on these two position papers and incorporating extensions by the author as well as experiences of the aforementioned follow-up project, this thesis defines Managed Forgetting as follows:

Definition: **Managed Forgetting**

Managed Forgetting is an approach performed by a computer system to replace the binary keep-or-delete paradigm with an escalating set of measures instead.

These measures range from

- *temporal hiding and inhibition*, to*
- *condensation, aggregation and summarization, to*
- *adaptive reorganization*, synchronization, archiving and deletion.*

The selection and application of measures, i.e. determining what to forget and what to focus on, is performed in a self-organizing and decentralized way based on observed evidences (typically gathered by user activity tracking).
Managed Forgetting relies on two variants of information value assessment inspired by human memory and cognition: Memory Buoyancy for short- and medium-term information value and the Preservation Value as a long-term counterpart.
The goal of Managed Forgetting is to complement not copy or replace human memory.

*) in extension to the initial definition

The initial definition of Managed Forgetting inherently addressed later phases of an information item's (or context's) life cycle. One idea of this PhD project was to emphasize more on self-organization aspects and especially extend the definition by adding *reorganization* to the set of measures. (*Inhibition* as the other extension is addressed below.) Support measures like reorganization can already be applied in early phases of a life cycle. As an example, think of receiving an email inviting the user to a meeting next week. Further imagine said email being a response to a previously sent inquiry for an appointment. The incoming response could automatically be assigned to the existing context already containing the first email. (Note: when speaking of "context" in this regard, actually *Context Spaces* are meant – more details follow in the next section.) In the spirit of early escalation steps like temporal hiding or inhibition to reduce cognitive load, reorganization can contribute to keeping a user's computer more tidied up by moving items to plausible contexts immediately when they emerge. Other, usually later actions could be the splitting or merging of contexts. The former could be helpful if existing contexts become too cluttered with information, while in case of the latter, too many similar contexts exist on the system, so a merge could reduce their number possibly applying condensation and summarization as well. Later sections come back to these ideas and address all mentioned measures in more detail. Note that Managed Forgetting was designed to deliberately not ask users whether certain support measures should be performed. Bringing items back into the user's view and mind by explicitly asking for permission would be against the "natural flow" of things slowly fading out of memory. However, realizing an undo function allowing users to revert unsatisfying system decisions is possible.

Memory Buoyancy and Preservation Value. The definition further mentions *Memory Buoyancy (MB)* and *Preservation Value* as two variants of *Information Value Assessment (IVA)*. Although aiming for different perspectives (short/medium- vs. long-term), both concepts incorporate similar dimensions in their calculations like user activity and user investment, gravity, social aspects, etc. They are presented more thoroughly in Chapter 7. For now, only a high-level idea of Memory Buoyancy is provided: As depicted in Figure 3.1, Memory Buoyancy follows the metaphor that items losing relevance gradually "sink away", while those that are important (again) are pushed closer to the user by their higher buoyancy [198, 270].

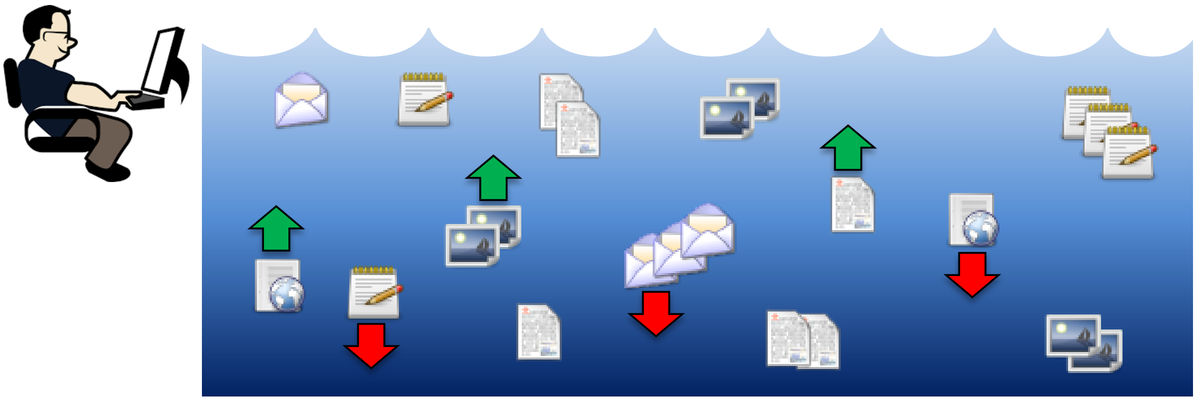


Figure 3.1: Memory Buoyancy Metaphor.
Items losing relevance gradually "sink away", while those that are important (again) are pushed closer to the user by their higher buoyancy.

Gradually fading out guided by Memory Buoyancy is one form of Managed Forgetting. The other one is triggered by sudden changes like a knowledge worker switching contexts. A concept guiding this second form is *Memory Inhibition* [222].

Memory Inhibition. In cognitive psychology, the term *Memory Inhibition* describes the temporal suppression of currently irrelevant or misleading information in order to facilitate processing of relevant information [222]. It can help to efficiently switch contexts by mentally segregating irrelevant, inhibited information from currently relevant information [365]. From a computer science perspective, especially with regard to "digital memories" of some form, mechanisms closely related to activation or highlighting the most relevant items are more frequently observed (e.g. information retrieval concepts for search [236] or spreading activation [70]) than ones trying to suppress "unwanted" information. Together with colleagues of the Managed Fogetting project [234], the author of this thesis has published a survey paper about *Memory Inhibition in Cognitive and Computer Science* [373]. Later sections, especially 9.3.3, address this topic in more detail.

Human–*Forgetful Computer* Interaction!? The following paragraph concludes this first introductory section on Managed Forgetting with a few thought-provoking remarks also sharing some insights on conversations and reactions the author experienced at workshops, conferences or invited talks during the time of this PhD project.

Since the time computers actually became *personal* computers (PCs), leading to a great number of people having one (or even multiple) for themselves, users could rely on information that had been stored on their computer, would remain there – hardware malfunction or manual deletion aside. Breaking this well-established assumption by introducing self-organizing and especially *forgetting-enabled ("forgetful") information systems (FIS)* regularly raised mixed emotions in audiences ranging from concerns and skepticism to fascination. Some considered digital forgetting unnecessary (e.g. since storage space is cheap – or was cheap at the time), others saw the dawn of a new era of computing. Anyway, over the last decade, more and more research on digital forgetting has been conducted (see Sec. 4.2 for an overview) possibly shifting the topic from its niche existence into the view of a larger audience. Rightful concerns by PC users especially with respect to usability, feeling of control and trust are addressed in later sections of the thesis. Especially (graphical) user interfaces are crucial in this regard. Consider the example of a user entering keywords into the search field of an FIS and no (or seemingly incomplete) results are shown. Users would then wonder whether they entered the "right" keywords or whether they really saved the information they are now looking for. The system could then, for example, inform the user that there is currently no search result in the "active" part of their data but something in the forgotten area providing additional information what and how things could be brought back. Well-established habits and ways of interaction possibly need to be rethought to some extent. That is why it may be worthwhile to speak of Human–*Forgetful Computer* Interaction (HFCI) instead of classic HCI [52] (or a similar term that sets them apart) to stress the possible impact of such a paradigm shift.

The Managed Forgetting introduction above already mentioned that gathering contextual information is crucial for its selection and application of support measures like temporal hiding, condensation, reorganization, etc. This leads to the third and last cornerstone of cSpaces, the eponymous *Context Spaces*, which are introduced in the following.

3.2 Context Spaces – Context as an Explicit Interaction Element

In the Managed Forgetting approach, deciding what information items to forget and on which to focus on heavily relies on *context*. The same item could be very important in one context while being completely irrelevant in another. When speaking of "context", this thesis specifically means the following:

> Definition: **Context**
>
> *"A context is understood as a 'sense-giving environment' for a (given) nucleus, i.e. a context tries to represent relevant information items and their relations describing the given situation. Such a nucleus can be an activity (e.g. writing a scientific paper), an event (a meeting) or an information item itself (a PDF document, email, etc.). Because of the dynamics of situations, a context evolves over time. The context of a large research task (later containing many documents, links, notes, etc.), for example, could spawn from a small context having only an email calling for participation as its nucleus."*
>
> H. Maus and C. Jilek in DFKI-internal documents (2020),
> first published in P. Gauselmann et al. (2023) [126]

Working *with* and *in* Context Spaces. This definition leaves great freedom to the user since every information item can become the nucleus of a context. Whatever belongs together in a person's mind, should be representable as a context (space) on their computer, e.g. a task, an event or the item itself like a file, email, website, etc. (In the more business-related use case of *Context Spaces in IT support* [see Sec. 9.2.1], the ticket of a failure incident reported by a customer or system health monitor was the nucleus of a context.) As a consequence of being in the same context, contained entities (items) should then also be displayed together, so their relatedness is clearly visible for the user. A counterexample showing a typical situation with traditional systems is given in Figure 3.2. It depicts a real-life example of the previously mentioned *project fragmentation problem* [27]. Purple highlighting shows several connections between items of the file, mail and web browser folder hierarchies and the calendar. For the sake of readability, not even all connections have been drawn. All connected items should actually be together in a context (space) and not unrelatedly spread across various hierarchy systems. Some contexts could possibly have further sub-contexts for fine-grained ordering within the context, e.g. the overall project could have sub-contexts for different meetings. But even with just an encompassing higher-level context, the general relatedness of the elements would already be explicated.

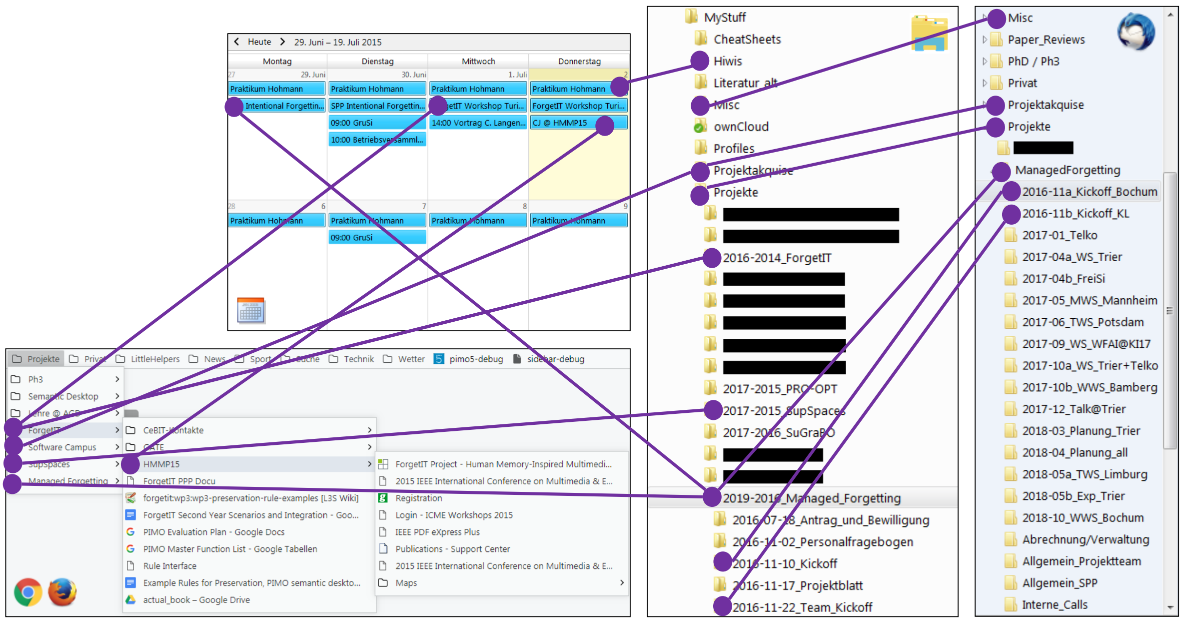


Figure 3.2: Real-life Example of the *Project Fragmentation Problem* [27].

Items being together in a single context hierarchy does not mean that users should stop using file, mail, web browsers etc. But instead of having unrelated hierarchies in each of these systems, their single context hierarchy should be available in ideally all of their applications. In this regard, remember Warren [391] and Voit et al. [387] (see Sec. 2.2) who suggested a unified file system (P3, T1) or a radically improved hierarchy browsing mechanism (V4), respectively, while staying compatible with user habits (V1). Context Spaces incorporate various item types like files, mails, bookmarks, etc. and basically allow for more complex hierarchy structures like a *Directed Acyclic Graph* (*DAG*) instead of a tree. How to enable users to actually work *with* and *in* these Context Spaces and especially how to make them available in arbitrary applications is research question RQ4 and addressed in Sections 3.3.3 and 8.2.

By using Context Spaces, users are actually aware of the *concept of context* [135] (which is a notable decision since there are also approaches going with the opposite). They can actively select a context to work *in*, which is typical for context-sensitive assistance systems. The environment then adapts itself to the selected context, for example by updating a list of recently used items accordingly [135] or by providing different suggestions as part of *Proactive Information Delivery* (*PID*) [243, 152, 343]. Beside working *in* contexts, Context Spaces also offer to work *with* contexts: contexts are also seen as a "tangible object" similar to a folder as mentioned above when discussed the unified folder hierarchy structure across all applications.

The basic motivation behind Context Spaces is the following:

> Motivation behind **Context Spaces**
>
> *Context Spaces pick up on the intuition that every user activity is performed in a certain context. Thus, every used information item is related to one or more contexts.*
>
> - *What if a support system knew about all these contexts, i.e. what they represent and what is associated with them?*
> - *What if users could actually work* with *and* in *these contexts to subtly and easily make more of their personal mental model explicit for the machine?*

The first "what-if" item hints at **advanced support measures that may become possible by exploiting the amount of contextual meta-information high above the usual** gained with this approach. Figure 3.3 illustrates this with the example of a business meeting. (The figure is at the same time a more business-related example of a user's PIMO.)

Figure 3.3: Folder Representation of a Context Space that Has a Meeting Event as a Nucleus.

The example shows a folder called *Consulting XY*. This folder, however, is not a "normal" folder and unrelated to anything else on the user's computer as it is usually the case (see above). Actually, the folder is a Context Space whose nucleus is a meeting given by the calendar entry *Meeting Mr. Smith (XY Ltd.) in Munich*. From other parts of the user's PIMO, the system can infer more about the mentioned entities of *Mr. Smith* and *Munich*. The latter *is a* city, while the former, beside being *an attendee* of the meeting, *is a* person who also *is Chief Executive Officer of XY Ltd.*, which *is an* organization whose *headquarter is located in* Munich. In contrast to many traditional systems, a lot more information about past activities performed in that particular "folder" is available, here especially the reading of a file called `report.pdf`. Now assume a point in time several months after the meeting took place. There has not been a follow-up conversation or project. Additionally, topics discussed at the meeting have not been "stimulated" for some time thus starting to lose relevance. A support system based on Context Spaces could now perform measures aiming at reducing the user's cognitive load by decluttering their personal information sphere. For example, it could move the folder to an archive "knowing" that there have not been any follow-up activities and its topics slowly becoming irrelevant, etc. This is again in contrast to comparable traditional systems that do not "know" more about the "folder" in question. More details on this matter are discussed in the next section about *Self-organizing* Context Spaces.

The second "what-if" item hints at **automatically capturing (more of) a user's mental model as well as users manually modeling their personal information sphere**. Using Semantic Desktop technology like application plug-ins (see Sec. 2.3), user activities can be tracked. Indulska and Sutton [174] speak of *virtual sensors* in this regard: software that monitors and reports (user) activities like actual hardware sensors. However, a certain sensor and interpretation gap may presumably still remain – at least as long as machines cannot read the full contents of a human brain. This is addressed more thoroughly in Schwarz [343], for example. As a consequence, a two-fold strategy is advisable: beside automatic capturing, users should be encouraged and incentivized (e.g. see Łazaruk et al. [218]) to make more of their personal mental model explicit for the machine – more on this matter follows in the next section. Again, features of the Semantic Desktop can be helpful, for example its tagging capabilities. In related studies by Sauermann and Heim [308] and Sauermann [303], users mostly used basic relations like *has-part* or *has-topic* (and their inverse counterparts *is-part-of* and *is-topic-of*, respectively) to model their personal information sphere although more complex relations were supported by the system. Accordingly, participants repeatedly stated that they "[did] not want to model the whole world" [308]. The exact nature of the relation was not needed to make navigation work, which was in line with another finding stating that users tend to only model when necessary and needed later. Context Spaces adopts these findings: By allowing users to explicitly work *in* a certain context, basic relations similar to the ones mentioned above can be inferred automatically. For example, by creating or dropping a file into a certain context space, relations like *is-related-to*, or more specifically *context-is-about* or *context-contains*, are stated without users tagging anything. (In regular Semantic Desktop scenarios, this would typically have involved two steps: creating/dropping the file and tagging it afterwards.)

Statements above already hinted at building Context Spaces on a Semantic Desktop ecosystem. The next section additionally takes Managed Forgetting into account to discuss support measures that can be enabled by combining all three of these concepts.

3.3 cSpaces – Combining Semantic Desktop, Context Spaces and Managed Forgetting

Having introduced all three cornerstones, the high-level idea of *Self-organizing Context Spaces (cSpaces)* can now be stated as follows:

High-level idea of **Self-organizing Context Spaces (cSpaces)**

Self-organizing Context Spaces, or cSpaces for short, are an approach to support information management and knowledge work by combining three fundamental ideas:

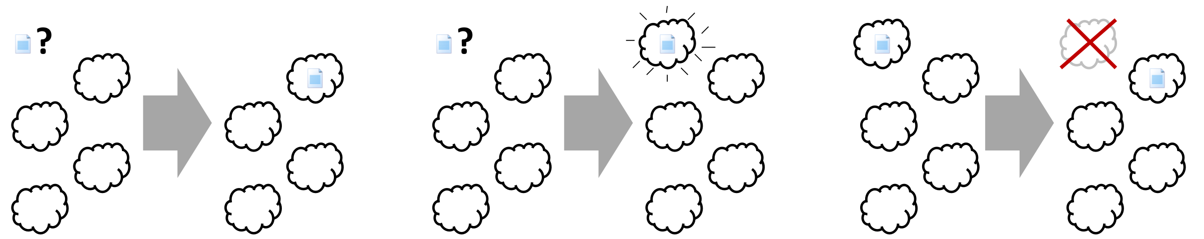
- *Have context as an explicit interaction element:* ***Context Spaces*** *allow knowledge workers to work* with *(i.e. a "tangible" object similar to a folder) and* in *contexts (i.e. immersion). These Context Spaces should be available in as many applications as possible fulfilling the vision of a single unified hierarchy structure (e.g. a tree or a DAG) replacing separate file, mail, web browser and similar folder structures. Additionally, they provide necessary contextual information to select and perform appropriate support measures.*
- *Apply measures of* ***Managed Forgetting*** *like temporal hiding, condensation, reorganization, etc. to support knowledge workers with new kinds of services and higher degrees of automation not available in traditional systems. Exploiting the high amount of contextual (meta-)information makes this possible. As mentioned in this thesis' motivation, Managed Forgetting is seen as an integral component for* unhindered *self-organization. Capabilities like condensation, summarization and ultimately deletion may significantly help with regard to long-term scalability.*
- *Use the* ***Semantic Desktop*** *as an ecosystem to capture and represent a user's personal information model (PIMO) as well as contextual meta-information. The latter is either gathered automatically by means of user activity tracking or manually by users working with Semantic Desktop features like tagging or dragging items to a certain Context Space. An enhanced PIMO explicitly contains all of a user's Context Spaces.*

How the Semantic Desktop can serve as an ecosystem for Context Spaces and Managed Forgetting has already been hinted at in previous sections, and it was briefly summarized in the statement above. The remainder of the section therefore first focuses on the combination of Managed Forgetting and Context Spaces and finally on combining all three of them. Ideas are discussed from a high-level perspective without implementation details, since they follow step by step in later sections – in this regard, remember the cSpaces interaction cycle with its eight different steps and aspects, respectively (see Sec. 1.4).

3.3.1 Self-organization

Self-(re)organization Measures. As introduced in the beginning of this chapter (Sec. 3.1), Managed Forgetting comprises several support measures like temporal hiding, condensation, reorganization, etc. Assuming users working *with* and *in* Context Spaces, Figure 3.4 shows a selection of examples of how support measures may look like. (Note: This also addresses the research question RQ2.)

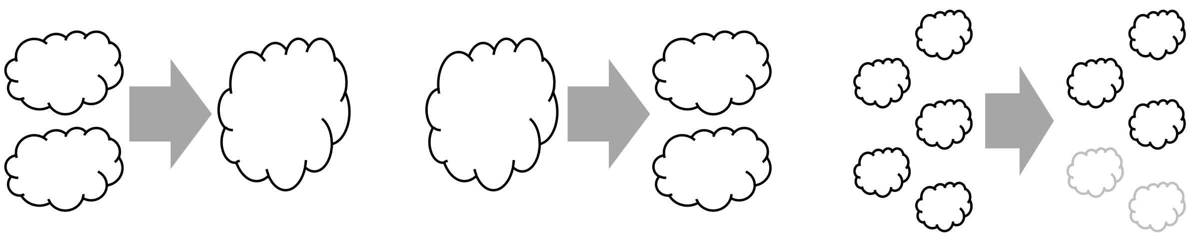
- **context assignment**: In the first example, a new information item emerges or gets into the system. cSpaces assigns it to one (or possibly more) plausible existing contexts (Fig. 3.4a).
- **context spawning**: If no matching context can be found, a new one is spawned and the item is added to it (Fig. 3.4b).
- **context retraction**: After spawning a new context, the system continues tracking user activities. A past decision can turn out to be false. Think of a case in which a company works with two different car manufacturers: tracked user evidence snippets indicate that the user is definitely in a automotive context. It is unclear, however, to which of the two customers, or maybe even a third new one, the current activities are related. After observing a few more evidences, the situation gets more clear. A newly spawned context is thus retracted and its element (or elements if more was added in the meantime) is assigned to the other context that now seems more plausible (Fig. 3.4c). However, if the user has already confirmed a newly spawned context, retraction is no longer an option.
- **context merging**: What is possible, however, is that contexts may be merged if indicators advise to do so. This could be the case if two very similar contexts have not been used for a longer time. In order to declutter the information sphere, the system could decide to thus merge them. Merging may happen on context level (Fig. 3.4d) or within a context, i.e. intra-context level (Fig. 3.4g). In the latter case, two sub-contexts of a common parent context are merged. During a merge, less important items could be set to be hidden by default making this operation not only a reorganization but also a condensation measure.
- **context splitting**: The opposite measure is a context split. If a context gets cluttered with information items, it may become reasonable two split it up. For example, think of a project context, in which travel management (booking hotels, flights etc.) happens over and over. At some point, separating traveling aspects from the actual project content could make sense. As in the case of merging, splitting may happen on context level (Fig. 3.4e) as well as intra-context level (Fig. 3.4h).
- **other measures**: Further measures are possible as illustrated in the following examples: On context level (Fig. 3.4f), the memory buoyancy of less important contexts could fall under a certain threshold and would then be set to be (permanently) hidden by default from that moment on (until they possibly regain relevance). On intra-context level, less important items or sub-contexts could be temporarily hidden for a certain amount of time (Fig. 3.4i).



(a) assign new item to existing context

(b) spawn context and add new item

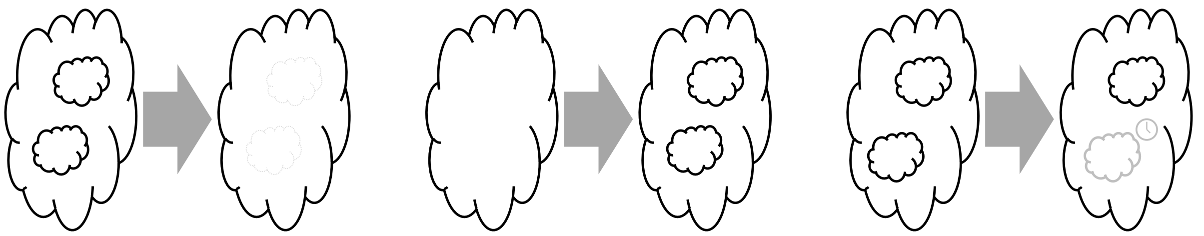
(c) retract context and assign its content to another context



(d) context merge

(e) context split

(f) other context-level measures like hiding forgotten contexts by default



(g) intra-context merge (dissolution of involved subcontexts)

(h) intra-context split (new subcontexts emerge)

(i) other intra-context measures like temporal hiding of contained items or subcontexts

Figure 3.4: Combining Context Spaces and Managed Forgetting Measures.

As mentioned earlier, these are just a few examples to illustrate the combination of Context Spaces and Managed Forgetting – further measures and combinations are possible. The mid section of Chapter 8, the main chapter on *Self-organizing Context Spaces*, is structured according to the different support measures investigated in this PhD project. They comprise condensation and summarization (Sec. 8.4), temporal hiding and reorganization as well as fading out and resurfacing (Sec. 8.5) and permanent reorganization (Sec. 8.6). Although a measure could often be mapped to more than one category, the author tried to choose the most fitting one.

Self-organization by Definition. Since *self-organization* is an eponymous concept of this thesis, some deepening and elucidating remarks shall be made here. Heylighen and Gershenson [165] defined self-organization (in computing) as follows:

> Definition: **Self-organization (in Computing)**
>
> *"A self-organizing system not only regulates or adapts its behavior, it creates its own organization. In that respect it differs fundamentally from our present systems, which are created by their designer. We define organization as structure with function. Structure means that the components of a system are arranged in a particular order. It requires both connections, that integrate the parts into a whole, and separations that differentiate subsystems, so as to avoid interference. Function means that this structure fulfills a purpose.*
> *Designers obviously create systems for a particular purpose: the function of a watch is to tell time, of a database to store data, of a spreadsheet to calculate. [...]*
> *Self-organization then means that a functional structure appears and maintains spontaneously. The control needed to achieve this must be distributed over all participating components. If it was centralized in a subsystem or module, then this module could in principle be removed and the system would lose its organization. Take the processor chip out of a computer and it becomes useless. Take any small piece of tissue out of a living brain (as commonly happens during brain surgery), and the brain will continue to function more or less like before."*
>
> Heylighen and Gershenson [165]

Note that cSpaces does not qualify as a self-organizing system according to this definition – which was not claimed. Remember the definition of Managed Forgetting (see Sec. 3.1), which stated that the selection and application of measures was performed *in a self-organizing and decentralized way* based on observed evidences. This means in other words: the content in cSpaces is maintained in a *self-organizing and decentralized way*. Having the exemplary measures described in the previous section and shown in Figure 3.4 in mind, one could see the different context spaces as the *structures* the definition speaks of. Their *function* is the representation of each of a user's contexts, which implies that they more or less constantly "reshape" (evolve) to stay in sync with the user's mental model (as far as this can be perceived). If arbitrary context spaces were removed, the system would still remain functional: selection and application of support measures would be performed based on and with the remaining ones. The decentralization aspect is relevant in this regard: not the whole content needs to be known in order to perform measures. Local areas of the whole space of context spaces, i.e. a context space itself and possibly its close neighborhood, are sufficient to select and perform support measures. In summary, it is (currently) not the cSpaces system as a whole but the way its individual context spaces are created and reorganized over time that resembles – not perfectly in detail but to a considerable extent – what is stated in the definition. A full decentralization of the decision-making component of cSpaces in combination with sharing global knowledge among autonomous entities seems feasible, which would make cSpaces further adhere to the definition.

3.3.2 Challenges of Self-organizing and especially Forgetting-enabled Information Systems

This section presents major challenges with regard to a self-(re)organizing and especially forgetting-enabled information systems (FIS). While some are discussed in detail, others are only briefly mentioned to provide an overview or guide. In particular, this is the case for those that are re-addressed in later sections.

Capturing and Storing Contextual (Meta-)Information and Privacy Issues (relates to RQ3 and RQ5)**.** Capturing a user's PIMO and especially their different contexts has been addressed in previous sections about the Semantic Desktop and Context Spaces. Chapter 5 deepens context modeling, storage and bootstrapping aspects, while Section 6.1 focuses on user activity tracking and privacy preservation, which is most relevant in that matter. A personal knowledge assistant like cSpaces captures and stores highly sensitive data, so caution should be taken right from the system's design phase.

Context-sensitive Information Value Assessment (RQ6)**.** To select the appropriate support measures to perform, Managed Forgetting relies on *Memory Buoyancy* and *Preservation Value* (see Sec. 3.1), which are forms of Information Value Assessment (IVA). IVA is addressed in Chapter 7: first in general, then different prototypes developed at DFKI and especially in this PhD project are presented.

Cautiousness. Benn et al. [26] found out that navigating folder hierarchies activates the same brain regions as real-world navigation. (The authors assume that this could also explain users' preference of navigation over search [391] mentioned in Section 2.2.) These findings were obtained with traditional systems having rather static folder hierarchies compared to a self-reorganizing system like cSpaces. "Wildly" moving items and folders around will thus very likely be unintuitive and unbeneficial. In the metaphor of a real-world map, places people knew were in the West should not suddenly appear in the East or other cardinal directions. As a consequence, cSpaces should act cautiously, which means preferring no activity over a possibly wrong activity when in doubt (e.g. multiple indicator scores are all just barely above a threshold, i.e. no stronger indication in any dimension). Additionally, undoing measures should be possible (at least as long as this is a research prototype and not a final product).

Trust (RQ8)**.** Cautiousness is just one measure for earning and preserving users' trust, not losing the feeling of control is another one. Section 8.8 especially addresses the topic of trust in highly autonomous agents based on artificial intelligence (AI) like cSpaces. Trust is particularly important with respect to searching an FIS.

FIS Search: Trust in Results and Visualizing the Forgotten (RQ7)**.** Section 3.1 mentioned searching an FIS when talking about HFCI. Remember the stated imaginary case of users wondering whether they entered wrong keywords or whether a certain item had actually been stored on their system, just because an FIS replies with zero query results. The topic of searching FIS is discussed more thoroughly in Section 8.7 presenting various prototypes developed in this thesis.

Incentivize Users' Willingness to Contribute (relates to RQ5). In the introduction of Context Spaces (see Sec. 3.2), a presumably remaining sensor and interpretation gap was mentioned. cSpaces will thus rely on users making (more of) their PIMO and Context Spaces explicit for the machine. As hinted at in Sauermann and Heim [308], users tend to only model when necessary and needed later – in other words: if they have a benefit from doing so. Thus, their contributions should be incentivized and rewarding (see Łazaruk et al. [218] for further examples). In the following, this thesis assumes an adapted and basic definition of the *return on investment* as *benefit* divided by *investment*:

$$\text{return on investment} = \frac{\text{benefit}}{\text{investment}}$$

Investment, in this case, refers to actions by the user, like entering additional information, tagging things, dragging something to a certain context, etc. Possible benefits could be easier re-finding or the system performing (sub-)tasks on its own, since enough information is available to do so, for example. Maximizing benefit and/or minimizing investment would be obvious choices for optimization. The author further argues for trying to design user interfaces in a way that investment can be decomposed into several sequential micro investments. For example, consider the case of adding a person to a meeting event in the calendar: First the event is clicked, than a tab with attendees is selected. Next, the user either has to open a list of known persons from their address book or they start entering a new name. Last, clicking on *OK* or *confirm* finally adds the person to the meeting. The system may now try to infer further facts by knowing that a certain person is an attendee of the meeting, e.g. involved companies, locations, topics, etc. But if the user stops at anytime during these steps (leading to the person not being added), the investment so far is likely lost completely. There is, of course, always a minimum sequence of actions needed to perform a certain task, which cannot be reduced any further. However, in the design of cSpaces, one could try to make even little actions like a single click or drag operation already mean a lot. In Section 3.2, such an example was given: Working in a particular context space already tags its items as being *related to* or *contained in* that context without any further user actions. Especially considering the aforementioned "multi-tasking craziness" [137], there will likely be times in which users would consider an investment worthwhile, but nevertheless just cannot perform it in that very moment due to pressure to continue their work. Offering sequences of "tiny" actions that can be interruptible at any time allows them to better regulate the amount of distraction they are willing to spend in that particular situation. In the example of creating a task, activities could be: create the task, set deadlines, add notes, web or file links, etc. The user can decide at any point whether they go another step or stop, whereas both, the system and the user, benefit from each additionally taken step. Even if the user stops early, at least some "modeling steps" still have been performed.

If possible, users should immediately have and see benefits from their actions. A mismatch of investment and resulting benefit may lead to a vicious circle of knowledge management [278], i.e. users investing (entering) less since they do not see a benefit from it, which in turn leads to less benefit since needed information is not available in the system and so on.

Evaluation. Evaluating an FIS is a particularly hard challenge mainly due to subjectiveness, missing datasets and, in several cases, the need that some time has passed, which either rules out short-term laboratory studies or forces bringing out effects artificially. Section 8.1.1 discusses these aspects in more detail. The chosen strategy, methods and tools are then presented in Section 8.1.2.

3.3.3 Further Conceptual Details

This section presents last conceptual aspects of cSpaces before detailed aspects and prototypical implementations are discussed in the main part of the thesis.

cSpaces Nucleus and Flavors. The definition in the beginning of this section (3.2) spoke of a context as a "sense-giving environment" for a given nucleus. This "environment" also has the characteristics of a collection: it contains or is associated with all of the context's items. Each context space can have a description, comment, wiki text, etc. as a basic text associated with it. When working with context spaces, users should also have the possibility to easily state that something has be to done with a particular context space, thus giving it a "task-ish" characteristic. The same is true for calendar events: it should be easy to associate a context space with an event in a person's calendar. This could be social events like a meeting, conference, etc., but also rather basic entries like the simple statement of a time span – as an example, think of a photo collection covering the first three years of a person's life. Thus, in addition to a nucleus, this thesis introduces a second characteristic, which is basically like a *type*. However, the "task-ish" or "event-ish" character of an actual collection is in the author's opinion better described as a *basically ephemeral user-attributed type* – metaphorically just called a *flavor* in the following (referring to the sensory perception of taste or smell, which may also vary over time). Figure 3.5 shows difference constellations of cSpaces nuclei and flavors. The flavor is deliberately drawn to be connected with the context space's "outer shell" in order to express that it determines its current "appearance". Since task-ish or event-ish flavors express an additional characteristic of the context space, which is still a collection as well, these context spaces are drawn with a special flavor icon, while the simple collection has an empty flavor field. The nucleus is drawn at the context space's core, although not right in the middle for readability reasons. Simple context spaces, i.e. just a collection, may (Fig. 3.5d) or may not (3.5a) have a nucleus. If a user creates a new task or calendar event, the context space would automatically get this task or event as a nucleus as shown in Figure 3.5b and 3.5c, respectively. As mentioned in the initial definition, a task could, for example, also emerge from an email: this is depicted in Figure 3.5e having the email as a nucleus and the flavor of a task. The same is possible for an event-ish context space that has another item like a document as its nucleus (3.5f). As mentioned before, flavors are not strict types. They rather vanish or maybe change over time and are meant to express basic additional characteristics of a context. However, task-ish context spaces could replace (lightweight) task management with the advantage of very low modeling effort: a task may start with a simple folder (i.e. a collection) and then evolve into more over time (e.g. entering a description, adding more items, setting a deadline, etc.).

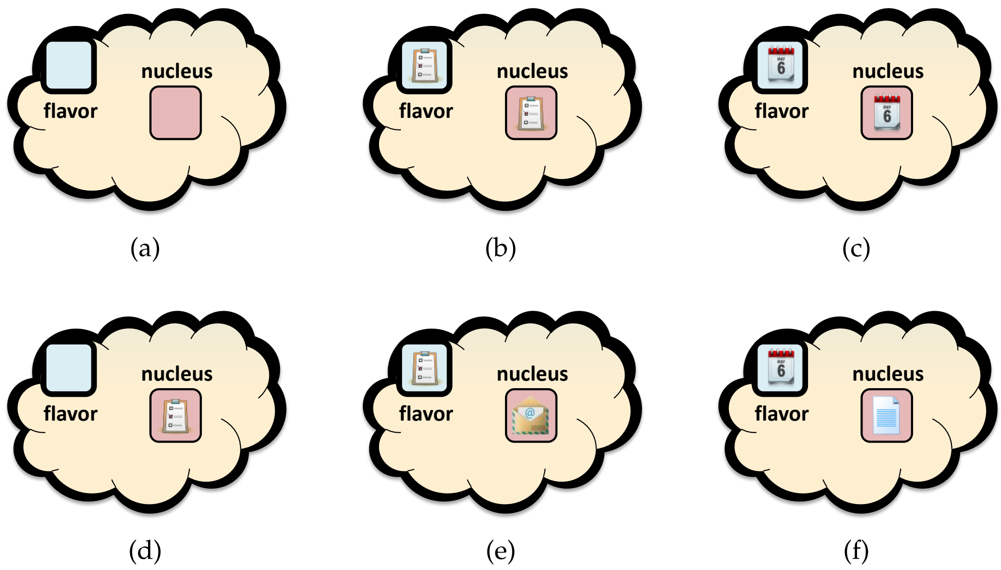


Figure 3.5: cSpaces Nucleus and Flavor Examples.

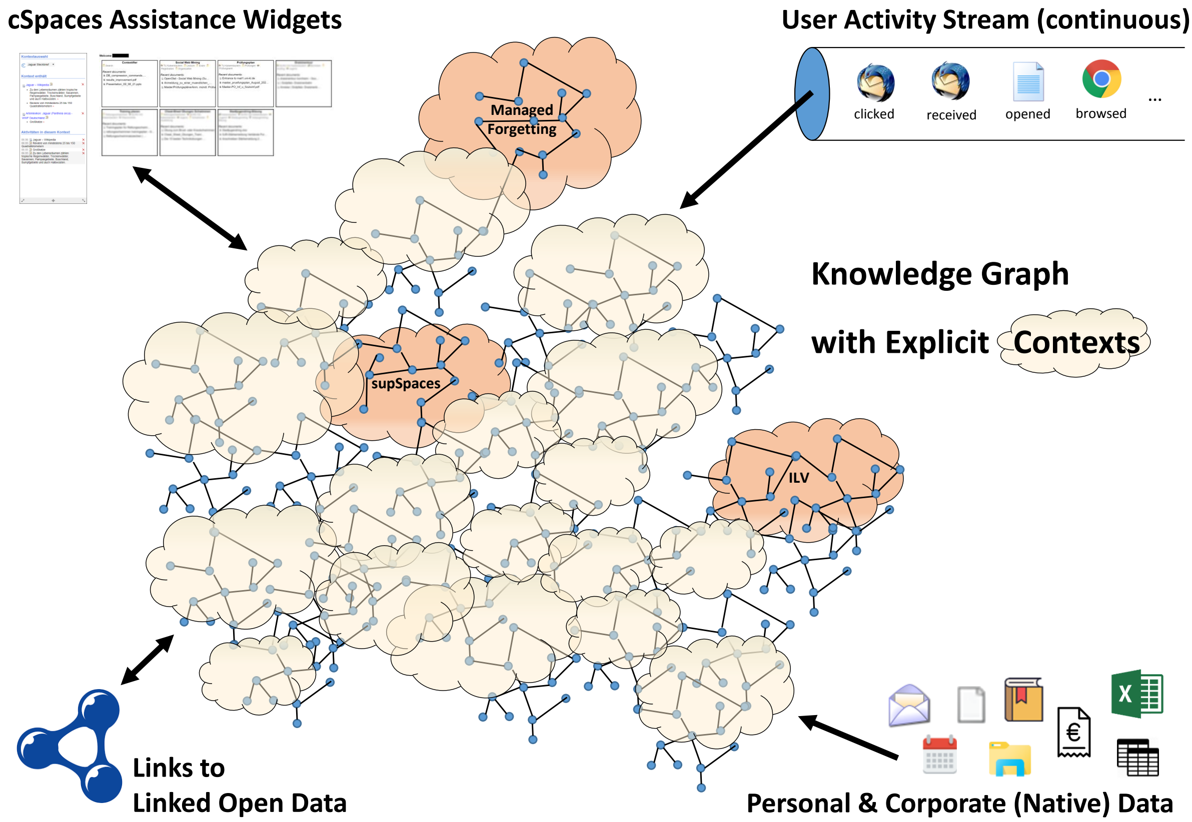


Figure 3.6: Technical Scenario of cSpaces.

Technical Scenario of cSpaces. As illustrated in Figure 3.6, the core of cSpaces is a personal knowledge graph (KG), the *Personal Information Model (PIMO)* enhanced with explicit *Context Spaces* (center of the figure). *Personal or corporate native data*, i.e. files, mails, etc., is typically available for and in cSpaces (bottom right). If semantic statements about an item shall be made, the item is "rebirthed", i.e. it is henceforth represented by a node in the KG. Note that cSpaces are a hybrid construct containing KG entities as well as native data. This is addressed in more detail in Section 5.2 about *ATIC knowledge graphs*. The KG may also be connected to *Linked Open Data* [32] sources like *DBpedia* [Sw34] or *Wikidata* [Sw121] (bottom left). cSpaces is fed with a *(continuous) stream of user activities* (top right). Last, cSpaces can be accessed using various (graphical) user interfaces (top left), which are discussed in more detail in the following.

cSpaces User Interfaces (relates to RQ4). cSpaces can be accessed using various user interfaces (UIs). In this PhD project, three UI variants were realized: a dashboard, a sidebar and transparent injections into existing systems like a file or web browser. As shown in the upper left of Figure 3.7, each of these interfaces offers a bit more familiarity to the user.

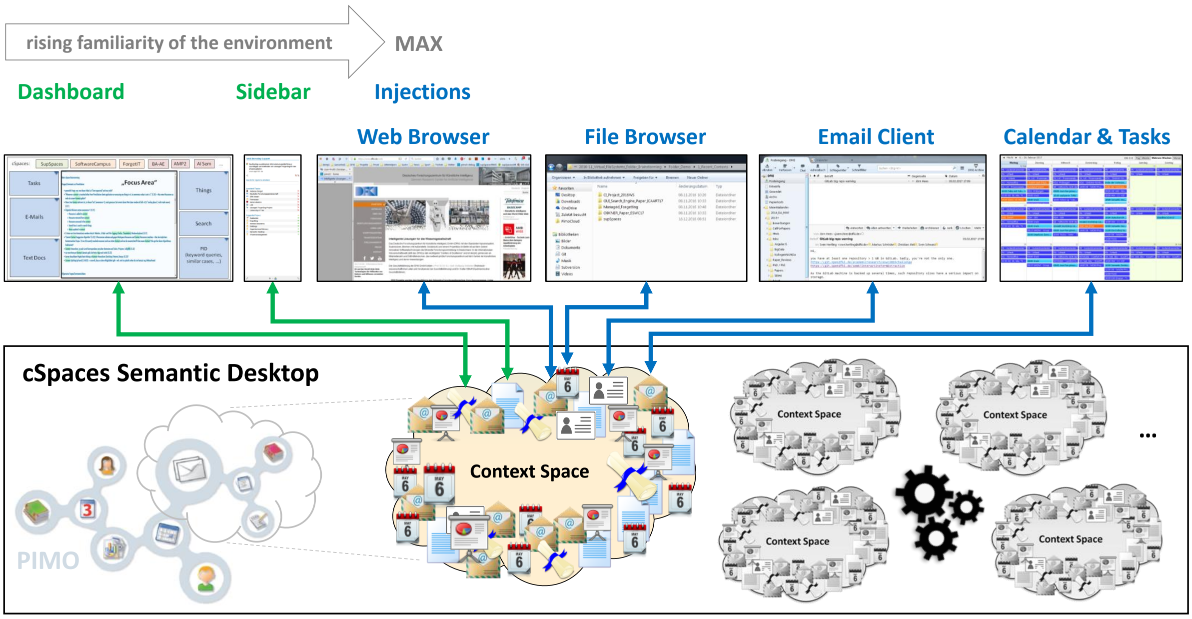


Figure 3.7: cSpaces User Interfaces.
(Note: A variant of this figure has already been published in Jilek et al. [187].)

First, there is the **dashboard**: with its potentially screen-filling size, there is lots of space to visualize information and offer interaction possibilities for search, tagging, etc. On the downside, using a dashboard is metaphorically a "hard cut" for the user: they have to leave their current application to switch to another app typically taking up the screen's whole space obscuring the user's view on all other applications.

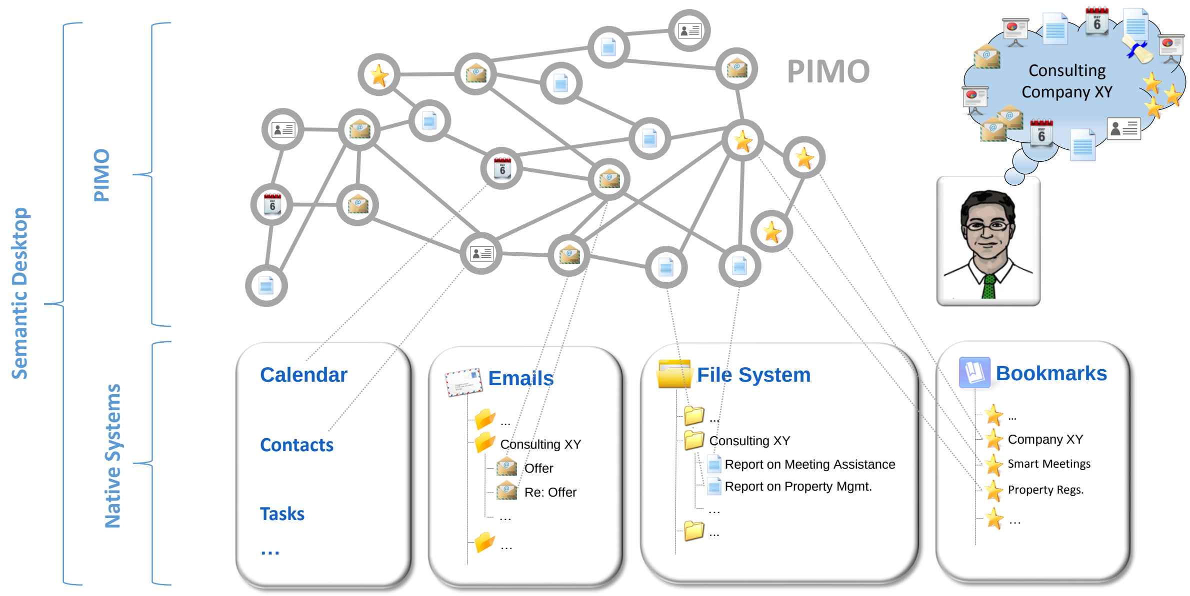
Using the **sidebar** instead of a dashboard mitigates the problem since the sidebar takes up far less space. This characteristic is at the same time its disadvantage: all widgets need to be more compact to fit into the sidebar, which typically means dropping visualization or input elements that would have been shown in the dashboard.

A dynamic solution proposed in this thesis is a so-called **sashboard**: The idea is to basically use a sidebar and whenever more space is needed, the user drags the inner side of the sidebar towards the center of the screen as if closing a (horizontal) sash window. As a consequence, the sidebar re-adjusts itself to a dashboard that is filled and pre-configured with the current cSpace's content. For example, a small quick search widget resizes into a full-fledged faceted search interface. The history list of previous actions in that context then contains up to ten elements instead of three. The same is true for a tagging widget and its suggested topics, etc.

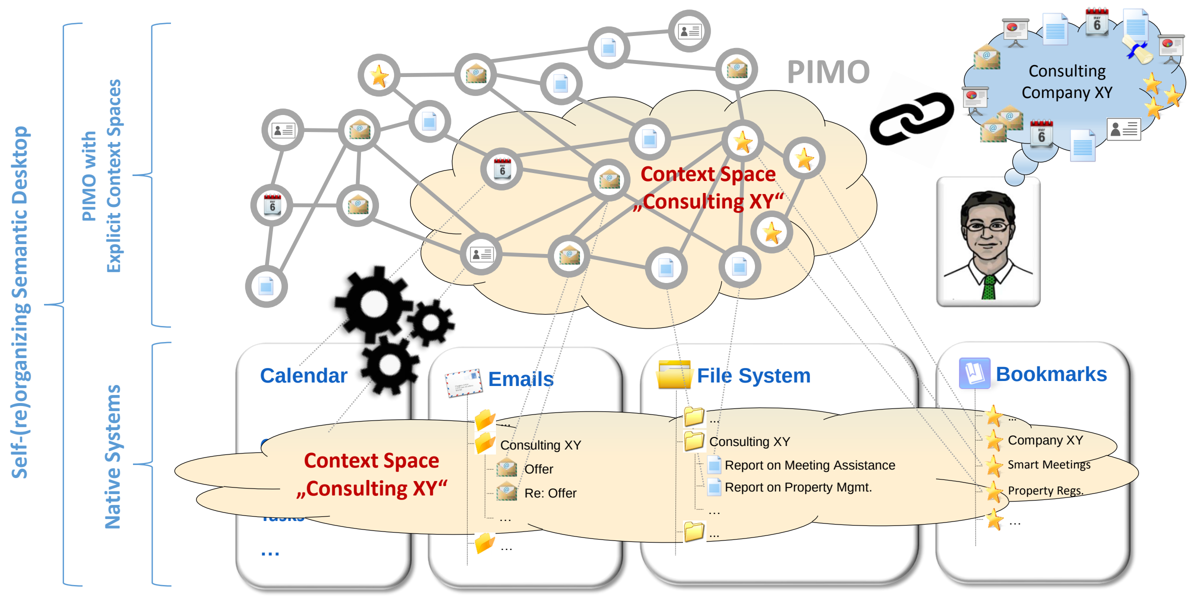
The highest familiarity with the environment is achieved using **transparent injections** of cSpaces into existing systems like the file, mail, web browser, calendar etc. In the next chapter about related work, several context-sensitive solutions are presented. Although many of them allow for capturing a user's context, especially adding items to a certain context, they typically do not take the final step of making these contexts available in (many, ideally all) existing applications, thus degrading their approaches to island solutions to some extent. In this regard, cSpaces uses standard protocols supported by all major operating systems to inject contexts on a very low level. Especially the file system is interesting in this regard since it is available in all applications and thus makes cSpaces available in all of them. Technical details follow in Section 8.2.

Traditional Semantic Desktop vs. cSpaces. Figure 3.8 shows the comparison of a traditional Semantic Desktop like *NEPOMUK* [267], *CoMem* [87] and others (top) and cSpaces (bottom). Explanations in the following refer to the bottom figure. As an addition, the PIMO explicitly represents *Context Spaces*. In the example, such a context (space) is drawn as a cloud labeled with *Consulting XY* (referring to the example from the beginning of this chapter), which ideally is in sync with the user's respective mental context (upper left). Data of native systems (files, mails, bookmarks, etc.) is also represented and contained in context spaces – see second cloud at the bottom of the figure. Note that, for the sake of readability, two clouds representing the context are drawn in the figure. Actually, there is only one context space containing both, KG entities (a subset of PIMO, top) and native data (bottom). Last, symbolized by the set of gears, support measures like condensation, summarization, reorganization etc. (see *Managed Forgetting*, Sec. 3.1) are performed to assist the user, which extends the degree of support and automation of a traditional Semantic Desktop. More technical details are addressed in Section 5.2 when discussing how cSpaces differs from CoMem and what they have in common.

This chapter presented the general approach and vision of cSpaces. More on detailed aspects follow in Chapters 5 to 8 in accordance with the cSpaces interaction cycle (see Sec. 1.4). Next, Chapter 4 presents related work, whereas especially the conclusion (Sec. 4.4) is a direct complement to the current chapter.



(a) Traditional Semantic Desktops like *NEPOMUK* [267], *CoMem* [87] and others



(b) Self-(re)organizing Semantic Desktop like cSpaces

Figure 3.8: Semantic Desktop vs. Self-(re)organizing Semantic Desktop like cSpaces.

4 Related Work

Regarding related work, this thesis follows a hybrid approach: The first part of this chapter provides general related work and overviews for the three cornerstones of cSpaces: the Semantic Desktop (Sec. 4.1), Managed Forgetting (Sec. 4.2) and Context Spaces (Sec. 4.3), whereas the last section also includes the overall idea of *Self-organizing* Context Spaces. From a high-level perspective, related ideas, research projects and groups are introduced, to which later sections may refer to. The idea is to avoid readers being overwhelmed with lots of very specific details before getting a proper introduction to each subtopic first. These detailed aspects are then addressed in related work sections dedicated to a specific subtopic or more briefly as part of that subtopic's introduction. After a conclusion on general related work (Sec. 4.4), an overview of these specific related work sections is provided in Section 4.5.

Projects mentioned in this chapter have spawned hundreds of publications. Although many of them are not directly relevant for this thesis, they may nevertheless serve as an overview or context providing paper. The chapter will thus mainly focus on the most relevant works from the perspective of this thesis possibly complemented by a few overview or background papers. Readers may thus briefly learn about approaches and findings in their respective project context (possibly with additional hints at broader overviews), before later sections refer to them again from the perspective of a certain subtopic. For example, many projects had to come up with a solution for user activity tracking or storing metadata. One project approached the problem with an eLearning background, another one from a corporate knowledge management perspective, etc. Metaphorically speaking, if one created a matrix having one row for each project and one column for each subtopic, then this overview is written from a row-perspective (projects), while later parts of the thesis typically focus on one column at a time as a subtopic-focused cross section of all projects.

Note: Parts of this chapter have already been published [187, 188].

4.1 Related Work on the Semantic Desktop and Similar Approaches

Section 2.3, which introduced the Semantic Desktop (SD) [80, 304] in more detail, already mentioned *Memex* [50] as an inspiring vision for many such approaches. Mainly because this PhD project was conducted at DFKI SDS but also since DFKI has been involved in many SD-related projects over the last two decades, that section had a slightly higher focus on DFKI SDS' ideas and realization of an SD. This chapter serves as a complement providing a broader view on the topic. However, detailed comparisons of the different solutions will not be presented here since cSpaces is just based on SD technology, not reinventing it from the ground up. Readers interested in more details are kindly referred to an overview by Dragan and Decker [96], which summarizes many of the mentioned approaches in a concise way serving as a good starting point for further reading. To the author's best knowledge, this overview is still the most recent one on SD approaches.

4.1.1 NEPOMUK and Adjacent DFKI Projects (2003–2012)

By far the most prominent SD project was "NEPOMUK – The Social Semantic Desktop" (2006–2008) [267], to the author's best knowledge, the so far highest budgeted European project dedicated to SD research (11.5 million EUR). It was coordinated by DFKI, and there were several directly adjacent DFKI projects: *EPOS* [110], *Mymory* [264], and some of the research was later continued during the time or as part of the *ADiWa* project [8].

EPOS – Evolving Personal to Organizational Knowledge Spaces (2003–2005). After two *Organization Memory*-focused projects at DFKI SDS, "KnowMore – Knowledge Management for Learning Organizations" (1997-1999) [208] and "FRODO – A Framework for Distributed Organizational Memories" (2000-2002) [124], EPOS [110] concentrated on the knowledge worker's perspective. This was also reflected by its title affix "Evolving Personal to Organizational Knowledge Spaces". Its objective was to leverage a user's efforts for their personal knowledge management for their own benefit as well as to evolve this within the organization [110]. Several works that started in EPOS were later continued in NEPOMUK [110], for example:

- the idea of the *Semantic Desktop* and its further elaboration [304, 81],
- the definition and first evaluations on the *Personal Information Model (PIMO)* [305, 306],
- findings by Dengel [88] about *subjectiveness* in PIM,
- works on context-specific *Proactive Information Delivery (PID)* [169, 170],
- works on user activity tracking, in particular *DragonTalk* [Sw107],
- Schwarz [341, 342] enhancing a *context model* by Maus [237] that was later additionally extended in this PhD project.

Follow-up projects of EPOS were Mymory and NEPOMUK.

Mymory – Situated Documents in Personal Information Spaces (2006–2008). Mymory [264] used many ideas and components of NEPOMUK [202], which took place in parallel. Its vision was 1) to employ technologies for unobtrusive user observation to create relations between information items that are meaningful to the user in their specific context, 2) to use attention evidence for more precise PID and 3) to provide mechanisms of meaning coordination to facilitate reusability of knowledge among different contexts [264]. Ultimately, Mymory should lead to a personal memory for knowledge workers, which not only was a passive storage, but also proactively supported context-driven structuring of its content and user-perspective interpretation and incorporation of arriving information [264]. Outcomes of the project involve topics like

- *Semantic Wikis* [202, 203],

- user activity tracking (e.g. *UserContext* [Sw106] as a collection of different solutions) and
- user attention as well as gaze-based filtering for document classification and annotations [203, 49].
- Additionally, *Skipforward* [201], a distributed recommendation system, introduced *applicability* and *confidence* values for semantic statements, which were also used in this thesis.

NEPOMUK – The Social Semantic Desktop (2006–2008). NEPOMUK [267] was a comparably large EU-funded project (budget of 11.5 million EUR) that brought together 16 research and industry partners from across Europe. It is an acronym for *Networked Environment for Personalized, Ontology-based Management of Unified Knowledge* which meant "a comprehensive solution for extending the personal desktop into a collaboration environment which supports both the personal information management and the sharing and exchange across social and organizational relations" [267]. An introduction was published by Groza et al. [141].

Apart from publishing more than 70 papers, several SD prototypes were created:

- From the perspective of this thesis, the most important prototype is the *NEPOMUK Semantic Desktop prototype* [Sw92] co-developed by DFKI SDS. However, there are also others, for example: *Gnowsis* [303] (a first reference implementation of some of NEPOMUK's functionalities [267] by Sauermann, who also was a member of DFKI SDS at the time), *DeepaMehta* [287], *X-COSIM* [122] and *Nepomuk-KDE* [268] as probably the most well-known prototype.

Other noteworthy contributions were

- the further elaboration of PIMO [306],
- first evaluations of Semantic Desktops [308, 121] that could show their superiority in supporting information management and knowledge work compared to traditional systems,
- first works on bringing index and structured search together [257],
- *TagFS – Tag Semantics for Hierarchical File Systems* [39],
- a desktop context detection method by Chirita et al. [62] and
- the PhD thesis by Maus [238] on *Workflow context to realize process-oriented assistance in Organizational Memories*, which was completed during the project.

ADiWa – Alliance Digital Product Flow (2009–2012). ADiWa [8] is an abbreviation for "Allianz Digitaler Warenfluss [Alliance Digital Product Flow]": 15 German partners from research and industry investigated applications that dynamically plan, control and execute complex business processes optimized by data from the *Internet of Things* [225] [8]. From the perspective of this thesis, relevant contributions are as follows:

- *ConTask* [152, 243], a context-sensitive task management system integrated in the *NEPOMUK Semantic Desktop* [Sw92] (discussed in more detail in Sec. 4.3)
- Relevant PhD theses were completed during or temporally close to the project:
  - Sauermann (2009) [303]: *The Gnowsis Semantic Desktop Approach to PIM*,
  - Schwarz (2010) [343]: *Context-awareness and Context-sensitive Interfaces for KW Support*,
  - Schmidt (2013) [317]: *Information Work Support Based on Activity Data*.

  The works by Schwarz and Sauermann (members or former members of DFKI SDS) greatly influenced this PhD project as mentioned in previous chapters and ones to come.
- The PhD thesis by Schmidt [317] provides a good overview of works by the research group of Schmidt, Stoitsev, Mühlhäuser et al., a collaboration between SAP SE[1] and TU Darmstadt[2]. Some of their related topics are:
  - an overview of user activity tracking methods [320], [317, Table 6.2],
  - task detection and mining [319, 318] and
  - activity-centric document clustering [345] (a few years after ADiWa).
- Development of *CoMem* [87] started and reached a stable prototypical state during the project (see Sec. 2.3.3 for details). First results were reported as part of Maus et al. [241].

4.1.2 Spiritual Successor Projects at DFKI (2013–2023)

There were several spiritual successor projects to NEPOMUK:

- *ForgetIT – Synergetic Preservation, Managed Forgetting and Contextualized Remembering* (2013–2016) [120],
- *supSpaces – Semantic Support (Knowledge) Spaces for Agile Knowledge Management in 3rd-level IT Support* (2015–2017) [370],
- *Managed Forgetting – Foundations, Methods and Effects of Managed Forgetting for Knowledge Workers (in Administration)* (2016–2023) [234] and
- *SensAI – Self-organizing Personal Knowledge Assistants in Evolving Corporate Memories* (2020–2023) [346].

The author of this thesis has been a core contributor to all of them. Thus, much of their content is presented in Chapters 5 to 9. However, since *ForgetIT* had a large consortium of research and industry partners and the *Managed Forgetting* project was part of a DFG priority program bringing it together with seven other topically related projects, a related work section for each of these projects follows (Sec. 4.2.1 and 4.2.2). Each section will give an overview of the respective project partners' work.

[1] https://www.sap.com/
[2] https://www.tu-darmstadt.de/

4.1.3 Similar or Related Works and Projects

Although several of the approaches presented in the following are not explicitly considered SD approaches, they nevertheless also deal with capturing metadata and context-sensitively supporting users. Some only cover partial aspects like search or task detection.

Works by Microsoft, in particular Microsoft Research (especially 2002–2006). Systems similar to a Semantic Desktop are *MyLifeBits* [131, 130] and *Stuff I've Seen* [98] as well as the announced but later canceled [66] *Windows Future Storage (WinFS)* [291]. Microsoft researchers also proposed an approach for predicting so-called *Memory Landmarks* [288, 172], "special landmarks or anchor events for guiding recall and for remembering relationships among events" [172]. Condensation measures later discussed in this thesis were inspired by this work.

In the context of works by Microsoft, two well-known software assistants are worth mentioning: First, there is the *Office Assistant* [Sw81] (also known as *Clippit* or *Clippy*) which was based on the *Lumière Project* [171] and available as part of the help functionality of *Microsoft Office 97 to 2003/2004* [Sw80]. The second assistant is *Cortana* [Sw77], which became available for *Windows 10* [Sw83] in early 2015 and for *Android* [Sw51] and *iOS* [Sw25] devices in late 2015 [394]. Cortana can be operated by voice and may help users with messaging, calendar and reminders, notes, etc. and search in general [393]. Over the years, however, Microsoft decreased its focus on Cortana, for example by shutting down its mobile variants in 2021 and reducing its presence in *Windows 11* [Sw83] [61]. Nevertheless, "Microsoft still sees value in conversational AI and the company is trying to reposition Cortana as an assistant that can improve Microsoft's enterprise-focused offerings" [394].

CALO (2003–2008). The *Cognitive Assistant that Learns and Organizes (CALO)* was a five-year project with a budget of 150 million US$ bringing together more than 300 researchers from 22 research institutions [360]. Semantic Desktops developed as part of CALO are *SEMEX (SEMantic EXplorer)* [94] and *IRIS (Integrate, Relate, Infer, Search)* [60]. The well-known *Enron Email Dataset* [206] [Sw33] also was collected and prepared in CALO. However, its most prominent outcome is probably the assistant *Siri* [Sw27], which started as a spin-off later bought by Apple Inc.[3] [360].

Works by Google (especially 2004–2011). Mentioning *Cortana* and *Siri* leads to the third major corporation having a similar solution: *Google Now* [Sw48], which was later (2016) succeeded by *Google Assistant* [Sw45]. In addition, there was also *Google Desktop* [Sw47] allowing for search on the local computer as well as widgets like a sidebar to show news, weather information etc. Google discontinued *Google Desktop* in late 2011 justifying their decision due to a "huge shift from local to cloud-based storage in the last few years" as well as "the integration of search and gadget functionality into most modern operating systems" [113].

TaskTracer (especially 2005–2009). Computer scientists of the Oregon State University[4], Dietterich, Shen, Stumpf et al., developed *TaskTracer* and several related applications to detect

[3] https://www.apple.com/
[4] https://eecs.oregonstate.edu/

tasks and task switches [97, 368, 350, 349] or recommend task-relevant resources [22, 348]. The PhD thesis by Shen [347] may serve as an additional overview.

APOSDLE – Advanced Process-Oriented Self-Directed Learning Environment (2006–2010). In an overview paper, Lindstaedt and Mayer stated that the goal of the *Advanced Process-Oriented Self-Directed Learning Environment (APOSDLE)* [13] was to "*enhance knowledge worker productivity by supporting informal learning activities in the context of knowledge workers' everyday work processes and within their work environments*" [227]. Researchers of the project defined the *User Interaction Context Ontology (UICO)* [282] to model user actions, resources on the computer, information needs, etc. in great detail. In Rath et al. [282], they explicitly compare UICO with PIMO [306] pointing out various differences (e.g. higher focus on user interaction) and some commonalities (e.g. representation of information items). They published several works on user activity tracking and task detection [284, 282, 283, 90]. An overview of their implemented tracking methods can be found in the PhD thesis by Rath [(281) as mentioned in (283)].

Works by Gyllstrom et al. (especially 2007–2010). Gyllstrom et al. published works on user activity tracking and task detection [150, 151] as well as an information value assessment approach called *LostRank* [149]. The PhD thesis by Gyllstrom [148] may serve as an additional overview.

ACTIVE – Knowledge-powered Enterprise (2008–2011). ACTIVE's goal was to "increase the productivity of knowledge workers in a pro-active, contextualised, yet easy and unobtrusive way. The aim [was] to convert tacit and unshared knowledge — the 'hidden intelligence' of enterprises — into transferable, interoperable and actionable knowledge to support seamless collaboration and to enable problem solving" [7]. Warren et al. [392] presented an overview of the project. Further publications involved their context model [111], context/task mining approaches [363, 362] or context-sensitive PID [135, 219, 364]. Section 4.3 presents further details especially with respect to context spaces.

SWELL – Smart Reasoning Systems for Well-being at Work and at Home (2011–2017). The objective of the SWELL project was to "develop user-centric sensing and reasoning techniques that help to improve physical well-being (mostly in a private context) and to improve well-working (in a work context)" [371]. "Well-working could be defined as 'being and feeling in control', with a positive impact on work efficiency and effectiveness, work pleasure, mental and physical health status" [371]. An overview of the project was published in Kraaij et al. [215]. Other Works were about context/task recognition [210, 209, 214, 301, 298], privacy and trust [41, 211] or context-aware recommendation [302]. Sappelli et al. [300] also presented an approach for e-mail importance estimation, which can be seen as a special case of information value assessment. Last not least, a dataset also containing PIM data was collected and made publicly available [212, 299]. The PhD thesis by Sappelli [297] ("Knowledge Work in Context") may serve as an additional overview.

DCON, PiMx(T) and (X-)iDeTaCt (especially 2012–2016). Over the years, the group of Abela, Staff, Handschuh, Scerri et al. (some of whom were also members of the NEPOMUK project) worked on several related topics: In Scerri et al. [311], they present *DCON*, a context ontology as an addition to PIMO to better model users' activity contexts. They also presented a *PIM Analytix framework (PiMx)* [5] "for the analysis and visualisation of evolving activity-data generated by individuals performing tasks on their desktops". Based on that, they developed *iDeTaCt*, a incremental density-based clustering algorithm to transparently generate task-clusters by exploiting document switching and revisitation [4]. In Abela and Staff [3], iDeTaCt was extended to *X-iDeTaCt* to also consider types of switching (e.g. clicking a hyperlink or the web browser's back or forward buttons, particular search patterns, etc.). The paper further presents *PiMxT (PIM for Tasks)* that uses X-iDeTaCt and allows for re-finding and resuming tasks.

Further Works. There are further Semantic Desktops like *Haystack* [199] or *MOSE* [(96) referring to (402)].

4.2 Related Work on Managed Forgetting

With regard to *Digital Forgetting*, the most prominent topic known to a general audience is presumably the *Right to Be Forgotten (RTBF)*. Although it is not a major aspect of Managed Forgetting, it is nevertheless worth mentioning here. (Especially with regard to corporate settings, Managed Forgetting could be extended to take RTBF rules into account.)

The Right to Be Forgotten (RTBF). The RTBF originated from Mayer-Schönberger (2009) [247] who proposed expiration dates for information items after which they should automatically be deleted (if not explicitly renewed by the user). In Europe, the RTBF is codified in Article 17 of the *General Data Protection Regulation (GDPR)* [112] and came into effect in 2016. Mayer-Schönberger also was a member of the ForgetIT project that is discussed in the following.

4.2.1 ForgetIT – Synergetic Preservation, Managed Forgetting and Contextualized Remembering (2013–2016)

The EU project *ForgetIT* (2013–2016) [120] brought together an interdisciplinary team of eleven partners from different universities, research institutes and companies across Europe. Their backgrounds covered various topics of computer science (especially PIM, information retrieval, multimedia analysis and cloud computing) as well as cognitive psychology, law and economics. Since the author of this thesis has worked in the project, its core topics, Managed Forgetting, Synergetic Preservation and Contextualized Remembering, are addressed throughout all chapters of the thesis. Thus, summarizing large parts of the thesis here is omitted. However, since the project consortium was comparably large, the work of partners can be seen as related work, especially those works with little or no contribution by DFKI or the author. At the end of the three-year project, the consortium published a book called "Personal Multimedia Preservation: Remembering or Forgetting Images and Video" [251] giving an overview of all contributions. Apart from what is presented in this thesis, the book also shows other perspectives and facets brought in by colleagues, e.g. higher foci on image processing [14], natural language processing [140] or cognitive psychology [230].

4.2.2 SPP 1921 – Intentional Forgetting in Organizations (2016–2023)

SPP 1921 Overview. The German Research Foundation (DFG) set up a two-phased priority program on "Intentional Forgetting in Organizations" [359] spanning more than six years. In this program, eight interdisciplinary research teams, mostly consisting of computer scientists, cognitive psychologists or ergonomists, investigated different variants and application scenarios of intentional forgetting. Each team had their own background and perspective on the topic ranging from knowledge representation and logics, to the use of ontologies in product design (e.g. in mechanical engineering), to processes and roles in socio-technical systems, to personal information and knowledge management, etc. Beside each project's individual publications, the consortium has published several cross-project papers on these different perspectives, some more from a computer science perspective [374], others more from the point-of-view of cognitive science [108, 107], roughly speaking. Additionally, a special issue of the German Journal of Artificial Intelligence [217] presented topics and contributions of the program, and by the time of writing this thesis, there is another paper in preparation [309] showing further aspects, dimensions and application scenarios of intentional forgetting.

The *Managed Forgetting* project [234], in which the author has worked, also was among the eight teams. Again, summarizing the content of various parts of this thesis at this point is omitted. The work of project partners is however related to varying degrees. Since only the most related works are discussed in the following, readers may use the (overview) papers mentioned above as well as the program's website [359] as starting points for further reading.

Dare2Del. The SPP project closest to Managed Forgetting was *Dare2Del*[5] by Niessen, Schmid, Göbel, Siebers, Nandini et al. from the Universities of Erlangen-Nürnberg and Bamberg. In Dare2Del, a deletion assistant was developed [355, 354] trying to infer deletion candidates among the files on a user's computer. It utilized *Inductive Logic Programming* [(355) referring to (263, 280)], a white-box method allowing for explaining the system's proposals, which facilitates users' trust in the system. A list of deletion candidates is shown and users ultimately have to decide what to delete. This is an alternative approach to Managed Forgetting and cSpaces in particular, which operates in a self-organizing way without explicitly asking for permission (undo operations are possible if decisions are dissatisfying). As mentioned in Section 3.1, this was a deliberate decision since bringing items back into the user's view and mind by explicitly asking for permission is considered to be against the "natural flow" of things slowly fading out of memory. Investigating effects of explanations on aspects like acceptance of suggestions or trust, Göbel et al. [133] pointed out another issue worth considering in this regard: participants felt more insecure whether to accept a deletion suggestion if they did not know or remember a file. The authors thus recommended that deletion suggestions should be shown in a timely manner when users are still able to remember a candidate file. They could further show that explanations lead to better acceptance, and deletion actually leads to less recognition of deleted files in a later inquiry, i.e. a forgetting effect sets in as desired. Note: as far as the author knows, findings in Göbel et al. [133] are first results of an experiment series still in progress. Readers are thus encouraged to look for respective follow-up publications.

[5] http://www.spp1921.de/projekte/p6.html.en

4.2.3 Further Works on Digital Forgetting

This section is about further works on digital forgetting. Note that this list is only a selection intended to show that forgetting has already been a research topic in various subfields of computer science. However, apart from a few "hot spots", the author would still classify it as a niche topic – but a versatile one, applicable in many scenarios as shown in the previous as well as following paragraphs.

Knowledge Representation and Reasoning. Projects of the *SPP 1921* (see above) already showed several quite different scenarios of (intentional) forgetting. Moreover, colleagues and fellow researchers of the program, Eiter and Kern-Isberner [103], have presented an overview on forgetting in knowledge representation and reasoning.

Robotic Memories. Gurrin et al. [145] propose a model of memory for robots, which "incorporates an element of forgetting to ensure that the robotic memory appears more human and therefore can address some of the challenges for human-robot interaction".

Life Logging. In their comprehensive summary, Gurrin et al. [146] named several challenges and issues life logging is facing. Beside challenges in the areas of capturing and accessing data, ownership or privacy concerns, they also discussed the aspect of forgetting. Some researchers see life logging as "an antithesis of forgetting, while others aim at modeling the human experience of forgetting in surrogate memories" [146].

Further Examples of Forgetting in PIM. In addition to the ForgetIT and Managed Forgetting projects, there were also other examples of forgetting in PIM: Schmidt's layered model for user context management applies an "aging mechanism as a form of controlled forgetting" to readjust confidence values. This "limits the impact of incorrect context data, which considerably increases the robustness of the system" [316].

Bergman et al. [28] propose a *GrayArea* as "an additional folder feature that allows the users to drag information items of low importance to a designated location at the bottom of a folder". This makes them less visible but their non-removal is important, so they can be retrieved in their original context, and users do not need to remember an additional storage place [28]. They also propose *Old'nGray*, a mechanism that automatically sets older versions of a file to gray, so users do not need to set file ordering by date in order to track which is the most recent version, for example.

Works on Machine Unlearning. The topic of *Machine Unlearning*, i.e. the removal of individual data points from a trained model (briefly speaking), has experienced a boost in recent years, for example by regulations like the GDPR [112] enforcing data protection regulations like the right of being forgotten (see beginning of this section). Overviews and often cited papers in this regard are Cao and Yang [51], Bourtoule et al. [42] or Gupta et al. [144].

4.3 Related Work on Context Spaces, in particular Self-organizing Context Spaces

This section contains related work about Context Spaces as well as the overall idea of *Self-organizing* Context Spaces.

Prior Work at DFKI SDS. Works at DFKI SDS most related to this PhD project are *ConTask* [152, 243] and the PhD thesis by Sven Schwarz [343], whereas the latter extended aspects presented in the PhD thesis by Heiko Maus [238], for example his context model (see Sec. 5.1).

Based on the *NEPOMUK Semantic Desktop Prototype* [Sw92] and the user activity tracking solution *UserContext* [Sw106], *ConTask* [152, 243] allows for context-oriented personal task management. Figure 4.1 shows its user interface having a *task diary* (left) to list open tasks, a *task editor* (center) and a PID sidebar on the right showing concepts of the user's PIMO and recently accessed resources [243].

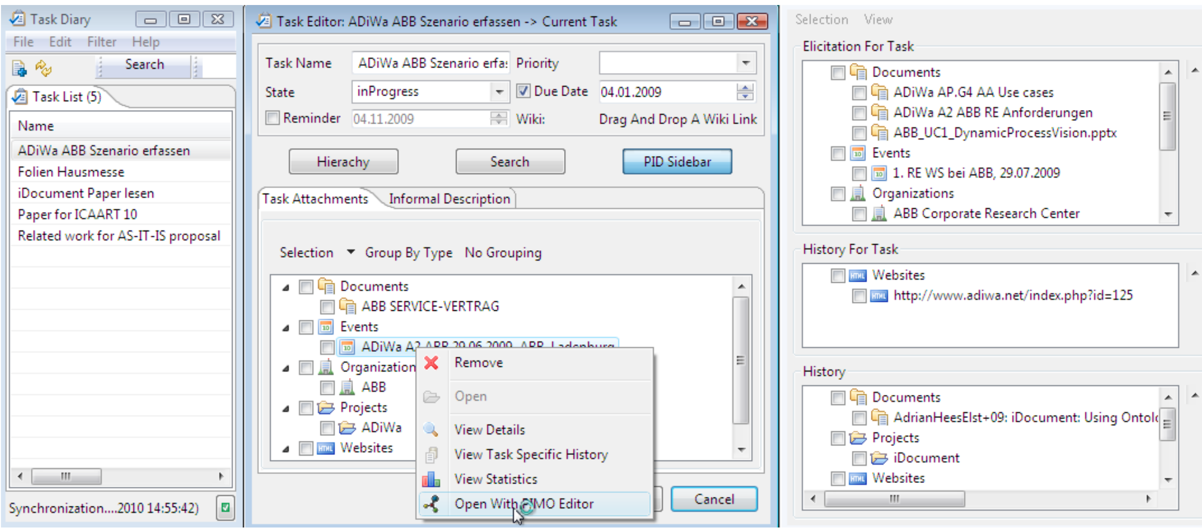


Figure 4.1: ConTask: GUI [152, 243].

(Note: This is a screenshot by Heiko Maus published multiple times [152, 243].)

The *UserContext* [Sw106] framework mentioned above was developed as part of the PhD project by Sven Schwarz [343]. His PhD about "Context-awareness and Context-sensitive Interfaces for KW Support" presents a full round-trip of a typical user support scenario ranging from user activity tracking to information extraction, to context elicitation and context-sensitive assistance, for example by PID. Schwarz explored several context-sensitive user interfaces, two of which were a horizontal and vertical assistance bar. The vertical variant is depicted in Figure 4.2. It consists of four parts: *textual entities* (e.g. documents), *topic entities* (topic categories leading to file/email folders of the user), *contact entities* and *process-oriented entities* [343]. The screenshot shows the situation of clicking one of the topic entities, which opens a context-relevant folder on the user's harddisk.

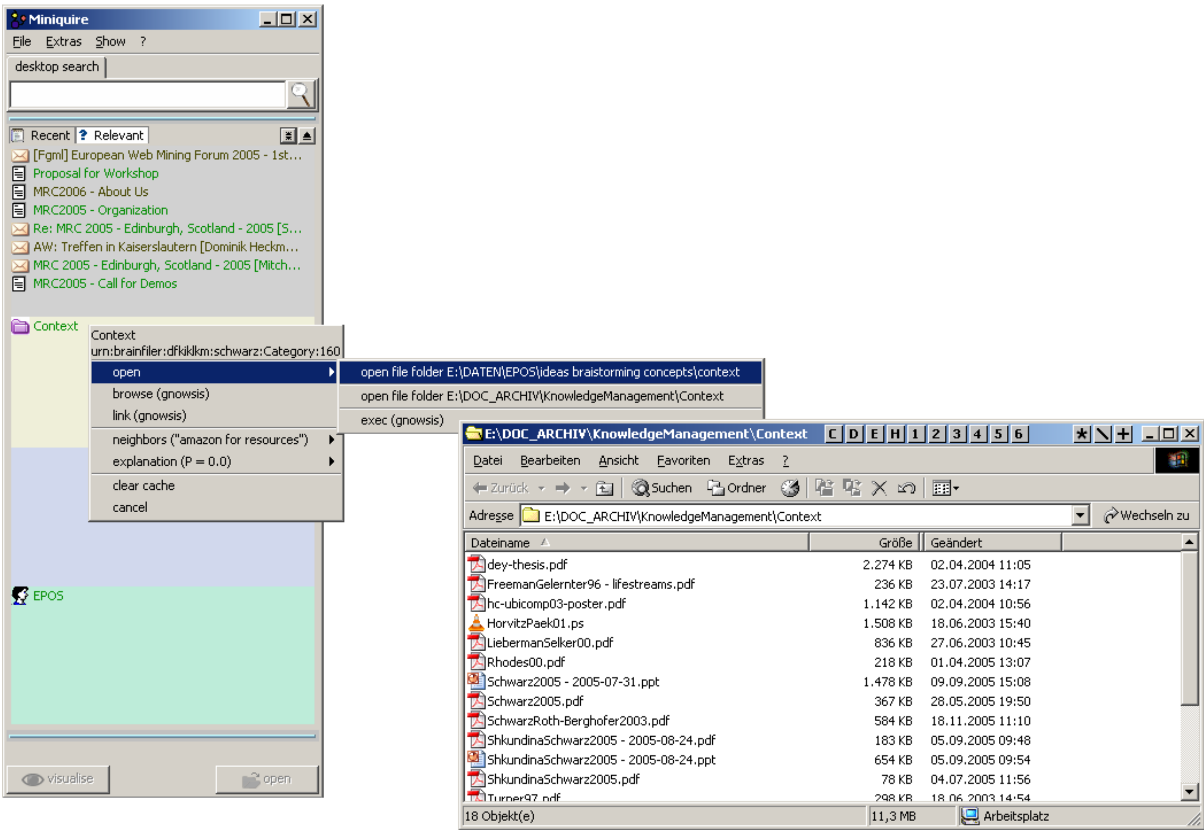
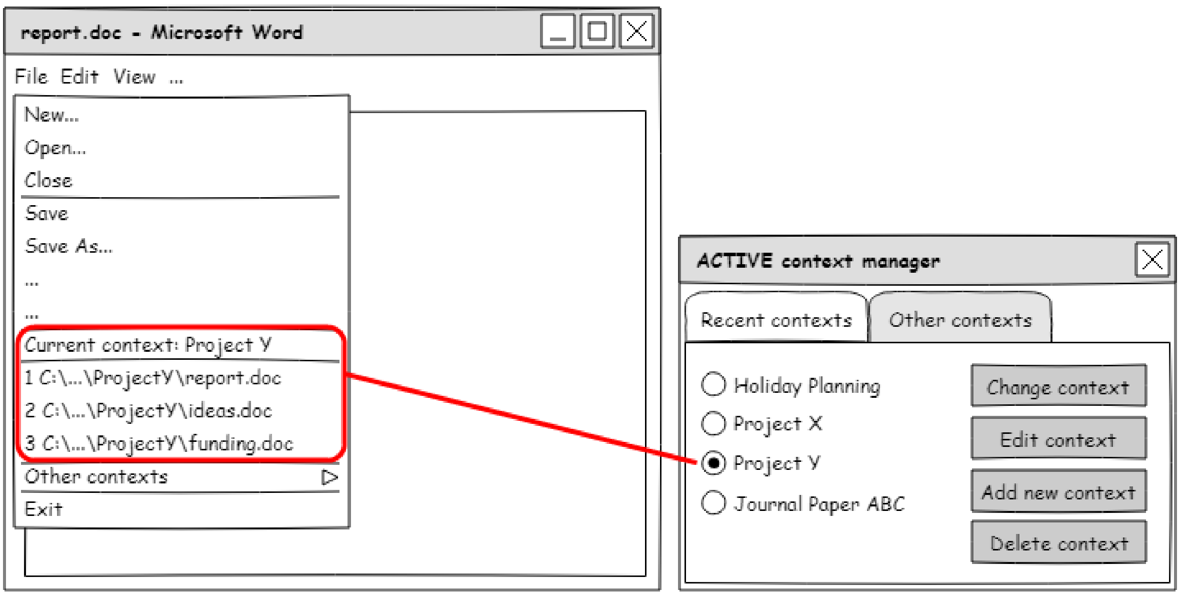


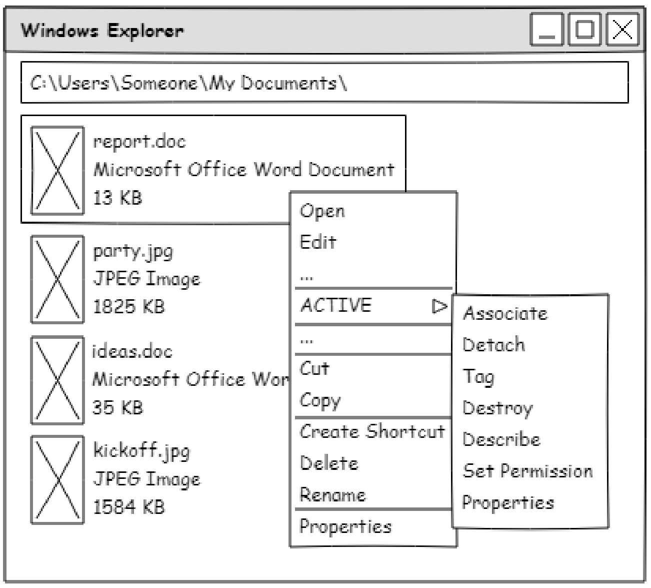
Figure 4.2: Sidebar Developed by Schwarz [343].
(Note: This is a screenshot by Sven Schwarz published in Schwarz [343].)

The ACTIVE Project. Of all Semantic Desktop or similar approaches presented in Section 4.1, the *ACTIVE project* is the one most related to this PhD project with regard to Context Spaces. The project follows an approach in which users are actually aware of the concept of context [135]. They may select a context from the *ACTIVE context manager* in which they would like to work next. As shown in Figure 4.3a, the system then updates the recently used documents list in *Microsoft Word* [Sw84] as a support measure, for example. In Warren et al. [392], they presented further features of ACTIVE like associating a file with a certain context or detaching it. Tagging or writing descriptions is possible as well. These actions can be triggered using the context/right-click menu of the *Windows File Explorer* [Sw85] as shown in Figure 4.3b.

Semantic File Systems. Allowing users to actually work in Context Spaces ultimately requires making cSpaces available in many (ideally all) of the users' applications. Especially the file system is interesting in this regard since it is available in all applications. Thus, it is an effective approach to utilize virtual (as in the case of this PhD project) or semantic file systems. To the author's best knowledge, the paper that first mentioned the idea of a semantic file system is the one by Gifford et al. (1991) [132]. Since then, more than 700 publications cited the paper, whereas only a fraction of them presents a semantic file system as well. Among them are the previously mentioned *SemDAV* [312] and *TagFS* [39]. A more recent overview analyzing



(a) Selecting a context in the *ACTIVE context manager* (right) updates the *Recent Documents* list in *Microsoft Word* [Sw84] (left) as shown in Gomez-Perez et al. [135, Fig. 1].



(b) Actions like associating with or detaching from a context, tagging, etc. were integrated into the context/right-click menu of *Windows Explorer* [Sw85] as shown in Warren et al. [392, Fig. 5].

Figure 4.3: Context-related GUI Updates and Context Menu Actions in the ACTIVE Project. (Note: This figure shows simplified drawings by the author to reflect the core statements of screenshots published by Gomez-Perez et al. [135, Fig. 1] and Warren et al. [392, Fig. 5].)

several of such systems in various dimensions is provided by Watson et al. [395]. Traversing a knowledge graph and thus potentially also cSpaces is theoretically possible with several of these systems. However, to the author's best knowledge, seeing a user's IM/KW context as an entity to traverse via the file system has not been realized by any of these systems. The same is true for self-organization: none of these systems is self-(re)organizing except for one: In their paper called "Self-* Systems" (encompassing self-managing, self-configuring, self-tuning, etc.), Mesnier et al. [250] do address and apply self-reorganization but in the scenario of low-level file system organization and not with respect to users' IM/KW contexts.

4.4 Conclusion on General Related Work

This section concludes the general related work on the three cornerstones and the overall idea of cSpaces presented above.

Semantic Desktop and Similar Approaches. Although proven to be successfully applicable to IM and KW support and even superior to traditional systems [308, 303, 121, 241, 87], SD approaches are still not widespread. As possible reasons Dragan and Decker [96] name

- a missing "killer app",
- the need to get familiar with new semantic applications replacing known traditional ones (not all SDs have this problem as stated by the authors) and
- the cold start problem.

Developers of *Nepomuk-KDE* [268] additionally name

- performance problems (NEPOMUK being a "CPU consumer"),
- RDF raising "the knowledge needed to contribute to a point where most developers decide to skip it",
- RDF being a "flexible way to store data" but "not the most efficient way" (e.g. complete normalization even if "quite often not required"),
- "no existing RDF storage [being] designed to work in a desktop" (e.g. "quickly [using] hundreds of megabytes of RAM")
- massive data duplication and synchronization problems (e.g. data residing in its original store as well as a copy in the triple store and another one in the index structure of the triple store),
- API duplication: data residing in both NEPOMUK and other stores makes it unclear from which stores to access it from [154].

In 2014, KDE [Sw65] dropped NEPOMUK in favor of *Baloo* [200], a file indexing and search framework. Note that the above citations are from KDE mailing lists [154] explaining this decision and additionally reporting about past experiences that led to it. The way some issues are stated is possibly too sweeping a judgment to stand up to scientific scrutiny. However, the core of each issue as well as the sum of reported problems gives an impression of what went wrong.

In summary, there are positive experiences with SD approaches as well as the problematic aspects just listed. With regard to research question RQ5, one can say:

- The Semantic Desktop is a suitable ecosystem for realizing cSpaces: there were several successful applications in IM and KW support scenarios [308, 303, 121, 241, 87] and other researchers came to the same conclusion [44]. In addition, cSpaces can benefit from several assets and experiences of projects mentioned above, in particular CoMem [87], which is based on NEPOMUK [267].

- Managed Forgetting and especially self-(re)organization could be a "killer app" mentioned by Dragan and Decker [96]. All systems mentioned above did not have such a high degree of automation and user support.

- To gather additional context meta-information, Context Spaces are build on top of the SD – see the interplay of SD, Managed Forgetting and Context Spaces mentioned in the previous chapter (Sec. 3.3).

- Given the reported problems mentioned above,

  – data and API duplication should be avoided (also see *ATIC knowledge graphs* in this regard, Sec. 5.2) and

  – measures reducing the cold start problem should be taken (see Sec. 5.3).

  – Using transparent injections of cSpaces into existing applications respects users' familiar environment.

  – Using only very few extra widgets that are additionally usable with a wide range of applications instead of a plug-in with its own GUI for each application reduces additional learning effort for users (further details in Sec. 6.1.1).

  – Knowing that phases of intensive calculation on a desktop can be perceived very negatively by users may help in the consideration and design of methods (e.g. possibly split calculations into smaller tasks and perform them in phases of no user activity, etc.).

  – SD background methods should be as real-time capable as possible making the system highly responsive.

  – SD approaches should protect users' privacy by design, which is also a future challenge mentioned by Dragan and Decker [96] (see Sec. 6.1.2).

- Despite the increase of cloud solutions, which Google named as one major reason to discontinue *Google Desktop* [Sw47] [113], performing IM and KW still requires local hard- and software. For cSpaces, the actual storage location of information is basically irrelevant. What is important is the user's access to and interaction with it. Thus, all cSpaces support measures are also applicable in cloud-based scenarios.

Managed Forgetting. Although there are several forgetting-related approaches, only few of them are similar to this PhD project. The most relevant one, the *Dare2Del* project, does not follow a self-organization approach. Instead, possible deletion candidates are presented to the user. From the perspective of this thesis, this has the disadvantage of bringing items back into the user's view and mind, although they might have already been forgotten. To reduce users' insecurity, members of the project recommended showing deletion candidates while users can still remember them. From the author's experience, this raises doubts whether users so closely after working with a file are able to ultimately decide whether the file is still needed or not. One hypothesis would be that they do not know and since in doubt reject the deletion suggestion.

Other approaches like a folder's *GrayArea* [28] need manual effort for every single information item. Although this is better than not offering any of such mechanisms, it is however doubtful whether the approach will scale. As mentioned earlier (especially see Sec. 3.3.2), users often do not find the time to tidy up.

Context Spaces, in particular Self-organizing Context Spaces. Of all presented Semantic Desktop (and similar) approaches, the *ACTIVE project* [7] is the one closest to cSpaces. The approach assumes that users are aware of the concept of context and allows them to select their current one. As a support measure, recently used document lists are updated according to the selected context. But similarly to *GrayArea*, adding and removing items from a context requires manual effort for each individual information item. In summary, ACTIVE and cSpaces share similar ideas, but cSpaces goes a bit further with respect to working in context spaces (e.g. by injecting contexts into existing systems) and self-organization (Managed Forgetting measures). The former reduces (or even eliminates) the problem of contexts being an island solution, whereas the latter raises the system's degree of automation and user support.

Schwarz [343] is the directly preceding context-focused PhD thesis in DFKI's SDS research department. cSpaces extends the project in various aspects, whereas the following are the most important ones:

- cSpaces adds self-organization thus raising the degree of automation and user support.
- As in the comparison with ACTIVE, cSpaces goes a bit further in terms of working in context spaces (see above).
- cSpaces extends its context model by three dimensions (see Sec. 5.1), one of them being hierarchical contexts.

Comparing *ConTask* [152, 243] with cSpaces basically yields the same results except for task (context) hierarchies, which are also possible with ConTask.

With regard to *Semantic File Systems*, traversing a knowledge graph and thus potentially also cSpaces is theoretically possible with some of the mentioned systems. However, to the author's best knowledge, seeing a user's IM/KW context as an entity to traverse via the file system has not been realized by any of these systems, and the same is true for self-(re)organization in IM/KW scenarios.

4.5 Specific Related Work on Detailed Aspects (Overview)

As mentioned in the beginning of this chapter, specific related work on detailed aspects is addressed in later sections of the thesis – either as dedicated sections or more briefly as part of a topic's introduction. Table 4.1 provides an overview of these sections.

Section	Topic
5.2.1	ATIC Characteristics
5.3.1	Crawling (Big) Personal Data
6.1	User Activity Tracking
6.2.1	Inflection-tolerant or Very Fast Named Entity Recognition (NER)
6.3	Context Mining Using Activity Data
7.1	Information Value Assessment (IVA)
8.1.1	PIM Datasets and Evaluation Challenges
8.3.1	Highlighting in Web Documents
8.3.3	Quantifying Task Resumption Lags in IM and KW Scenarios
8.8	Trust in AI-based Assistants

Table 4.1: Overview of Sections Addressing Specific Related Work (either as dedicated sections or as part of a topic's introduction).

This overview concludes the related work chapter and the thesis' introductory part. Next, the main approach (aligned with the cSpaces interaction cycle) is presented in Part II.

Part II

Approach

5 Context Modeling, Storage and Bootstrapping

Part II addresses this thesis' main approach in four chapters, Chapters 5 to 8.

In the current chapter, context modeling, storage and bootstrapping aspects are discussed, which are step/aspect 0 of the cSpaces interaction cycle (*Data Storage and especially Knowledge Graphs*). First, Section 5.1 presents the context model of cSpaces. Next, *ATIC knowledge graphs* are introduced as this PhD project's knowledge representation and storage solution (Sec. 5.2). Section 5.3 then proposes a bootstrapping approach for cSpaces consisting of three steps, each with its own dedicated tool: *PIM Crawler* (5.3.1), *Semantifier* (5.3.2) and *Contextifier* (5.3.3). Last, the Excursus Section 5.3.4 addresses knowledge graph construction in general as this PhD project's major side topic.

> Note: Parts of Sections 5.1 to 5.3.3 have already been published [182, 185, 325, 160].
>
> The Excursus Section 5.3.4 mainly summarizes collaborative work with and led by Markus Schröder in his PhD project [323]. Thus, its content has already been published by him [322, 324, 326–336]. These publications are not considered an explicit part of this PhD thesis (see declaration of side publications in the preface).

5.1 Context Model

The context model used in this PhD project is an extension of the model by Schwarz [341, 342], which itself is an extension of the one by Maus [237]. In his PhD project, Maus investigated *Workflow context to realize process-oriented assistance in Organizational Memories* [238]. This is reflected by the first five aspects adopted by Schwarz's model:

1. **Organizational aspect**: current role of the user, relevant projects and persons
2. **Historical aspect**: previous tasks, workflow audit
3. **Causal aspect**: tasks/workflows and task concepts (goals)
4. **Informational aspect**: current/recent documents, relevant topics and domains
5. **Operational aspect**: active applications and tools, recently used services

Schwarz's PhD project was about *Context-awareness and Context-sensitive Interfaces for KW Support* [343]. He thus added three more aspects:

6. **Behavioral aspect**: native operations, user actions

7. **Environmental aspect**: location (room), hardware (PC, notebook)

8. **Attentional aspect**: text scope (cursor), read/skimmed text

Further extensions by the author to realize cSpaces were as follows:

9. **Hierarchical aspect**: Context spaces are ordered hierarchically (e.g. as a tree or DAG), i.e. contexts can have super- and sub-contexts.

10. **Forgetting aspect**: Parts of a context (several information items) could be forgotten, condensed, temporarily hidden, etc. depending on the current situation and the support scenario.

11. **Focal aspect**: Presenting things in their respective context is usually already helpful. This aspect additionally shows which items were particularly in focus when last working in the context. The idea is that a context may present its "golden thread" to the user upon revisitation.

Figure 5.1 gives a summarizing overview of these eleven aspects:

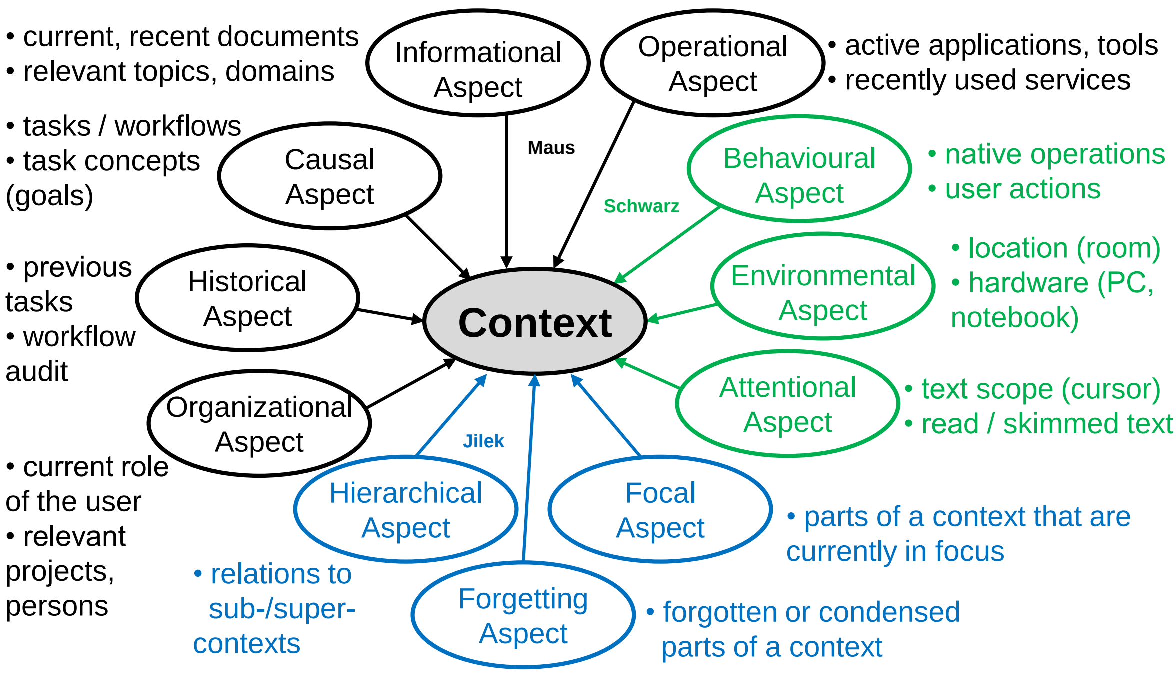


Figure 5.1: Context Model of cSpaces.
Colors: *black: initial model by Maus [237] / green: extensions by Schwarz [341, 342] / blue: further extensions by the author.*
(Note: A variant of this figure has been published in Jilek et al. [185].)

Note that some aspects were more relevant for this PhD thesis than others. However, it was a deliberate decision to extend previous context models of the SDS research department to make the model more versatile and be prepared for future work, e.g. if future researchers bring cSpaces closer to process/workflow management, the model would already support this.

Having introduced cSpaces' context model, the next section is about *ATIC Knowledge Graphs*, its knowledge representation and storage solution.

5.2 ATIC – Specifically Tailored Knowledge Graphs as Storage

cSpaces is based on Semantic Desktop technology and thus also has the *Personal Information Model (PIMO)*, a personal knowledge graph (KG), as its base for knowledge representation (see Sections 2.3.1 and 2.3.2). However, cSpaces extends PIMO to explicitly also represent context spaces as discussed in Section 3.3. This extension is also reflected technically: cSpaces adopts parts of CoMem's design and functionality. Major intentions behind this decision were as follows:

- to benefit from years of experience in the SDS department,
- reuse of ideas and components,
- contributing to CoMem-based research and industry projects during the time of this PhD project and
- by staying (mostly) compatible to CoMem allowing for re-integration of cSpaces ideas and assets into future CoMem versions.

Brief Technical Summary of CoMem. CoMem uses a client-server architecture. Its back-end is written in *Java* [Sw95], while its front-end is web-based using *HTML5* [275]. Clients typically call back-end methods by means of *JSON-RPC* [194, 195] in a REST-like way [118].

CoMem utilizes ontologies that emerged from the *PIMO ontology* [305–307] and stores knowledge in the form of triples. These triples are technically split into resources (nodes in the KG) and properties (edges in the KG), whereas the latter are additionally divided into object properties (*subject-predicate-object, SPO*) and literal properties (*subject-predicate-literal, SPL*). This approach is similar to *Sesame* [48] (especially see Section 6.3 of Broekstra et al. [48]).

CoMem uses a *Relational Database (RDB)* [67] as its major store: *MySQL* [Sw100] in early years and *MariaDB* [Sw72] later. According to discussions with Sven Schwarz, one of CoMem's architects and main developer (see Sec. 2.3.3), choosing an RDB was mainly done to benefit from its transaction concept and the availability of very stable implementations.

Another technical overview complementing the statements above can be found in Maus et al. [241].

cSpaces vs. CoMem. Adding explicit context spaces to PIMO was already mentioned as one extension of cSpaces to CoMem. Other extensions are related to CoMem being a comprehensive Organizational Memory solution with the Semantic Desktop as just one of its pillars and the goal of becoming an industry-ready product. This imposes requirements to user privacy, user access control, reliability and stability of added components, etc. For example, sending every mouse click, accessed email, file etc. to the CoMem server would be too privacy-harming for many users. However, cSpaces relies on such highly sensitive data.

As a solution, cSpaces operates completely on the user's local desktop without sending any information to another computer. On the other hand, CoMem needs user access control. Since not all resources are shared with others and remain in private use of their owner, protecting them from being accessed by unauthorized users is necessary. This aspect can be completely neglected in cSpaces since it is a personal application installed on a user's local machine. Other users do not have access to the system – only if folders, ports etc. are explicitly shared by the user by adjusting operating system settings accordingly (not via cSpaces).

cSpaces as envisioned in this PhD thesis – at least in its first prototypical state – is thus more desktop-centric exploring new ways of user support directly embedded in their personal information sphere before possibly transferring and scaling solutions to group or organizational levels. However, such organizational and group aspects are not completely out of scope, they are just a secondary priority: Some solutions developed during the work on cSpaces have already been transferred to CoMem and industry partners (see Sec. 9.1).

The author of this thesis has been in several discussions with Markus Schröder and Sven Schwarz about requirements for a knowledge graph (store) suitable for cSpaces as well as CoMem (future versions). Results are presented in the next section.

5.2.1 ATIC Characteristics

The most important characteristics of a knowledge graph (store) suitable for cSpaces as well as CoMem (future versions) happen to start with one of the letters A, C, I or T. Therefore "ATIC" was chosen as a term to summarize these characteristics. It refers to the idea of CoMem and cSpaces being a meta-model over several data stores to represent a user's personal information sphere – just like an *attic* is the architectural superstructure of a building.

The ATIC characteristics are a follows:

- **Applicability-attributed statements**: Relations between items, e.g. a context and its tags or elements, should not be binary but more fine-tunable (metaphorically speaking not black or white but in grayscale). It should be possible to associate statements with a certain degree of applicability (given as a value between -1 and 1) as proposed in Kiesel and Schwarz [201]. Finding contexts dealing with a certain topic should be easier if one may state that the topic is a major or minor topic or one of "average" relevance for the context. If one would like to express, for example, that a certain context C is primarily about skiing but also refers to certain types of snow (and is therefore also about snow but not primarily), they could do this using different values of applicability:

  - *Context C –is about→ skiing* (1.0) and
  - *Context C –is about→ snow* (0.6).

  One could also express that context C is definitely not about surfing using negative applicability:

  - *Context C –is about→ surfing* (-1.0).

- **Time-dependency**: To do research on and especially evaluate features like Managed Forgetting, setting the KG back to an arbitrary point in time would be helpful. Thus, resources and statements have to be (virtually) added or removed as if it actually was that particular point in time. There are approaches like *Temporal RDF* [147] which go into that direction but differ in the concrete elaboration of such temporal aspects (e.g. statement validity).

- **Information management-focused**: PIMO is used as a core element of ATIC. It comprises mental concepts as well as statements (made with these concepts) about resources on a person's computer like files, mails, bookmarks, etc. and folder hierarchies. This imposes the challenge of how to handle the border between items conceptually represented in the KG and items that are not yet "rebirthed" since they only exist for technical reasons and/or the user is not aware of them. As an example, think of a person opening a Microsoft Word [Sw84] document: This triggers the creation of a temporary file while the document is in use by the application [252]. Regular users might not even be aware that such files are created. Nevertheless, the system has to deal with it, e.g. by observing file content updates, possibly even serving the file itself (if cSpaces are accessed using a virtual file system), etc. Another example are shared documents: They are available in a user's information sphere (e.g. via a network share) although the user is still unaware of their existence, not to mention their content. Is it really justified to represent such items in the *personal* information model of that user? The author therefore defines a **Shadow Knowledge Graph** to cope with those parts of a user's information sphere that are not explicitly expressed in the actual KG but nevertheless need to be handled by cSpaces. The metaphor of a Shadow KG refers to the *shadow cabinet* in politics, which has already been set-up by the opposition, and the only thing still missing is the mandate by the voters to set them in office as the actual cabinet. In the case of Shadow KGs, only the rebirth is missing, which basically just assigns a KG ID to a native resource making it available as a new node in the KG. Everything else like NER results (as virtual tags), index structures, etc. have already been there. More details are presented in the next section.

  What belongs to the actual (core) KG and what to the Shadow KG is an open question. For the scope of this thesis, the author decided to take *user awareness* as a distinction whether things should remain "in the shadows" or move into the core.

- **Individual-centric**: As mentioned in Section 2.2, information and knowledge emerges with an individual while performing typical PIM activities. Support systems as well as their respective stores should therefore be tailored to a person's individual needs and subjective views [88, 28], for example with respect to vocabulary, but also with regard to privacy. Additionally, the system/store has to cope with such knowledge (at least parts of it) later spreading into surrounding groups like a team, the department, the organization – see the last item (*corporate memory*) in this regard.

- **Continuously changing (evolutionary)**: The KG is changing (evolving) constantly, i.e. new resources or statements are added or removed very often, possibly several times per minute or even second. Especially in the cSpaces scenario, a balance between real-time user support and constantly evolving, ever growing KGs needs to be found.

- **Context-focused**: As mentioned before, PIMO is extended to explicitly also represent context spaces. It thus not only contains a user's individual view on their data but especially their different contexts and which items are actually associated with each context space. Every resource and statement may be associated with one or more contexts. With regard to the third item in this list (*IM-focused*), remember that cSpaces are a hybrid construct, containing KG entities as well as native data (see Sec. 3.3.3).

- **Confidence-attributed statements**: In our scenario, cSpaces tries to predict user contexts, switches between them and which items belong to which contexts. Since the ultimate answer to these questions is only available in a person's mind, the aspect of uncertainty has to be incorporated into the system. In some cases, the system is able to collect strong evidence from user activities to base its decision on. In other cases, its reasoning is more speculative. Statements should therefore be assigned with a score reflecting the system's confidence (certainty) as proposed by Kiesel and Schwarz [201] (in addition to applicability scores, see first item above).

- **Corporate Memory**: The KG primarily represents the PIMOs of individual persons, but parts of them will be shared with others. Thus, the KG as a whole is actually a GIMO consisting of all individual PIMOs, their shared overlap as well as other sources of information (e.g. intranet shares or world knowledge), together forming the basis of an organizational/corporate memory. As mentioned in Section 2.2, the knowledge flow between an individual and the corporate memory is bidirectional: on the one hand, knowledge emerges with the individual and spreads into the group (by sharing parts of their PIMO). And on the other hand, an individual may use and adopt information and knowledge already available in the corporate memory (from sources named before). In this mixture of private and shared data, user access control becomes relevant. There are several works with regard to this matter, for example by Jain and Farkas [177] or Flouris et al. [119] – just to name a few.

These ATIC characteristics are so far just a list of requirements or helpful features cSpaces and CoMem would benefit from. A first prototypical implementation has been developed as part of this PhD project and is presented in the next section.

5.2.2 ATIC Knowledge Graph Store Prototype

There are already solutions for several of the features mentioned in the previous section. However, to the author's best knowledge, there is no store that comprehensively provides all of the above functionality in an efficient way, especially as a free or open-source solution. Given the high amount of different types of data and metadata, the most promising solutions to explore are presumably *Multi-model Databases* (MMDB) [231] or *RDF**-based [155] solutions. The author decided to go for the former. With the additional major constraint to stay (mostly) compatible to CoMem, ATIC is a CoMem-inspired, -adopting and -compatible solution with enhancements in the direction of MMDBs. Its basic architecture is depicted in Figure 5.2, while Figure 5.3 shows the corresponding class diagram.

The main part of the system (center) is the actual **Core KG**. As described before, the **Shadow KG** (also called **Native Metadata Store**) begins "at the Core KG's border". Native

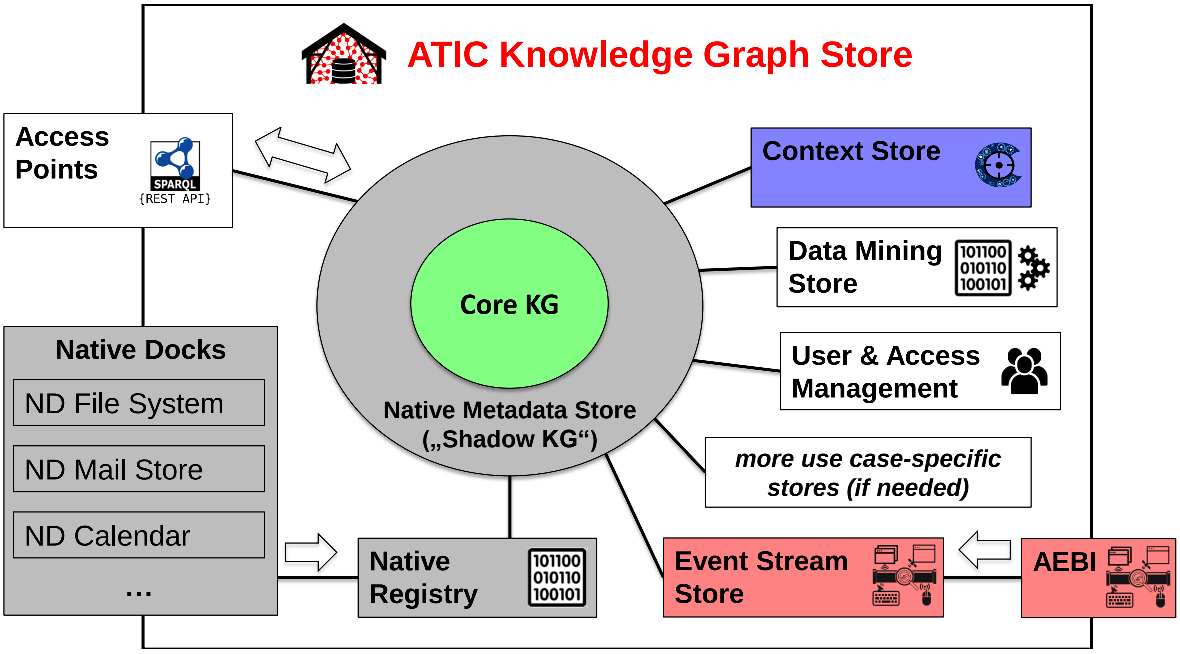


Figure 5.2: ATIC Knowledge Graph Store: Architecture Diagram.
(Note: Colors deliberately match those of the corresponding class diagram in Figure 5.3.)

Data (resources) are files, mails, bookmarks, etc. existing on the user's computer, the intra- or internet. To avoid data duplication, one of the stated reasons for the Nepomuk-KDE failure (see Sec. 4.4), ATIC only indexes the data, performs NER on it, etc. In this regard, a feature envisioned by Markus Schröder called *deep-linking of (desktop) resources* [324] is interesting. This technology allows for addressing all parts of a resource with mechanisms similar to using the fragment identifier ("#") of a URI [35, Sec. 3.5]. For example, one could link to a figure shown on the sixth slide of a presentation file. Since the *deep link* is itself a URI, further statements, e.g. using RDF, can be made.

In ATIC, native data is accessible using so-called **Native Docks (ND)**, i.e. a ND for the file system, one for the email store, etc. (see bottom left of the figure). All resources get an entry in the **Native Registry** which serves as a mediator between the Shadow KG and the Native Docks. Native data can be accessed directly but using virtual triples is conceivable as well. With the latter, a single query, for example in SPARQL [279, 344], could address both, the Core as well as the Shadow KG. The first ATIC prototype was developed as part of the cSpaces prototype using *Java*-internal methods as well as JSON-RPC [194, 195] for communication.

In addition to the KG and native data, there are also **sub-stores** for the context spaces (right-hand side, highlighted in blue), data mining results, additional structures for user and access management and further use case-specific stores (e.g. for recommendation, research experiments, etc.). Last, there is a store for gathered evidence snippets from user activity tracking (bottom right, red highlighting). These evidences come into system using an *Activity Event Bus Infrastructure (AEBI)* described in more detail in Section 6.1.1.

One last aspect worth mentioning here is **one of the major design decisions in cSpaces**: every folder (or more generally: every collection-like entity), whether native or non-native, is

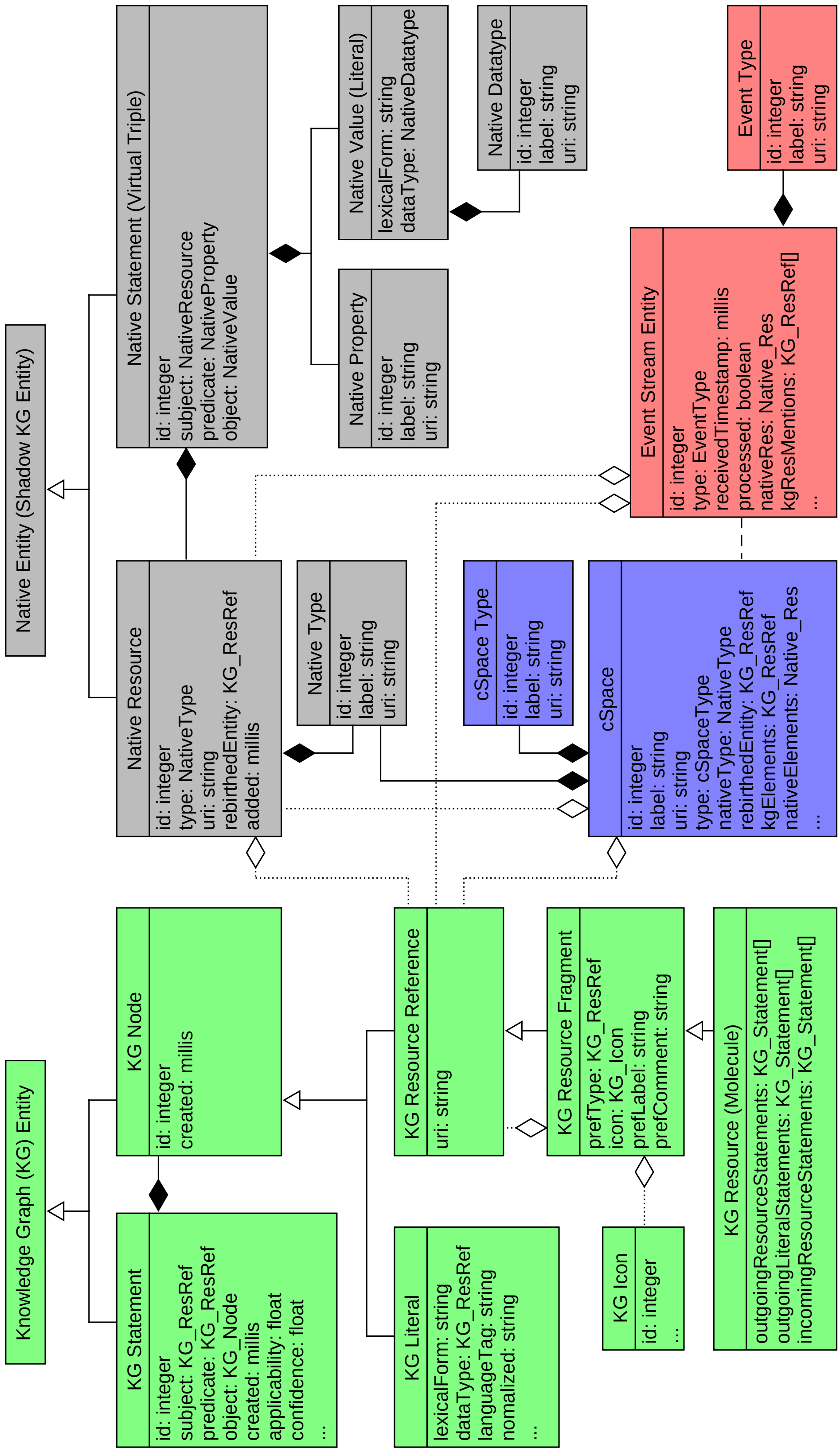


Figure 5.3: ATIC Knowledge Graph Store: Class Diagram.
(Note: The depicted classes only show the most important fields, methods and minor aspects were omitted. Colors deliberately match those of the corresponding architecture diagram in Figure 5.2.)

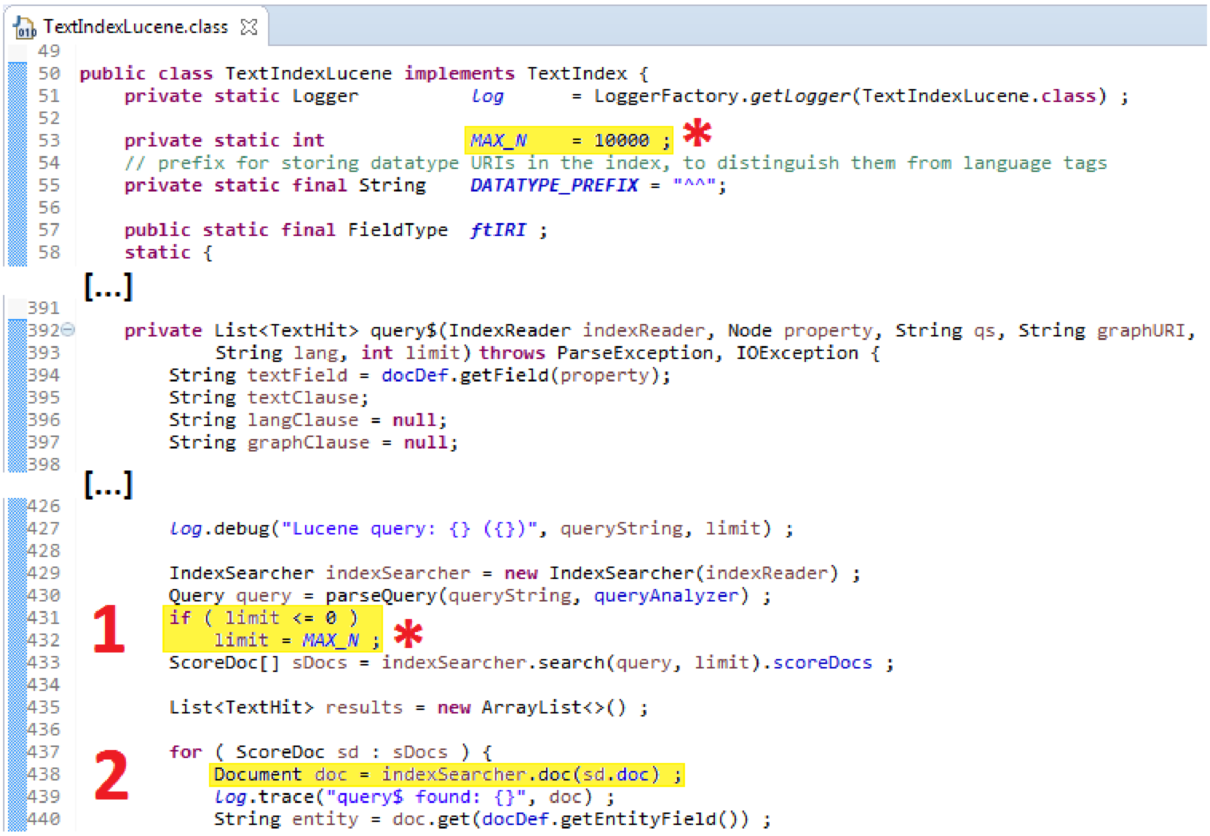
basically a cSpace. When cSpaces are rebirthed, they additionally get a KG ID and a flavor (see Sec. 3.3.3). Otherwise, they can be seen as a simple folder. Their native type shows where they originated from, e.g. the file system, the mail client, the web browser, etc. cSpaces may also be non-native, which means that they did not originate from a native system but were created within the cSpaces application itself, e.g. a users creates a new cSpace in the sidebar.

5.2.3 Efficiently Searching ATIC Knowledge Graphs

Two important features on a desktop are navigation and search. cSpaces already offers a new and unique way for the former, this section presents a solution for the latter.

Combining Structured and Indexed Search. Several triple stores, for example *Jena* [Sw9], utilize *Lucene* [Sw12] to realize Full-text Search (FTS) on graph literals. This raises the problem of separated indices: FTS hits point to nodes in the graph, which have to be further processed – or vice versa: the number of literals to be searched needs to be restricted by giving graph information beforehand. Depending on the size of the KG and the number of results, thousands or even millions of IDs possibly need to be transferred between stores. One solution observed in practice is to cut-off FTS result lists at a certain point, e.g. after 10,000 results as shown in Figure 5.4 (first highlighted section). But if further processing with graph information is needed for ranking, this solution will wrongly eliminate relevant results. Every solution combining heterogeneous indices will have to tackle this problem. From the author's experience, the problem would presumably be solved best using a single index structure covering all stores. However, this is an expensive endeavor since typically decades of experience went into the development and especially optimization of relational databases, triple stores, search engines, etc. Other solutions like data duplication may lead to synchronization or performance problems, e.g. Lucene only operates with adding or removing documents: with a KG changing potentially multiple times per minute or even second, Lucene's adding and removing overhead would be considerable. ATIC currently follows the **approach** of combining an RDB-based triple store (in this case *SQLite* [166], primarily to allow for lightweight deployment) and the FTS engine *Lucene* [Sw12] with the additional idea to hold an external mapping of Lucene document IDs to KG IDs as proposed by Minack et al. [257]. The problem with this solution is that Lucene document IDs are "ephemeral" and "clients should thus not rely on a given document having the same number between sessions" [12]. Lucene documents are numbered by non-negative integers: every newly added documents is assigned the next higher number. Upon removal, documents are flagged to be deleted (some documentations speak of "tomb-" or "headstones"), whereas their number is not reassigned. To the author's best knowledge, only merge operations (triggered according to the configured merge strategy) lead to such reassignments. ATIC search thus applies the following strategy:

- In all Lucene queries, the **maximum limit** of $2^{31} - 1$ documents is used.
- ATIC **listens to internal Lucene loggers** to be informed about merge operations. If a merge occurs, the **external ID cache needs to be updated**. To reduce these costs, the Lucene index is split up into n smaller indices. Using Multi-index readers, the internal numbers of later indices are simply added to the former ones as if it was one large




```
public class TextIndexLucene implements TextIndex {
    private static Logger          log      = LoggerFactory.getLogger(TextIndexLucene.class) ;

    private static int             MAX_N      = 10000 ; *
    // prefix for storing datatype URIs in the index, to distinguish them from language tags
    private static final String    DATATYPE_PREFIX = "^^";

    public static final FieldType  ftIRI ;
    static {
[...]

    private List<TextHit> query$(IndexReader indexReader, Node property, String qs, String graphURI,
            String lang, int limit) throws ParseException, IOException {
        String textField = docDef.getField(property);
        String textClause;
        String langClause = null;
        String graphClause = null;
[...]
        log.debug("Lucene query: {} ({})", queryString, limit) ;

        IndexSearcher indexSearcher = new IndexSearcher(indexReader) ;
        Query query = parseQuery(queryString, queryAnalyzer) ;
        if ( limit <= 0 )
            limit = MAX_N ; *
        ScoreDoc[] sDocs = indexSearcher.search(query, limit).scoreDocs ;

        List<TextHit> results = new ArrayList<>() ;

        for ( ScoreDoc sd : sDocs ) {
            Document doc = indexSearcher.doc(sd.doc) ;
            log.trace("query$ found: {}", doc) ;
            String entity = doc.get(docDef.getEntityField()) ;
```

Figure 5.4: Full-text Search in Apache Jena [Sw9].

Highlightings: *1: either limit the results (potentially misses relevant hits if graph information is needed for ranking) or return "everything" (i.e. potentially a huge amount of results) / 2: reading Lucene documents from harddisk typically slows down retrieval by orders of magnitude.*

(Source: Apache Jena SPARQL Text Search V3.6.0 [Sw10], TextIndexLucene class.

Note: In later versions, including the most recent V4.5.0 [Sw11], the highlighted sections can still be found but are spread across several methods. Also note that the implementation is fine, the screenshot is meant to illustrate a general problem when combining stores using the well-known Jena as an example.)

index. This has the disadvantage of needing a bit more memory due to increased index overhead, which can be neglected as long as the indices do not get too large. Since ATIC in cSpaces is a research prototype for a person's desktop, this should still be acceptable for the scope of this thesis. In large-scale industry applications, improved strategies are presumably necessary. The advantage is however that if merges occur, only $\frac{1}{n}$ of the ID cache needs to be updated. Further optimization is possible, up to the point of not re-reading document IDs at all after merge (since the merge outcome can be fully predicted, e.g. by applying the same strategy in the cache itself). However, since merge behavior may change in every new version of Lucene and large-scale use cases presumably need a different strategy anyway, no further optimizations have been performed as part of this PhD project, yet.

For readers interested in further details, Listing 5.1 provides an example of querying a

Lucene index with 3.8 million documents with and without the external cache. Note that these numbers are meant to give a better impression of the different size ratios (i.e. relative comparisons within the given scenario) without serving as an actual benchmark. The used query is chosen to generate considerably more effort than typical queries in cSpaces scenarios. Several optimizations that would have been done for productive use like multi-threading for cache (re-)building, more RAM, faster CPU and harddisk, etc. were not considered in this demonstration.

- ATIC avoids reading the actual Lucene documents from harddisk (as opposed to the depicted case in Figure 5.4, second highlighting), and instead **only uses the actual index structure**, which keeps query times considerably lower.
- Additionally, since ranking is currently mainly done using graph information, it is enough to have Lucene return a **bit vector** stating which KG nodes have been hit by the query (via their literals).

In CoMem's scenario (see next paragraph), further measures like checking users' access rights to these resources can thus be done in milliseconds using binary AND-operations on these vectors.

```
Single-index:
- 3,818,886 docs
- query: "a* OR h* OR m* OR n* OR s*"
- (Re)Building caches... building whole cache took 166,482 ms ==> 166 s/cache
- Querying... Found 1,441,968 hits.
- search time (w/o cache): 4736 ms for core lucene, 57,808 ms for ID lookup
- search time (w/  cache): 4736 ms for core lucene,     30 ms for ID lookup

Multi-index of 10:
- 3,818,886 docs
- query: "a* OR h* OR m* OR n* OR s*"
- (Re)Building caches... building whole cache took 181,357 ms ==> 18 s/cache
- Querying... found 1,441,968 hits.
- search time (w/o cache): 6701 ms for core lucene, 66,266 ms for ID lookup
- search time (w/  cache): 6701 ms for core lucene,     41 ms for ID lookup
```

Listing 5.1: Querying Single- and Multi-index Lucene with and without External ID Cache. Used hardware: Intel Core i5-661 3.33 GHz, 4 GiB RAM, Windows 7 32-bit [Sw83].

Integration into CoMem. A version of this search (only including some of the ATIC characteristics) has been integrated into CoMem (mostly as back-end code, the front-end has been developed by Sven Schwarz). Using the aforementioned vectors, connections in the graph can be represented, too. This is similar to the approach of *RDF molecules*, for example proposed in Fernández et al. [114]. Facets for further drill-down can thus be calculated dynamically in a few seconds. Figure 5.5 illustrates this using an example: Searching with the keyword

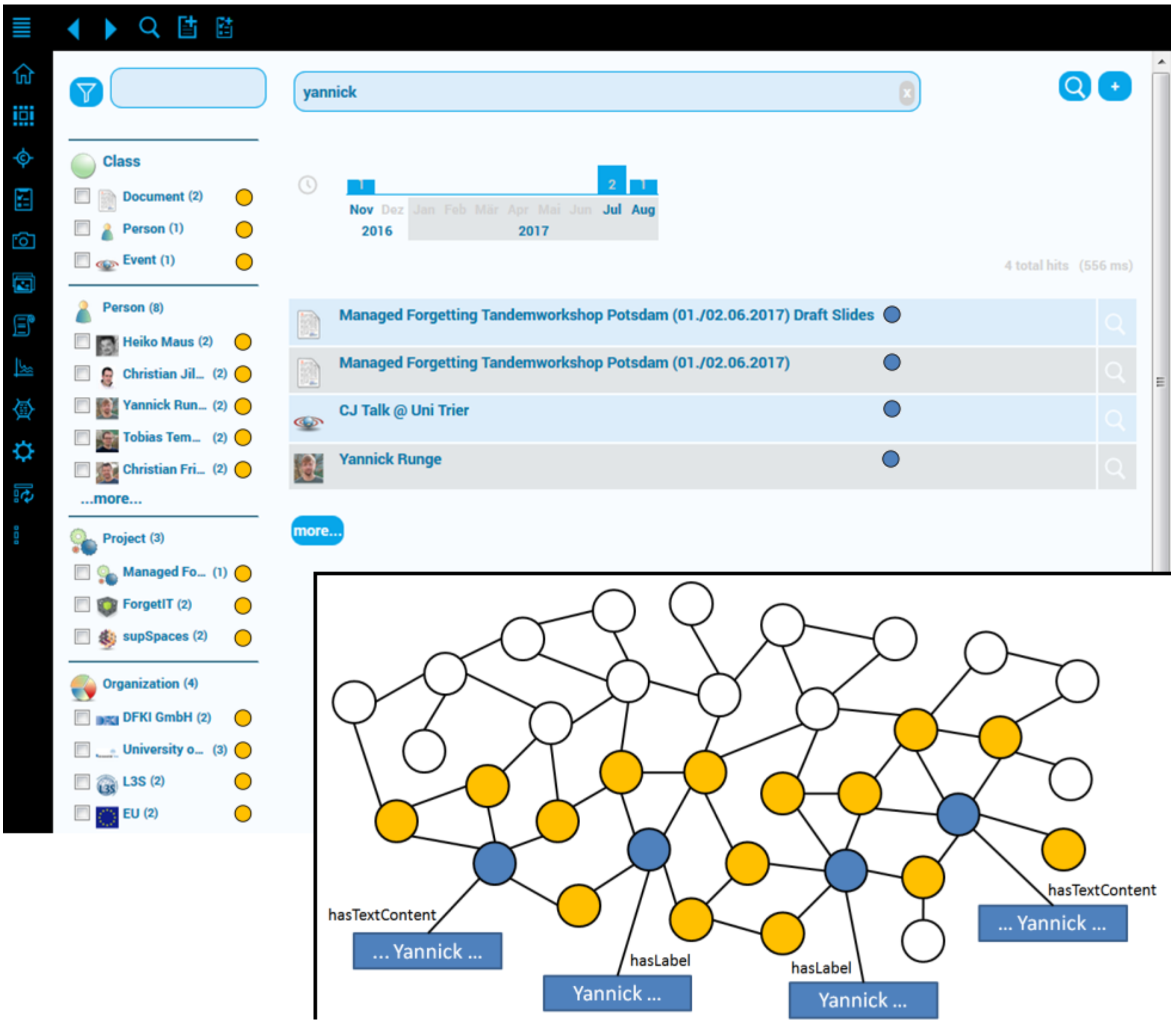


Figure 5.5: CoMem Search: Combining Structured and Indexed Search.
Color highlighting: *blue nodes: search results obtained by keyword search / yellow nodes: neighboring nodes in the knowledge graph used for faceted search and drill-down.*
(Note: A variant of this figure has already been published in Jilek et al. [182].)

"yannick" resulted in four hits (highlighted in blue). Their neighborhood in the graph, e.g. each thing's type and relations to other things (in yellow), is obtained by combining their pre-calculated vectors. Facets are then shown (on the left) allowing for further drill-down by either class/type or via connected things like persons, projects, etc.

Working on this search started in the *supSpaces* project [370] (see Sec. 9.2.1) and has meanwhile been in productive use in CoMem 24/7 for about six years (see Sec. 9.1).

Additionally, in Hertling et al. [163], top-k shortest paths potentially useful for semantic search have been explored, but nothing has been integrated into any of the prototypes, yet.

After introducing the context model of cSpaces as well as its knowledge representation and storage solution in the form of ATIC knowledge graphs, this chapter's last section addresses the bootstrapping of context spaces to "jump-start" the system.

5.3 Bootstrapping Context Spaces

The Semantic Desktop or personal knowledge assistants in general suffer from the so-called *cold start problem* (e.g. [96]), i.e. a system being "empty" in the beginning since no information about the user, interaction data etc. are available to draw conclusions from and perform support measures accordingly. To overcome or at least reduce this problem, this section presents a three-step approach to bootstrap context spaces (addressing research question RQ5c). Each step has its own tool: the *PIM Crawler* (5.3.1), the *Semantifier* (5.3.2) and the *Contextifier* (5.3.3), which are discussed in the following.

Note: These bootstrapping algorithms can (theoretically) be fed with an arbitrary amount of content, which can typically add up to hundreds of thousands of items, possibly millions. The literature speaks of "Personal Big Data" [146] or "Big Personal Data" [351] in this regard. Since all presented tools in this section are not heavily optimized, this thesis only hints at the (theoretical) possibility of processing large amounts of items by using the bracketed term of "(Big) Personal Data". Given a certain, larger amount of data, further optimizations are presumably necessary for the tools to keep working.

5.3.1 PIM Crawler – Crawling (Big) Personal Data

PIM Crawler (5.3.1) collects "PIM data", i.e. information items like files, mails, bookmarks and calendar entries on a user's computer. Files are read directly from harddisk. For emails, the *MBOX* format [153] as well as the *Internet Message Access Protocol (IMAP)* [71, 248] are supported. Regarding bookmarks, parsing the data store of *Mozilla Firefox* [Sw87], called `places.sqlite` [262], is possible. Calendars in the *iCalendar* format [78, 89] are supported and parsed using *iCal4j* [Sw76].

PIM Crawler was initially created to just crawl "a bit of data" for some experiments without having a comprehensive crawling toolkit in mind. Over the years, more and more student and research projects made use of it (e.g. [274, 64, 159]), so spending additional effort to support more formats would be justified (future work).

The tool has been largely developed by Markus Schröder, while the author of this thesis added minor aspects like an option to treat archives like a folder, so their content is accessible as well. This feature is realized using *7-Zip* [Sw29, Sw30].

PIM Crawler can output RDF or an RDB, in particular *SQLite* [166]. Its database schema is depicted in Figure 5.6.

Related approaches are *Aperture* [Sw24] or *Apache Tika* [Sw19, Sw21]. And just like these approaches, PIM Crawler creates a 1:1 representation of the items on the user's computer. This is as intended but does not go far enough for cSpaces and CoMem scenarios: the next step is the extraction of higher-level concepts performed with the *Semantifier* as explained in the next section.

5.3.2 Semantifier – Semantic Leveraging of (Big) Personal Data

The PIM Crawler presented in the last section creates a 1:1 representation of the items on a user's computer. **Higher-level concepts**, for example found in file names, email subjects or, in general, in the content body of information items, are not extracted. As shown in

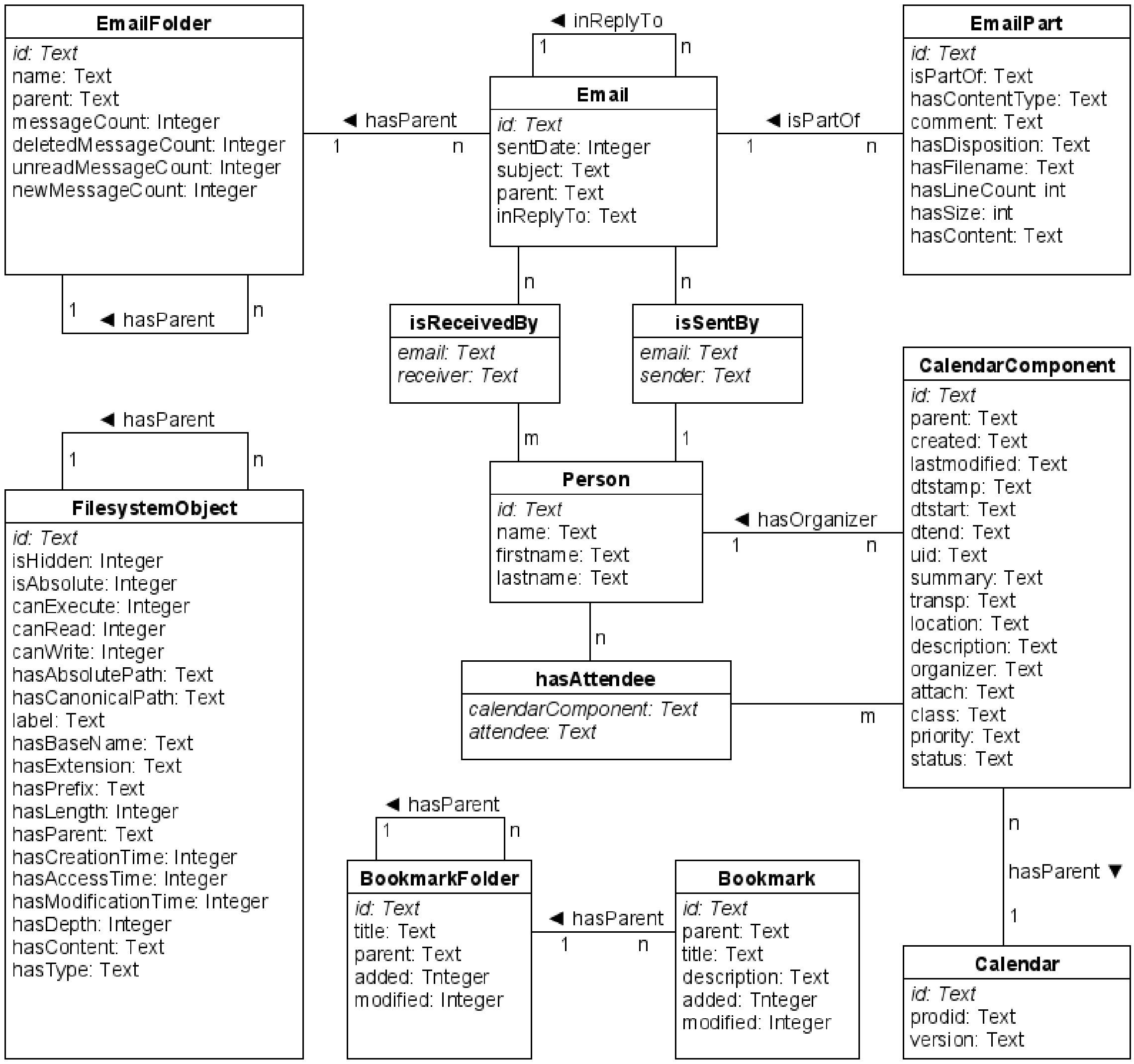
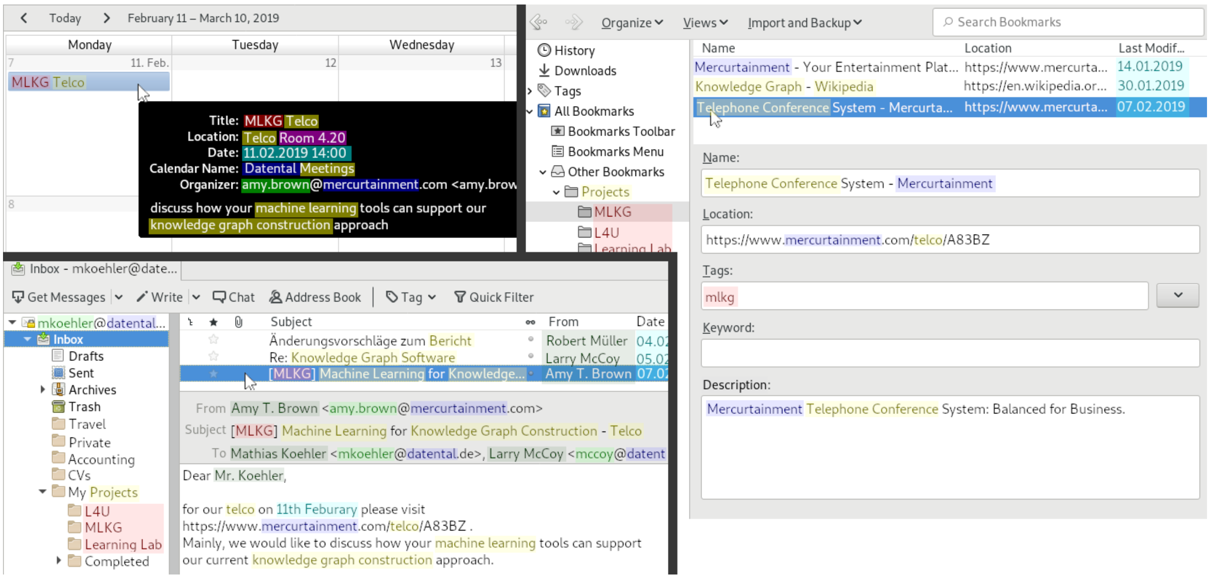


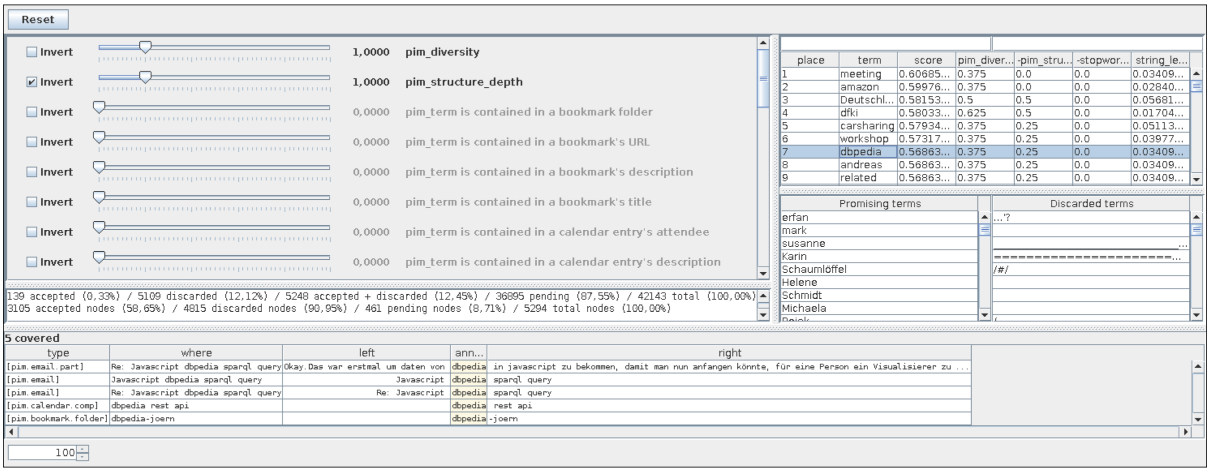
Figure 5.6: PIM Crawler: Database Schema.
Using data types of *SQLite* [166], and primary keys are written in italics.

Figure 5.7a, these concepts can be projects (highlighted in red), persons (green), organizations (blue), times (cyan), places (purple), general topics (yellow), etc. In Jilek et al. [184], the author experienced better performance of a semantic system if such additional vocabulary was utilized. Sparing these concepts is thus not advisable. On the other hand, accepting irrelevant concepts may clutter the KG and lead to non-helpful relations between items. Identifying the "right" concepts cannot be fully automated since each user has a subjective view on their data [88, 28]. The Semantifier therefore follows a human-in-the-loop approach.

The system comes up with an initial suggestion based on various heuristics. One of the basic ideas is to exploit the more structured parts of PIM data, e.g. email or calendar fields, to enrich rather unstructured texts (e.g. folder names). This follows the assumption that particular text fields contain particular named entities more likely and with a similar format,



(a) Concepts to be found in the calendar (top left), email client (bottom left) and bookmarks (right). Color highlightings: *red: projects / green: persons / blue: organizations / cyan: times / purple: places / yellow: general topcis*.



(b) GUI to browse and filter concept candidates.

Figure 5.7: Semantifier: Concepts to Be Found and GUI.
(Note: These are screenshots by Markus Schröder published in Schröder et al. [325].)

e.g. first and last names in an email address or a company name in its domain part. While the title of a website is typically given by the website's author, folder names or the titles of calendar entries are typically given by the users themselves. Especially the latter are a special kind of summary. They may thus more likely contain the user's own vocabulary like technical terms, specific abbreviations or acronyms. Additional heuristics are based on spatial distance in folder hierarchies or temporal relationships. Once a new concept is found, similar occurrences can be discovered in other items using NER (see Sec. 6.2.2).

Finally, the identified concepts are presented to the user as shown in Figure 5.7b. Since terms come in various shapes and facets, sorting concepts by a single metric is not sufficient. Instead, the Semantifier offers different perspectives on the candidate list by calculating a weighted harmonic mean score of appropriate combinations of multiple metrics. To avoid users being overwhelmed with the high variety of settings (see sliders on the left-hand side of the figure), meaningful presets obtained in user studies can be selected.

The above statements are a brief summary of a more detailed introduction of the Semantifier available in Schröder et al. [325]. Work on the Semantifier was led and mainly conducted by Markus Schröder, whereas the author mainly contributed conceptual ideas and experience from previous DFKI projects like *supSpaces* [370], in which concepts had to be found using alternative (term extraction) approaches, for example available in the *JATE* toolkit [Sw125].

5.3.3 Contextifier – Retrospective Context Mining on (Big) Personal Data

In a third and last bootstrapping step, the *Contextifier* tries to identify user contexts spread across the different data systems. For example, a file could have been the attachment of an email, things stored in a certain folder all belong to a specific meeting mentioned in the calendar, a bookmark in the web browser originated form a an email the user has received, etc.

Retrospective Context Mining. Several related approaches do context/task mining on-the-fly on a running system (see Sec. 6.3). Some of these approaches, like Costache et al. [69], also take past data into account making them hybrid approaches. To the author's best knowledge, however, there is no purely retrospective context mining approach – which is the gap, the Contextifier tries to fill. A reason for this could be the difficulty of the problem. Using a catchy phrase, one could say that retrospective context mining is not just data mining but rather a kind of "data archaeology": From possibly millions of interactions with a file only a few timestamps remain: the creation, the last modification and the last access date. Some of them may even be missing depending on the file system. All information about the interaction between creation and last modification or access, respectively, is usually lost. This is just one example to illustrate that trying to retrospectively reconstruct what is or may have been a particular user context is very speculative. Therefore a human-in-the-loop approach is followed.

Approach. First exploratory work has been done by Jessica Chwalek as part of a student research project in combination with a seminar paper [64]. Based on these findings, a bachelor thesis had been advertised by the author, which was accepted and conducted by Desiree Heim [159]. The rest of this section briefly presents the outcome of this work.

The Contextifier tries to **identify relations between information items**, for example those mentioned in the beginning, e.g. a file being the attachment of an email. This example already illustrates one of the basic trade-offs to be made: In order to determine whether a file on the harddisk is identical to the attachment of an email, different amounts of effort can be spend, whereas results typically become more certain with each additional amount of effort spent. Escalation steps could, for example, be as follows:

- comparison of file sizes: not much effort (integer comparison) but high uncertainty since very different files could have the same size

- comparison of file names: also not much effort (string comparison) but similar uncertainty since a file could have been renamed,
- comparison of (extracted) file content: more effort but also more certainty and possibly still less effort than the next item (especially if the content has already been extracted previously),
- binary comparison of files: highest effort but an absolutely certain result.

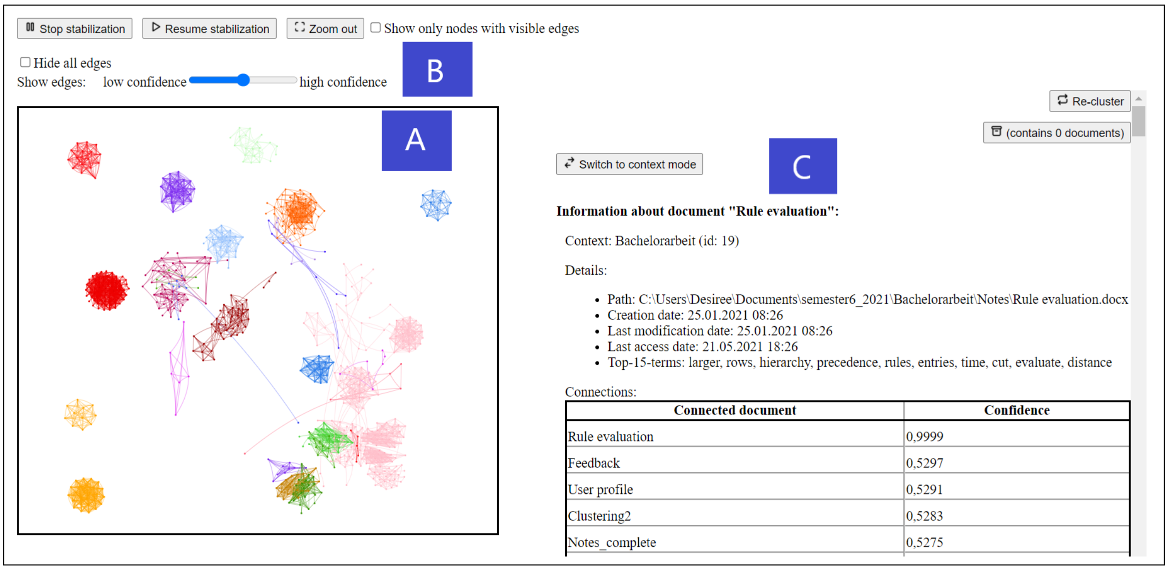
Since Contextifier is an interactive tool, more effort also means longer waiting times for users, which is only tolerable to a certain extent. Contextifier compares items of the same type or items of different types as in the example above. Every item pair is thus assigned several scores, one for each relationship indicator (i.e. basically a set of rules or heuristics). The problem of identifying contexts can then be modeled as a **multi-graph clustering problem** having the information items as nodes and the different comparison results as (multiple) weighted edges between them.

Again, a **human-in-the-loop approach** is followed: users can give feedback in the beginning by selecting interesting context candidates upfront or by giving feedback on the Contextifier's suggestions. Results can be viewed using a graph view or a tree view as depicted in Figure 5.8. Users may delete, rename or split a context, merge two contexts or move an item from one context to another. Figure A.1 (appendix) shows another GUI to adjust rule weights individually or declare some rules to be particularly meaningful, which basically boosts their score. Additionally, further examples of rules and heuristics can be seen in the figure.

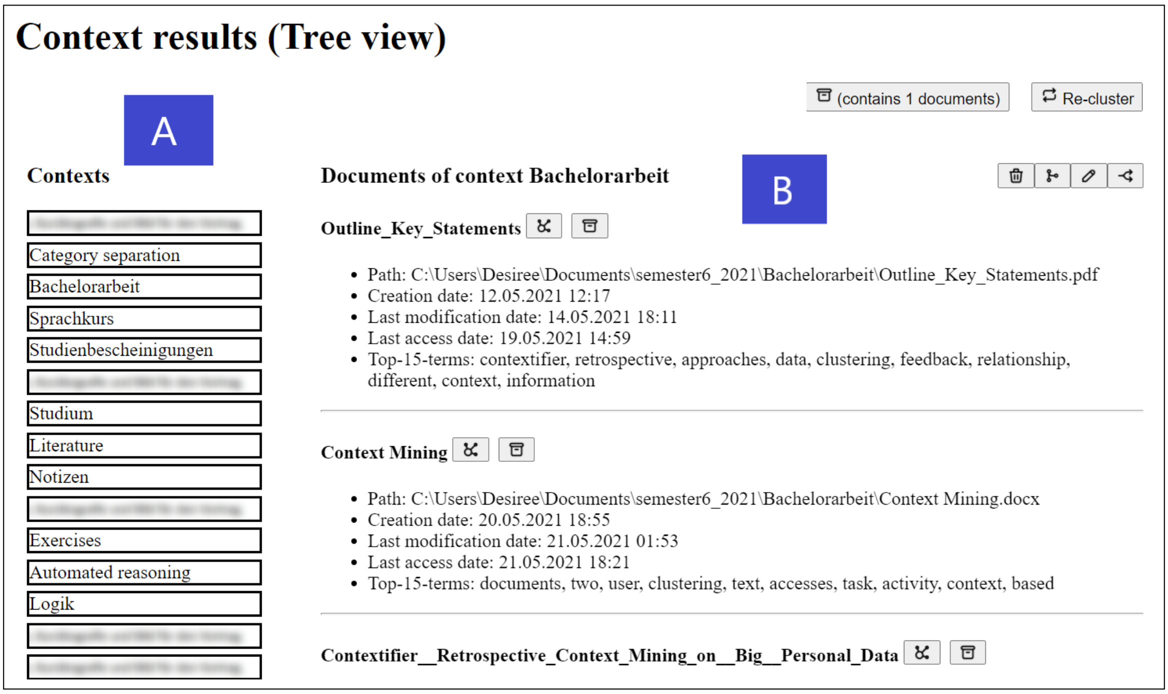
Evaluation. The approach has been evaluated in a short-term user study with 14 participants that could test the app with their own data and answered a questionnaire afterwards. Six of the participants were members of DFKI SDS (one female, five male, in their mid-twenties to early fifties), and the others were students (three female, five male, between 21 and 27 years old, mostly with a computer science background). The first part of the questionnaire addressed **usability** with a selection of questions of the *Computer System Usability Questionnaire (CSUQ)* [223] (business-related questions were omitted). Except for clear error messages (Q_{C6}) and recovery from mistakes (Q_{C7}), all items are mostly on the positive side of the rating spectrum. These two more negatively assessed categories can be explained by technical problems of Contextifier: There were performance and scalability issues that led to slow processing and unresponsiveness of the system. Some users ultimately had to reduce the number of items to be processed by the system. Detailed results on the system's usability can be found in Figure 5.9.

More importantly, Contextifier's **main functionality** was rated very positively – with some reductions regarding the generation of context labels (Q_{C17}) and ease of *finding* information (Q_{C22}). Especially items stating that documents in a context actually belong together (Q_{C16}), comprehensibility of relationship scores (Q_{C19}) and satisfaction with context views (Q_{C23}) and direct context manipulation functions (Q_{C24}) received high agreement among the participants. As with the CSUQ items, agreement or disagreement could be stated on a 7-point Likert scale [226]. Detailed results are depicted in Figure 5.10. A last part of the questionnaire asked whether participants would prefer the graph view or the tree view, which did not lead to a clear preference. Results were actually close to 50:50 as shown in Figure A.2 (appendix).

As mentioned before, some of the figures can be found in the appendix (Ch. A.2). Details about the Contextifier can be found in [160].



(a) Graph View.
Highlighted sections: *A: graph visualization / B: interaction possibilities / C: explanations*.



(b) Tree View.
Highlighted sections: *A: list of contexts / B: elements of selected context*.

Figure 5.8: Contextifier: Graph View and Tree View Realized in Heim [159].
(Image source: Heim [159].)

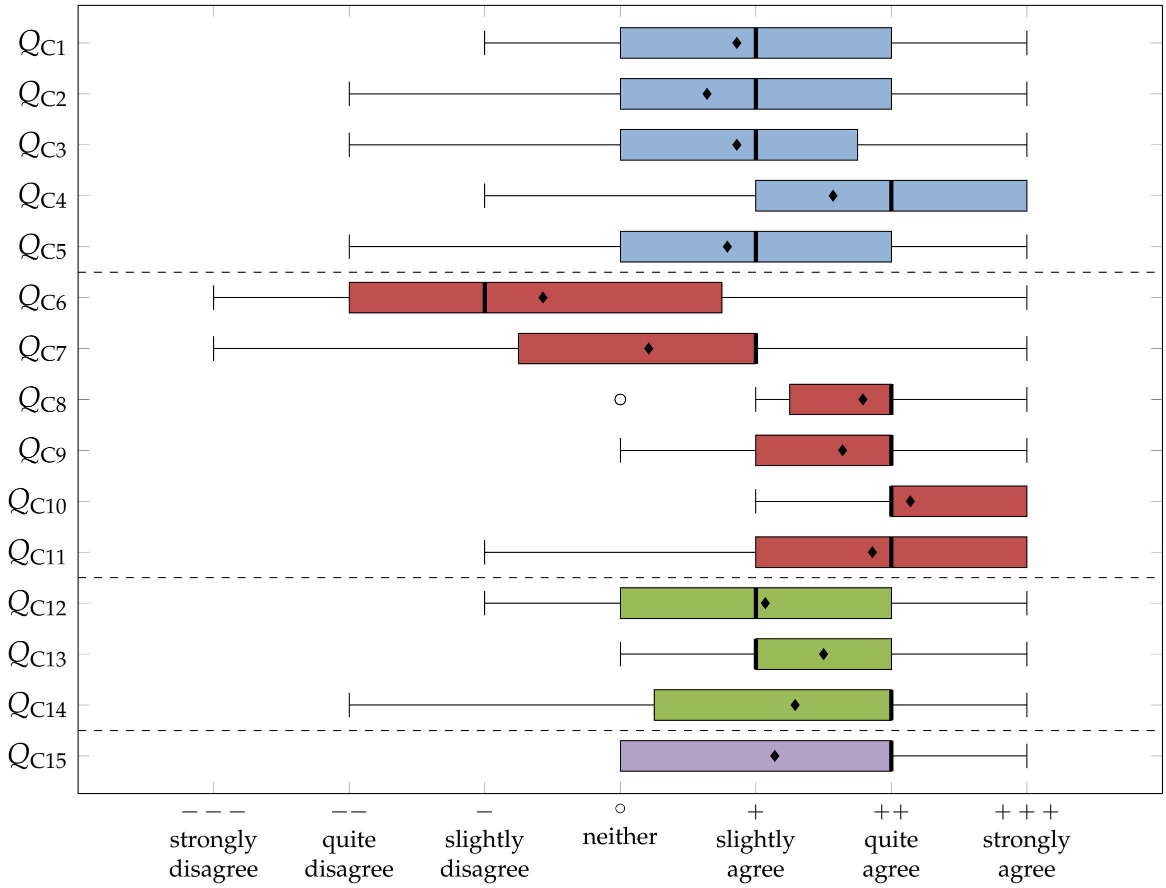


Figure 5.9: Contextifier: Survey Result for CSUQ Items [223] as Reported in Heim [159].

Q_{C1}–Q_{C5}: System Usefulness,
Q_{C6}–Q_{C11}: Information Quality,
Q_{C12}–Q_{C14}: Interface Quality,
Q_{C15}: Overall.

Q_{C1}: Overall, I am satisfied with how easy it is to use this system.
Q_{C2}: It is simple to use this system.
Q_{C3}: I feel comfortable using this system.
Q_{C4}: It was easy to learn to use this system.
Q_{C5}: I believe I became productive quickly using this system.
Q_{C6}: The system gives error messages that clearly tell me how to fix problems.
Q_{C7}: Whenever I make a mistake using the system, I recover easily and quickly.
Q_{C8}: The information (such as on-line help, on-screen messages and other documentation) provided with this system is clear.
Q_{C9}: It is easy to find the information I need.
Q_{C10}: The information provided with the system is easy to understand.
Q_{C11}: The organization of information on the system screens is clear.
Q_{C12}: The interface of this system is pleasant.
Q_{C13}: I like using the interface of this system.
Q_{C14}: This system has all the functions and capabilities I expect it to have.
Q_{C15}: Overall, I am satisfied with this system.

(n=14; questions from CSUQ [223]. details on box plot semantics in Sec. B.2)

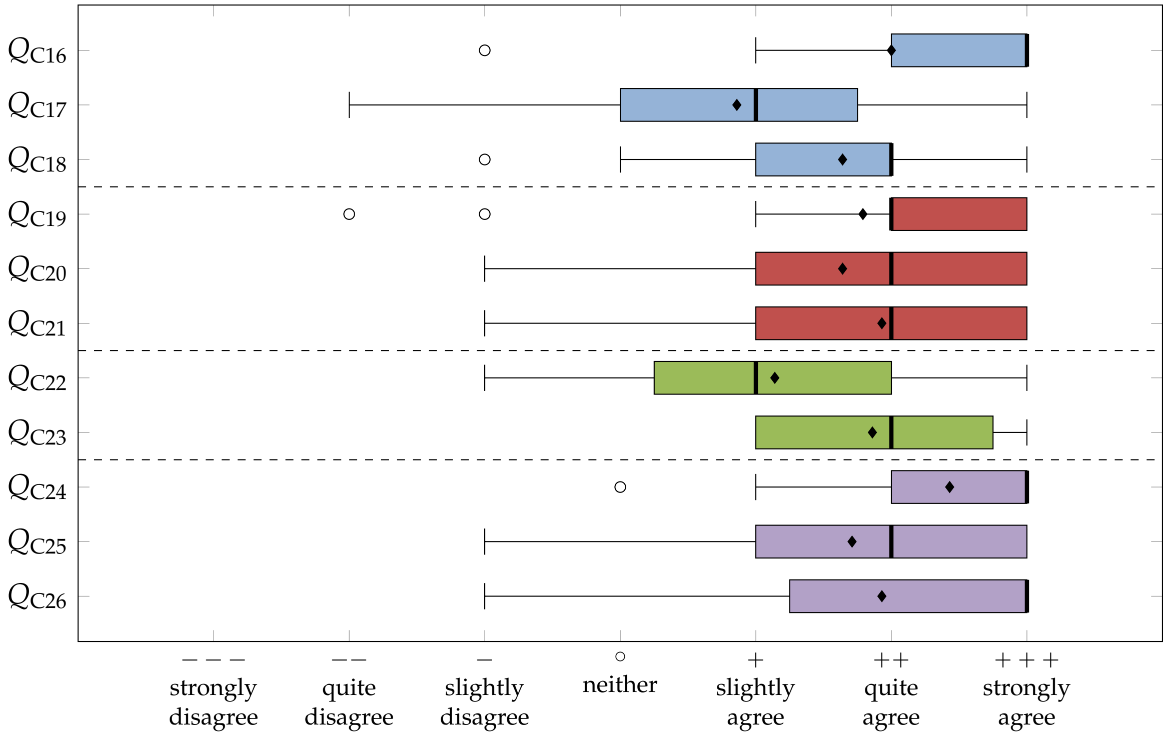


Figure 5.10: Contextifier: Survey Result for Custom Items as Reported in Heim [159].
Q_{C16}–Q_{C18}: Quality of Initial Results,
Q_{C19}–Q_{C21}: Transparency of the Calculation Process,
Q_{C22}–Q_{C23}: Ease of Finding Specific Information,
Q_{C24}–Q_{C26}: Interaction Possibilities.

Q_{C16}: I think that documents in a context belong to each other.
Q_{C17}: In my opinion, the context labels are appropriate.
Q_{C18}: The calculated relationship confidences seem reasonable to me.
Q_{C19}: I can comprehend why a document has a certain relationship confidence to another document.
Q_{C20}: I can comprehend why documents are in a certain context.
Q_{C21}: I can comprehend how the change of rule weights or properly important documents reflects on the clustering.
Q_{C22}: It is easy to find the information I am interested in.
Q_{C23}: The different views (i.e. the context tree view, the context graph view, and the old/new cluster graph view) provide all the information about the contexts, documents, and relationships I wished to have.
Q_{C24}: The direct context operations (rename, split, merge, delete, move documents from one context to another using the tray) work as expected.
Q_{C25}: The indirect context operation, i.e. re-clustering with changed rule weights or changed particularly meaningful rules, works as expected.
Q_{C26}: The available context operations offer all the modification possibilities I wish to have.

(n=14; details on box plot semantics in Sec. B.2;
Note: This figure has already been published in Heim et al. [160].)

5.3.4 Excursus: More General Work on Knowledge Graph Construction

This PhD project's most prominent side issue was contributing to research on personal or corporate *Knowledge Graph Construction (KGC)* in general. The presented work was led and mainly conducted by the author's colleague, Markus Schröder, as part of his PhD project [323]. All publications mentioned in this excursus section except for the *Semantifier* [325] are not considered an explicit part of this PhD thesis as declared in the preface (see list of side publications).

The collaborative work with Markus Schröder can roughly be summarized in three clusters: 1) KGC from spreadsheets, 2) KGC from PIM data, especially files and 3) KGC evaluation and tooling.

KGC from Spreadsheets. Real-world spreadsheets can become very messy: Some of the problems one can face are multiple surface forms, different style usage (e.g. crossed out statements), mixed date representations, multiple entries in a cell, multiple types in a table, etc.

A first approach to construct KGs from such messy spreadsheets was an interactive one [334]. It was very expressive but using the GUI was rather time consuming and there was no formalism.

A second approach [328] utilized mapping rules with the *RDF Mapping Language (RML)* [92]. RML so far had not supported *Microsoft Excel* [Sw79], which was added by the approach. Typically only CSV files were supported, but as illustrated by the problems stated before, a lot of (meta-)information can be lost if a spreadsheet is "down-converted" to a CSV file. Although this second approach was an improvement compared to the first one, creating the mapping was still time consuming.

Thus, a third approach [329] added mapping prediction using rules and heuristics and especially addressing aforementioned messiness. Prediction performance still needs further improvement though.

While all three approaches are primarily aimed for knowledge engineers, an RDF Spreadsheet Editor for non-technical users was presented in Schröder et al. [332, 333]. Since spreadsheets are well-known by many knowledge workers, they quickly feel familiar with such an editor and can thus fully focus on the actual modeling. Typically each row contains an entity, while properties are stated in the columns. A similar scenario (in Excel, not the interactive editor) is depicted in Figure 8.1.

KGC from PIM Data, especially Files. With regard to PIM data, earlier sections already introduced the idea of *deep-linking desktop resources* [324] (Sec. 5.2.2) and the *Semantifier* [325] (5.3.2) for concept mining. In Schröder et al. [330], a third PIM-related approach is presented, in which KGs are constructed from file names in a semi-automatic way with humans-in-the-loop. The approach consists of four major stages: domain term extraction, ontology population, taxonomic and non-taxonomic relation learning. An evaluation was conducted with four experts of enviaM [109] showing that file systems are promising sources for KGC and using the presented tool already yielded useful personal KGs with moderate effort spent. Additionally, a tendency that helpful statements can be automatically suggested could be observed.

KGC Evaluation and Tooling. With regard to KGC evaluation and tooling further work has been conducted:

- Schröder et al. [327] introduced *Data Sprout*, a pattern language and dataset generator for the evaluation of KGC from spreadsheets.
- Another generator was provided in Schröder et al. [335] with a special focus on the extraction of person names.
- Schröder et al. [336] presented *RDF2RDB-REST-API*, a solution for generating RDBs (and server code to access it) from RDF.
- A Linked data application framework to enable rapid prototyping was provided in Schröder et al. [326].
- Schröder et al. [331] presented an extension of the *deep-linking approach* for data scientists easing data understanding by basically linking data to its documentation.

All mentioned tools and publications are part of Markus Schröder's *Hephaistos Toolkit*, which is also documented online[1] with additional images, videos, demos, etc.

[1] https://www.dfki.uni-kl.de/~mschroeder/hephaistos/

6

Context Elicitation

This chapter addresses steps 1 to 3 of the cSpaces interaction cycle: evidence collection (Sec. 6.1), information extraction (6.2) and context elicitation (6.3). As in the previous chapter, a balance had to be found between ever-growing, continuously evolving knowledge graphs potentially being updated in high frequency and real-time capable analysis and querying to ensure responsiveness of the assistance.

Note: Parts of this chapter have already been published [182, 185–187, 240, 244].

6.1 Evidence Collection by User Activity Tracking

Semantic Desktop [80, 304] or similar approaches typically need a form of user activity tracking: To provide appropriate user support, the system has to "know" what the user is currently doing. Tracking is closely related to context/task detection and mining, which is addressed in the section after next (6.3).

Related Work and Existing Solutions at DFKI SDS. Some of the groups and projects mentioned in Section 4.1 developed their own tracking solutions (e.g. [283, 281, 5]), others used existing tools like *uLog* [Sw94] as in the case of the *SWELL* project [299]. Such solutions differ in aspects like the definition of "user context" (e.g. [341, 342, 282, 111, 311]), how they store captured information or the scope of what is tracked. An overview is provided in Schmidt et al. [320] or Schmidt [317, Table 6.2], for example. Members of DFKI SDS and research colleagues also created *UserContext* [Sw106], a tool collection in this regard. For reasons stated in previous chapters, it made sense for this PhD project to base solutions on CoMem and reuse and extend what is already available there. In the beginning of this PhD project, this mainly was the *User Observation Hub* [Sw108] that was primarily fed by *DragonTalk* [Sw107], both part of the aforementioned *UserContext* [Sw106].

However, in 2015, Mozilla announced to deprecate *XUL-based* [260] extensions [265], rendering several of these solutions as well as the related *FireTag* [Sw119] (mentioned in Sec. 2.3.3) useless in the near future without major overhauls. The announcement was one of several reasons, why the author invested a bit of time in re-engineering the existing environment. Further reasons as well as the outcome are discussed in the next section.

6.1.1 Plug-outs – Semantic Desktop Re-engineering

Previous sections have shown that the Semantic Desktop (SD) can be a useful ecosystem for IM and KW support. So far, **two categories of SD applications** could be observed: newly created semantic applications and plug-ins to enhance traditional, non-semantic ones [96]. For both categories, application development is rather costly: either a whole new application or a deep integration into an existing one needs to be developed – one may be worse than the other depending on the concrete case. Additionally, designing and implementing a corresponding GUI is needed in both cases. In this regard, software developers will likely agree that front-end development may be very different from back-end development – both may show very different facets of the same programming language. There is thus potential to reduce the costs of plug-in development for the SD while at the same time tap the potential of further synergies.

Together with Markus Schröder and Sven Schwarz, the author had the idea of realizing so-called **plug-outs**. They are *headless plug-ins* having very little functionality, often just the capability of *sending out* in-app events to a surrounding system like an event bus. Such basic functionality can often be realized with very few lines of code. Their corresponding GUI elements and logic are located in a general SD widget, for example a sidebar, where they can be easily reused. Plug-outs for various applications can thus share the same GUI for tagging, commenting, recommendation, etc. Depending on the currently browsed item, those general GUI elements can be complemented by item-specific ones like sender, recipient and attachment information for an email. Moreover, a general SD GUI like a sidebar is also a single interaction point for users. They do not need to learn a new (plugged-in) interface for each of their applications. Instead, they can just keep using them the usual way having only a sidebar as a new user interface to become familiar with. This reduces development and learning effort.

The scenario of **different plug-outs and a shared sidebar** is depicted in Figure 6.2. The upper screenshot (6.2a) illustrates a web browsing scenario, whereas the lower one (6.2b) shows browsing of an email. In the web browsing scenario, the sidebar shows the title and URL of the currently browsed website as well as widgets for tagging. When browsing an email, the panel showing website-specific content like title and URL is replaced with the email's subject, sender, recipient and attachment information. Tagging widgets are identical. Note: since the second screenshot is newer than the first one, it additionally shows a PID section (bottom right) developed later. In general, screenshots of cSpaces widgets throughout the thesis may all look slightly different since it has been an application under heavy development and tailored according to several experiments. Feel free to compare the scenario of multiple plug-outs with a single sidebar (Fig. 6.2) to the earlier one having a dedicated plug-in including a GUI for each application as shown in Figures 2.3 to 2.5.

Plug-outs can be seen as a net of software sensors [174] reporting user activities to a central entity. The author called this configuration **AEBI**, which stands for **Activity Event Bus Infrastructure**. AEBI is depicted in Figure 6.1. A sample evidence snippet can be found in Listing 6.1. The snippet signals that a user has visited DFKI's website given by the URL in line 5. The exact type of the event is a *document visibility change* (line 9), i.e. typically a browser tab regaining focus. As shown in line 7, the website's full content is also send to AEBI.

Since plug-out development is rather cheap, extending and thus increasing the density of the net is easier possible than in former times. With the additional transition from *XUL* [260] to the widely supported *WebExtensions* [261] (as hinted at in the previous section), the number of supported applications rose from *Firefox*, *Thunderbird* and the *Windows File Explorer* [Sw85] in the beginning of this PhD project to all major web browsers like *Firefox* [Sw87], *Chrome* [Sw46], *Edge* [Sw78], etc., *Thunderbird* [Sw88] and *Outlook* [Sw82] as well as *Windows File Explorer* [Sw85] and *Excel* [Sw79]. The development of some of these plug-outs was supported by Rudolf Koch (née Novik) and student research assistants. Rudolf Koch also supported the transition of plug-outs and sidebar components to CoMem taking over their development with respect to further enhancements for industry use (see *PIMO/CoMem Desktop Application* in Sec. 8.2).

In the *SuGraBo* project [369], colleagues of DFKI SDS and the author also explored **plug-in-free solutions based on accessibility interfaces** like *Microsoft Active Accessibility* [356] and its successor the *Windows Accessibility API* [253]. All components and UI elements of an application supporting this technology are typically available as a tree of elements to be browsed. Having the content of each UI widget directly available as text can be less error-prone than the alternative approach of on-screen *Optical Character Recognition (OCR)* [258]. However, both technologies have their pros and cons. (Hint for further reading: Led by Sven Hertling, an early prototype of a search engine for GUI elements has been enhanced during the SuGraBo project [164].) Bringing AEBI and *SuGraBo* tools together is an envisioned idea.

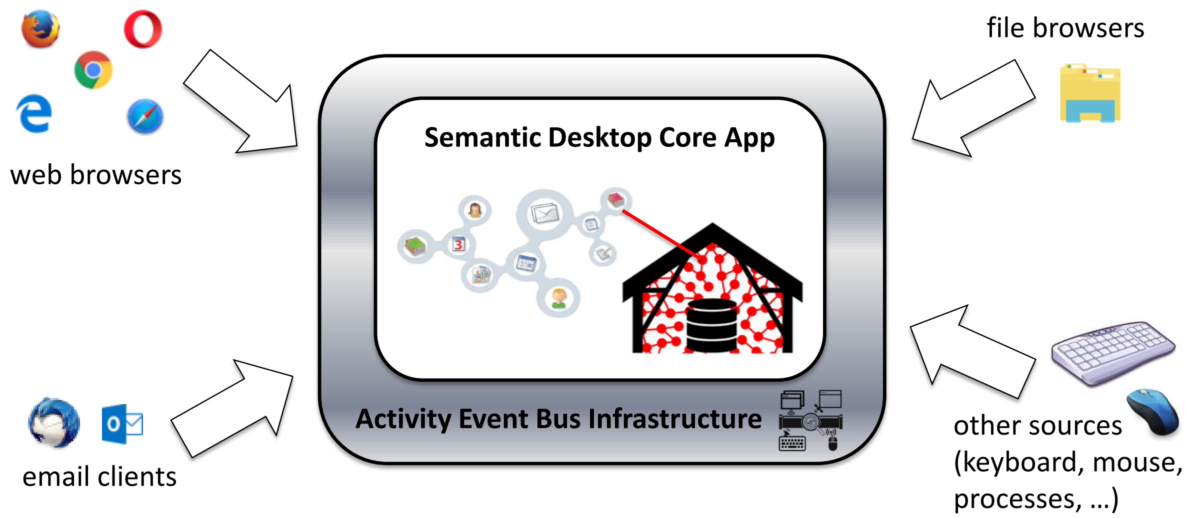
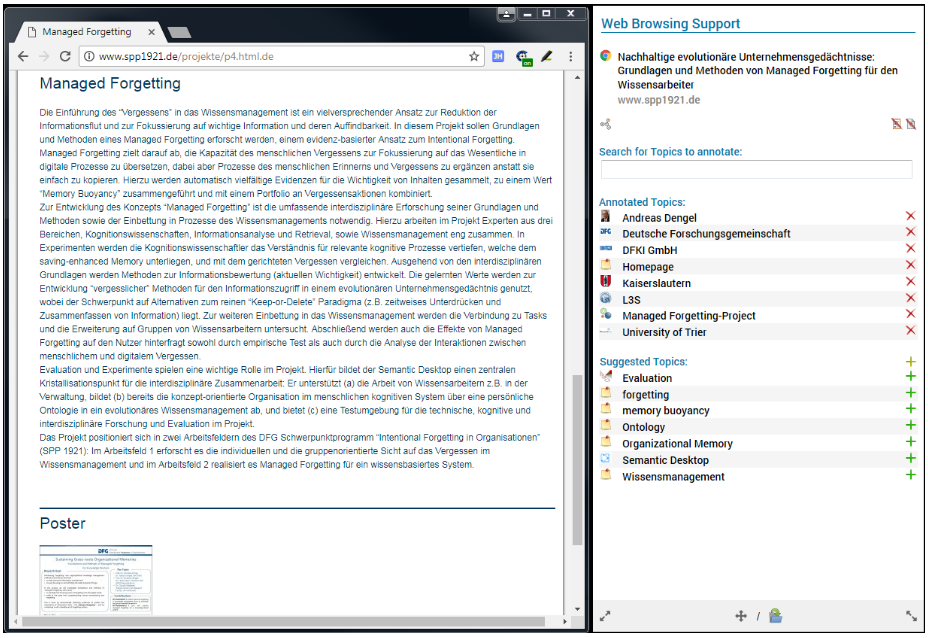


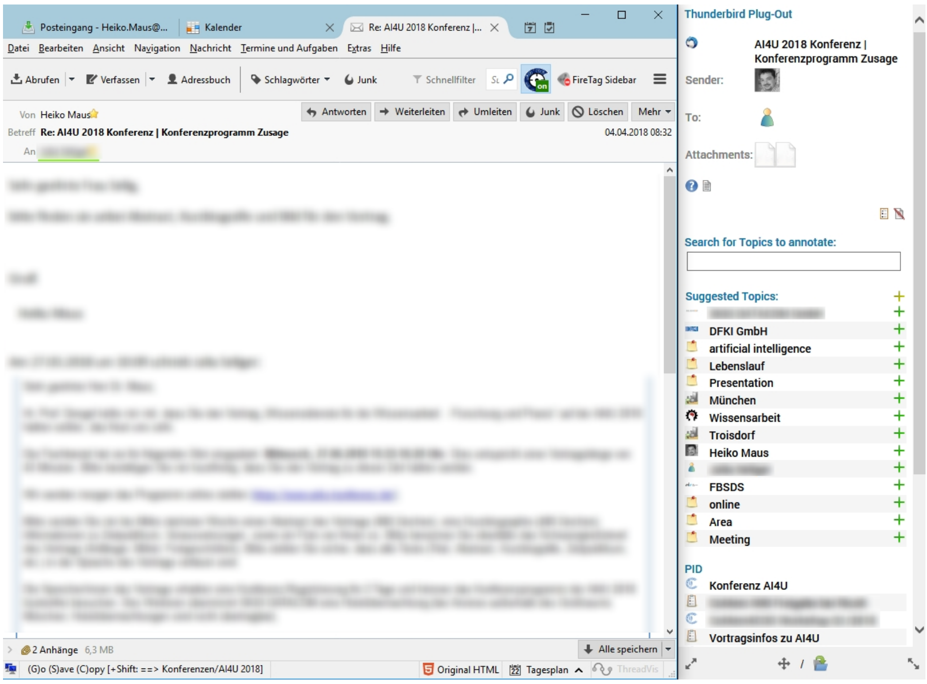
Figure 6.1: Activity Event Bus Infrastructure (AEBI).

```
{
    browserID:     1627979532973,
    browserName:   "chrome",
    timestamp:     1627983321892,
    url:           "https://www.dfki.de/en/web",
    title:         "German Research Center for Artificial Intelligence",
    content:       "<!DOCTYPE html><html lang="en-US"><head> [...] </html>",
    visible:       true,
    cause:         2, // document visibility change
    version:       1
}
```

Listing 6.1: cSpaces Web Browser Plug-out: Sample Evidence Snippet.



(a) cSpaces web browser plug-out (works with all browsers supporting *WebExtensions* [261] like *Mozilla Firefox* [Sw87], *Google Chrome* [Sw46], *Microsoft Edge* [Sw78], etc.).



(b) cSpaces email client plug-out for *Mozilla Thunderbird* [Sw88].

Figure 6.2: Example of Two cSpaces Plug-outs and A Common Sidebar.
The depicted sidebar prototype has been particularly tailored to work with CoMem and is nowadays known as the *PIMO (CoMem) Desktop Application*.
(Note: The lower screenshot was created by Heiko Maus. The upper one was created by the author and has already been published multiple times [182, 185].)

Two last reasons for the re-engineering is the **involvement of students and further synergies in GUI development** (also allowing for more rapid prototyping). Professor Dengel's chair at TU Kaiserslautern, the *Artificial Intelligence Group*[1], is carried and supported by members of DFKI SDS an their individual research topics. Earlier Semantic Desktop tools were mainly developed in *Java* [Sw95], both back- and front-end. Especially the latter was often conceived tedious by students. Colleagues and the author had good experiences using *Java* [Sw95] in the back-end and *HTML* [275], *CSS* [193] and *JavaScript (JS)* [101, 102] in the front-end. A starter package for students having a simple back- and front-end in the respective programming languages has been created by Markus Schröder and the author. After only a few hours to a day, students typically came up with first own code given the package and a first task, which suggests easy familiarization and understandability. Since then, the starter package has been used in several student projects, e.g. [63, 274, 64, 65, 159]. In particular, this also serves the educational purpose of developing and promoting students' skills in commonly used programming languages. According to the *Stack Overflow Developer Survey 2021* [361], the aforementioned languages as well as *Python* [Sw103] and *SQL* [57, 175] (both also used in typical SDS projects) actually were the top-5 of most commonly used programming languages[2].

The combination of *Java* and *HTML5* (i.e. HTML, CSS and JS) has another advantage: the author promoted the **replacement of *Java* front-end components on the desktop by embedded browser components** like the *Java Chromium Embedded Framework (JCEF)* [Sw53, Sw52], *JavaFX* [Sw96] or *DJ Native Swing* [Sw36]. By doing so, the front-end components of web and desktop apps can be developed in the same programming languages, ultimately allowing for a single codebase, higher reusability and a coherent look and feel (familiarity). Note that more functionality is available on the desktop due to browser security restrictions. Developers may either host their own JS environment or utilize the one typically coming with the embedded browser to gain access to these additional features of the local system.

Re-engineering as described above comes with two **drawbacks**. First, the sidebar is not immediately inside a respective app. At least one colleague had the feeling of "not having things immediately at the fingertips anymore". The author so far does not have any indication that this is a majority opinion but definitely a legit one. A second drawback is the further introduction of platform-dependent code with some of the embedded browser implementations. However, from the author's experience, using *OpenJDK* [Sw95] in combination with *OpenJFX* [Sw96] allows for running all so far developed CoMem and cSpaces solutions on *Windows* [Sw83], *macOS* [Sw26] and *Linux* [Sw114] thus covering a wide range of systems.

Summary. In summary, re-engineering measures discussed in this section...

- provided a renovation of the existing user activity infrastructure,
- enabled colleagues and the author to easier increase the scope of user activity tracking (see number of developed plug-outs),
- allowed for more rapid prototyping especially due to increased reuse and reduced GUI development effort (see number of realized UIs in this PhD project),

[1] `https://agd.cs.uni-kl.de/`

[2] Most commonly used programming languages according to the *Stack Overflow Developer Survey 2021* [361]: 65% *JS* [101, 102], 56% *HTML* [275] & *CSS* [193], 48% *Python* [Sw103], 47% *SQL* [57, 175], 35% *Java* [Sw95].

- enabled easier ramp-up for contributing students (see student projects and their outcome) and
- realized a more coherent look and feel of the Semantic Desktop, especially for novice users (see screenshots throughout the thesis).

A topic closely related to user activity tracking is user privacy, which is addressed more thoroughly in the next section.

6.1.2 Privacy Preservation

Systems collecting and storing potentially very sensitive user data should be privacy protecting by design. In a study of the *SWELL project* about **privacy and user trust in context-aware systems**, Koldijk et al. [211] found that "privacy information had a positive effect on perceived privacy and trust in [the] system." Said "privacy information" involved informing participants about purpose limitation (i.e. make clear for which purposes gathered data is used and that it is not used for anything else), control (e.g. tracking can be disabled), data minimization, data aggregation (e.g. local processing, only summaries are saved), adequate protection (e.g. hidden from unauthorized access) and data subjects right (e.g. participants have full control over their data, can view or delete it) [211]. They also found that the "attitude towards using [the] system was related to personal motivation, and not related to perceived privacy and trust", which is why they recommend to "implement privacy by design to adequately protect the privacy of the users" in such systems [211].

In the **scenario of cSpaces, possible solutions** are, for example:

- allowing to temporarily disable the observation (see Fig. 8.25, highlighting C),
- only store sensitive data on the user's local machine (i.e. no data leaves the user's computer unwillingly, e.g. via server/cloud uploads) or
- only stimulate the semantic graph (e.g. "activation" of respective parts) without storing any details permanently.

The last item goes into the direction of privacy preserving models, which was out of scope of this PhD project but remains an interesting topic for future work. The first two items were the main measures taken in a cSpaces-related multi-month user study, in which seven participants used the system for up to five mounts. In the beginning, they were introduced to the privacy measures, which involved several aspects also mentioned by Koldijk et al. [211] like clarifying that participants may turn off the observation at any time. Since everything that the system stored permanently was available as an *SQLite* database [166], they could use tools like the *DB Browser for SQLite* [Sw102] to view or delete their data (or parts of their data). No data should leave the users' computers unless send by themselves. For this, an anonymization tool was provided, but submitting data to the experimenter was also voluntarily. In the anonymization, basically all strings were removed. Connections in the graph could still be browsed, but all words, labels, etc. were deleted. Again, all parts of the database could be browsed before and after the anonymization. Asked in the final interview whether they felt comfortable with these privacy preservation measures (Q_{M15}), all participants but one *quite* *agreed* or *strongly agreed* as shown in Figure 6.3 (mean: *quite agree*, median: *strongly agree*). As a side question, they were also asked whether they sometimes even forgot that they were observed by the system (Q_{M16}). Answers were on a wide spectrum between *strongly disagree* and *strongly agree* with the mean and median slightly on the positive side.

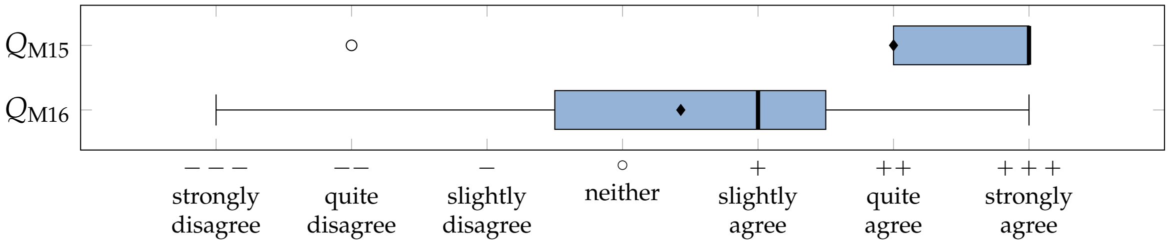


Figure 6.3: Multi-month User Study: Survey Result for Privacy Preservation.

Q_{M15}: I feel comfortable with the current prototype's privacy preservation measures.
Q_{M16}: I sometimes forgot that I am being observed by the system.

(n=7; full questionnaire document in Sec. A.4; details on box plot semantics in Sec. B.2)

Having discussed how evidence snippets are obtained by user activity tracking, the next step (2) of the cSpaces interaction cycle (see Sec. 1.4) is the further processing of these snippets by means of information extraction as presented in the next section.

6.2 Information Extraction by Specifically Tailored Named Entity Recognition (NER)

Note: Section 6.2 is in large part a revised version of Jilek et al. [186].

Before the actual context elicitation, i.e. the interpretation of user activities (see next section), an information extraction task has to be performed. cSpaces mainly focuses on extracting named entities known from the user's PIMO from the obtained evidence snippets. The idea is to track which parts of the knowledge graph (KG) are currently stimulated by user activities, especially with the intention to constantly update *Memory Buoyancy* (see Sec. 7.2) and perform support measures accordingly. This process is illustrated in Figure 6.4: Things of the user's PIMO are detected on a browsed website (left-hand side, highlighting A). These concepts (as well as the website's textual content) are used for context elicitation. The concepts are also shown as suggested topics in the sidebar (right-hand side, highlighting B). Users may confirm them making them annotated topics (tags) of the website (highlighting C).

Note on terminology: The literature speaks of *Named Entity Recognition (NER)* in this regard. However, cSpaces does not look for new entities not yet in the PIMO (at least not currently; possible future work), so the task is rather the recognition of *known* named entities, which is also called *Named Entity Linking (NEL)* in literature. In various publications, the term NER also comprises NEL.

Since cSpaces' target is real-time user support, the NER task should take as less time as possible: According to Miller [255] and Card et al. [53], as cited in Nielsen [271], 100 ms is

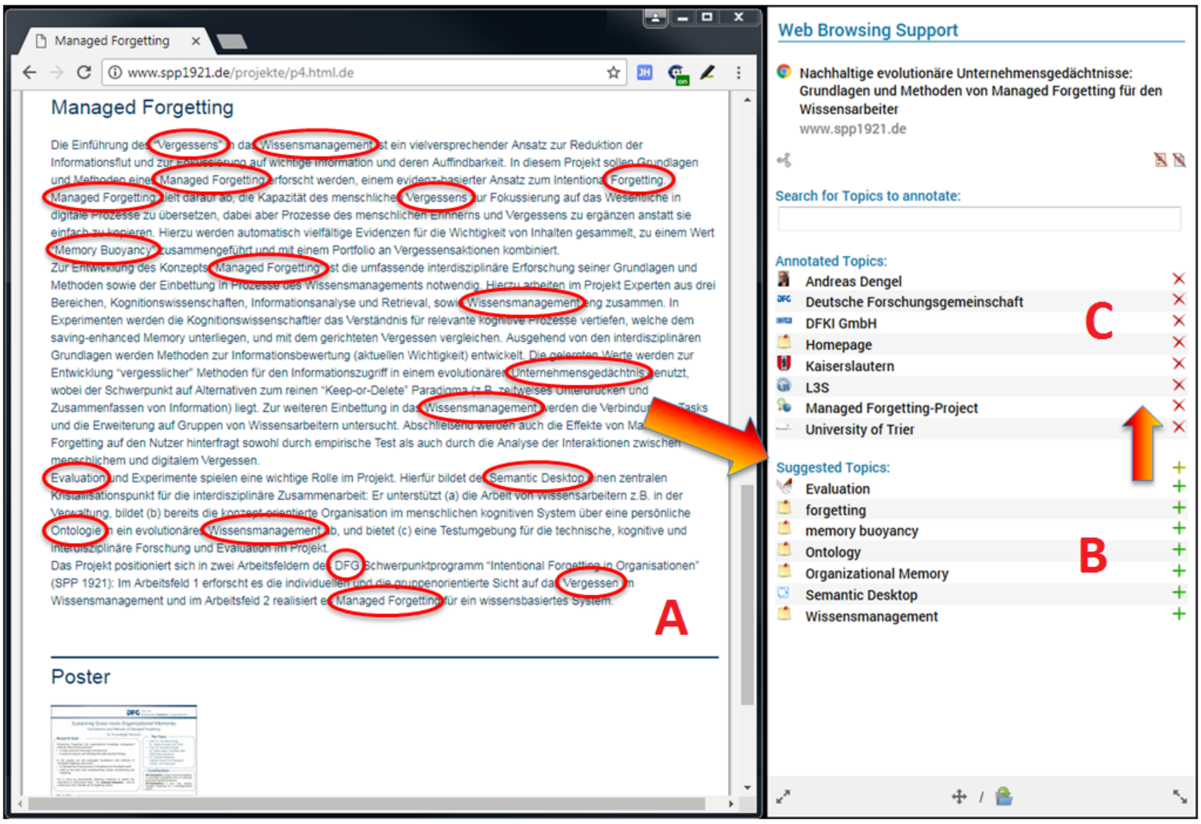


Figure 6.4: Named Entity Recognition on the Content of a Browsed Website.
Highlighted sections: *A: found entities in the website's text / B: found entities deduplicated and ranked [suggested topics] / C: topics confirmed by the user [annotated topics].*
Note: This screenshot [without highlighting] has already been published multiple times [182, 185].)

"about the limit for having the user feel that the system is reacting instantaneously" and 1000 ms is "about the limit for the user's flow of thought to stay uninterrupted". Since the system also needs some time for calculating and performing the support measures, the actual NER process should only take a small two-digit number of milliseconds. Additionally, a certain robustness regarding lexical variation of entity names is required, so the system does not lose track of users' activities (by not detecting mentioned entities). This is especially relevant in highly inflectional languages like German (but also other ones like Spanish, Latin, Hebrew, Hindi, Slavic languages, etc.). Figure 6.5 illustrates the problem of inflections in NER using the example of *DBpedia Spotlight* [Sw74], a well-known and often used recognizer to detect *Wikipedia* [398] or *DBpedia* [Sw34] entities, respectively, in given text snippets. (For readers unfamiliar with the topic: in short, *DBpedia* is a more structured, thus easier machine-readable version of *Wikipedia*.) The first paragraph of the German Wikipedia article of *Aussagenlogik [Propositional calculus]*[3] (top) was fed to the recognizer. Green entities are non-linked entities to be found by the system. As highlighted in yellow in the middle section of the figure, twelve entities (in just three sentences) were not found, ten of them due to lexical variations induced

[3] https://de.wikipedia.org/wiki/Aussagenlogik

by inflection. For example, *Wahrheitswert [truth value]* was found, whereas its inflected (plural) forms ending with -e and -en were not. Lowering the confidence from 0.5 to 0.0 (bottom part of the figure) increased the number of found entities, however, false positives started to come up (highlighted in red).

Figure 6.5: Illustrating the Problem of Inflections in Named Entity Recognition.
The first paragraph of the German Wikipedia article of *Aussagenlogik [Propositional calculus]* (top) was fed to *DBpedia Spotlight* [Sw74], a tool to detect *Wikipedia* [398] or *DBpedia* [Sw34] entities, respectively, in given text snippets. Confidence values of 0.5 (center) and 0.0 (bottom) were used.
Color highlighting: *green: not linked entities to be found / yellow: entities not found (ten of twelve entities not found due to inflection) / red: false positives (coming up if confidence gets too low).*
(Note: This screenshot has already been published in Jilek et al. [186].)

The envisioned NER approach thus needs to focus on two major aspects: 1) inflection tolerance and 2) real-time capability. And since it is also ontology-based with PIMO as its source for entity names to detect, the author called it IT-RTC-OBNER standing for *inflection-*

tolerant, real-time capable ontology-based named entity recognition. The IT-RTC-OBNER approach is introduced in detail in Section 6.2.2 after presenting related work in the next section.

6.2.1 Existing Inflection-tolerant or Very Fast NER Approaches

The search for related work focused on inflection tolerance or real-time capability, preferably both at the same time.

Existing Inflection-tolerant NER Approaches. Dey and Prukayastha [91] gave an overview on different NER methods especially targeting Indian languages presenting gazetteer-based and machine learning approaches as well as hybrid solutions. Al-Jumaily et al. [196] presented an NER system for Arabic text mining. They used a token-based approach involving stemming as well as pre- and postfix verification tailored to the Arabic language. Although they aimed for real-time applications, they did not give any details about their system's runtime performance. Savary and Piskorski [310] investigated solutions for Polish, also a highly inflectional language. As one sub-component of their information extraction platform *SProUT*, they filled a gazetteer by "explicitly listing all inflected forms of each entry", which is adopted by IT-RTC-OBNER.

Existing Real-time Capable NER Approaches. Such strict temporal requirements as stated above usually rule out very sophisticated *Natural Language Processing (NLP)* pipelines (higher quality solutions but slower), leaving only rather simple (lower quality) but very fast methods often based on pre-defined rules or gazetteers.

Dlugolinskỳ et al. [93] presented an overview on different gazetteer-based approaches, especially referring to various versions included in the *General Architecture for Text Engineering (GATE)* framework [Sw42]. They distinguished between character- and token-based variants and stated that the latter usually have "longer running time and low processing performance" [93]. They thus focused on character-based gazetteers and presented several versions [93, 269]. Since some of their implementations are available online [Sw93], they are used as a baseline in the evaluation of IT-RTC-OBNER (see Sec. 6.2.3). Al-Rfou and Skiena [286] proposed *SpeedRead*, an NER pipeline which they tested to run at least ten times faster than a comparable *Stanford CoreNLP* pipeline [Sw113]. They only reported runtime performance in terms of tokens per second. In their final results, SpeedRead achieved about 153 tokens/second. Using word length statistics published by Norvig [273] and assuming an average token length of five, this can be translated to 765 characters/second. Even if an average token length of twelve is assumed (although 90% of all English words are shorter [273]), the speed of 1836 characters/second would still be too slow for cSpaces scenarios.

Summary. In summary, the author found several approaches dealing with either inflection tolerance or real-time capability. Only one paper mentioned both but did not report any concrete speed measures. Doing NER extremely fast is apparently rarely discussed in literature, yet. This may be because usual NER methods operate in only a few seconds, which may be sufficient for many use cases, but not for the ones of cSpaces. The author therefore envisioned and realized an own approach, IT-RTC-OBNER, which is presented in the following.

6.2.2 IT-RTC-OBNER – Combining Inflection Tolerance and Real-time Capability

IT-RTC-OBNER focuses on the very fast recognition of named entities given as instance labels of KGs. These labels should still be recognized by the system, even if they lexically vary slightly as induced by inflection. To achieve this, KGs like PIMO or LOD sources like *DBpedia* [Sw34] or *Wikidata* [Sw121] are exploited to get more details about the entities, e.g. their specific type. Based on this type and language information, e.g. coming from *Wiktionary* [399], different lexical variations per instance can be accepted. For example, not too many variations of person names should be accepted, whereas the system can be more tolerant with topic, project, organization or location names, especially if they contain adjectives like the "*Technical* University of Kaiserslautern" or "*German* Research Center for *Artificial* Intelligence". As an example, Figure 6.7 (word w4) shows all 18 inflected forms of the German adjective *künstlich* [*artificial*].

Overall Architecture – A Multi-FST. The overall architecture of IT-RTC-OBNER is depicted in Figure 6.6: the upper half of the figure shows its core, a named entity recognizer as a combination of several *Finite State Transducers (FST)* [292] having different configurations, for example with regard to tolerance levels, languages, case sensitivity, etc. All FSTs operate in parallel and in the end, the overall system, the *Multi-FST*, collects their individual results and decides (via a voter) which ones to accept, rank high, etc. To acquire the entity labels as well as background information, it is connected to KGs (bottom left) and language information sources (bottom right) as described before.

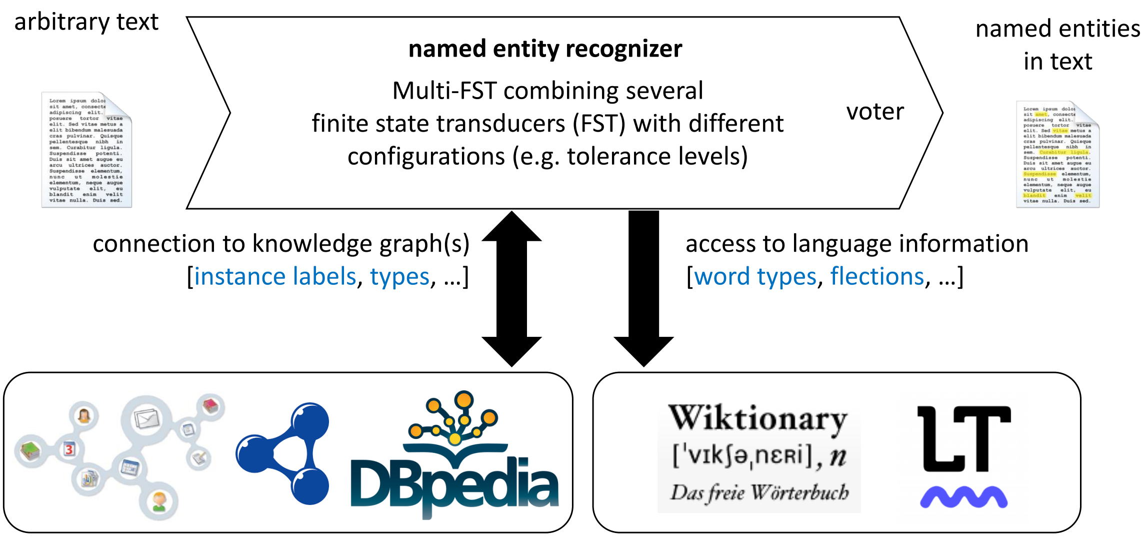


Figure 6.6: IT-RTC-OBNER Approach: Architecture.
(Note: An earlier version of this image has been published in Jilek et al. [186].)

To meet the aforementioned strict runtime requirements, IT-RTC-OBNER basically follows a gazetteer-based approach. Since the additionally required inflection tolerance is not well compatible with the typically static character of a gazetteer, the approach is enhanced as explained in the following.

FST-based NER. Each FST that is part of IT-RTC-OBNER utilizes the well-known string matching algorithm by Aho and Corasick [9]. It operates on *tries* [79, 123], i.e. trees whose nodes represent characters, which are traversed synchronously to the processing of each character of the input text. Whenever the traversal ends in an accepting state, there is a string match. Since these strings are the labels of named entities, their ID or URI is returned in case of a match, which makes the system an actual FST [292] – as opposed to a finite state *machine* that does not return a particular output and just stops in an accepting or non-accepting state. The algorithm basically has linear runtime complexity as explained in the subsection about real-time capability. Since cSpaces scenarios involve highly dynamic, evolving KGs, in which instances (and in particular their labels) can be added, deleted or updated potentially in high frequency, further optimizations like suffix compression were omitted in favor of fast and easy to update FST structures.

Multi-layer FST. IT-RTC-OBNER follows the approach of explicitly listing all inflected forms of an entity label as proposed in Savary and Piskorski [310]. Without further measures, this can easily lead to memory performance problems due to a considerable increase of the FST, especially for multi-word terms: The more words such a multi-word term consists of, the more potential combinations exist. Although inflection tolerance is discussed more thoroughly in the paragraph after next, a short example is also discussed in the following: Accepting each combination of inflected forms of the term *Deutsches Forschungszentrum für Künstliche Intelligenz* [*German Research Center for Artificial Intelligence*], although lots of them are grammatically not correct, would end up with 576 variations ($= 6 \cdot 3 \cdot 1 \cdot 16 \cdot 2$; see upper part of Fig. 6.7). Inspired by Abney, who proposed the idea of *finite state cascades* [6], an additional layer is therefore introduced into the FST to separate character from word processing. This makes the FST a multi-layer FST as illustrated in Figures 6.7 and 6.8 for the cases of high and low tolerance, respectively. Additional layers could be added in the future, for example scanning for phrases that indicate to-dos/tasks, appointments, *Hearst patterns* [157], etc.

Once a word is identified in the first layer (i.e. the FST is in an accepting state; highlighted as a gray node), its ID is passed to the second layer, which checks whether this word may be accepted at this position, either as a single-word or part of a multi-word term. If a term match is detected, its ID/URI is returned. As a consequence, each word and its inflected forms, no matter how often or at which positions (in multi-word terms) they appear, only exist once in the FST thus preventing it from growing too fast in size.

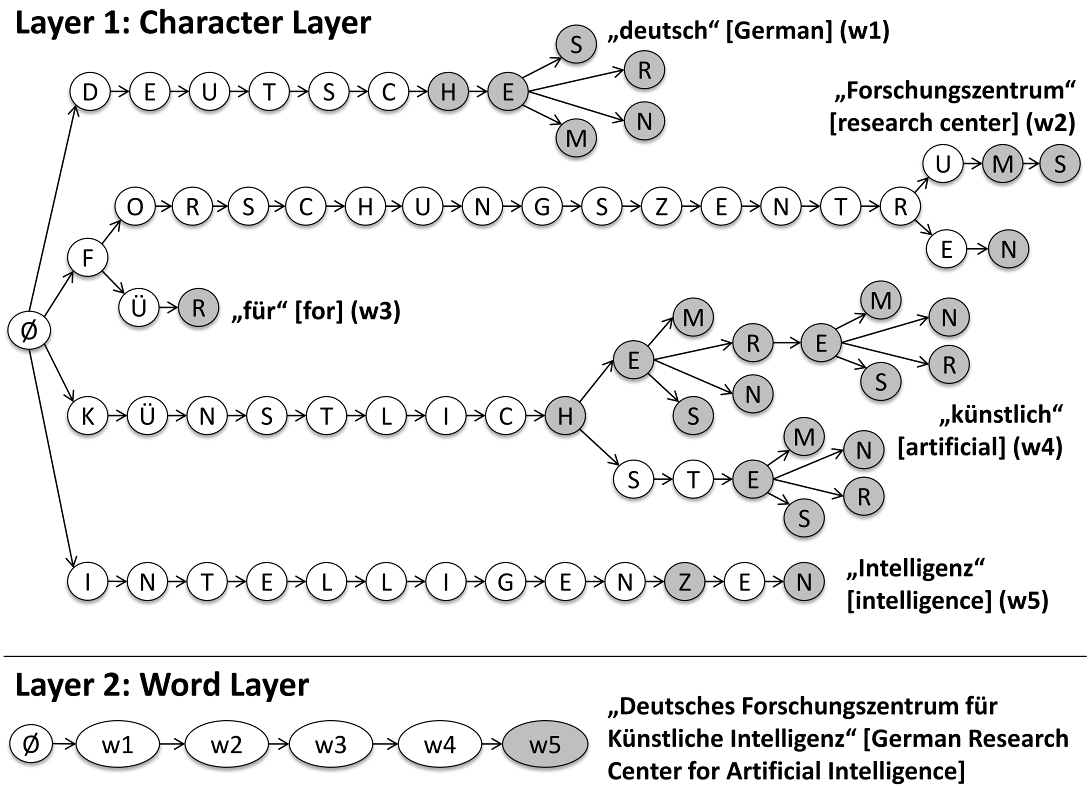


Figure 6.7: IT-RTC-OBNER Approach: Multi-layer FST with High Inflection Tolerance. It was fed with the term *Deutsches Forschungszentrum für Künstliche Intelligenz* [*German Research Center for Artificial Intelligence*].
Ø: start nodes / w_i: word IDs / gray nodes: accepting states.
(Note: This figure has already been published in Jilek et al. [186].)

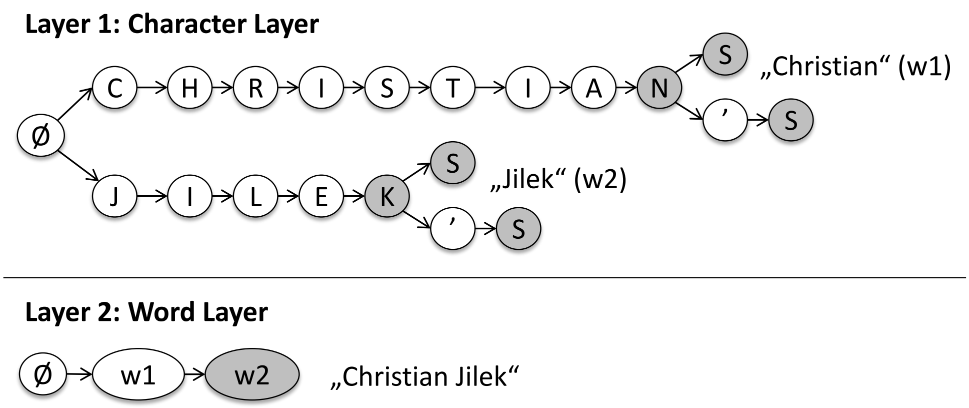


Figure 6.8: IT-RTC-OBNER Approach: Multi-layer FST with Low Inflection Tolerance. It was fed with the Term *Christian Jilek*.
Ø: start nodes / w_i: word IDs / gray nodes: accepting states.
(Note: This figure has already been published in Jilek et al. [186].)

To avoid backtracking in the word layer, the system processes several options in parallel as shown in Figure 6.9.

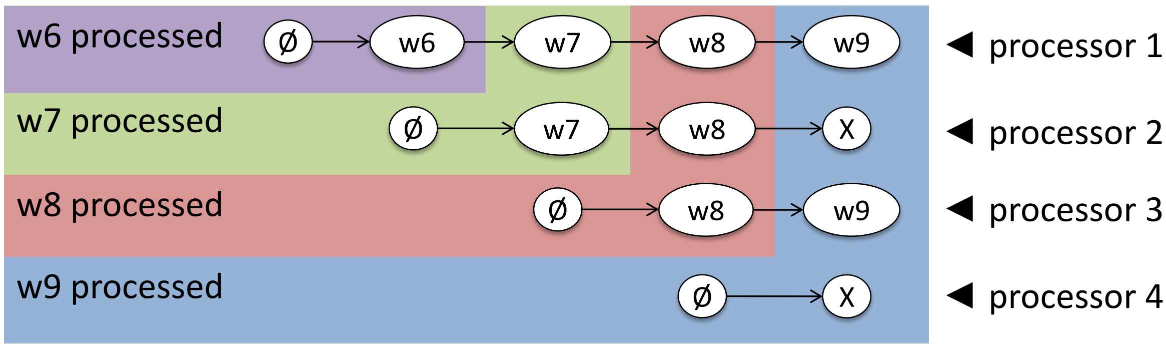


Figure 6.9: IT-RTC-OBNER Approach: Processing in the Word Layer.
Ø: start nodes / w_i: word IDs / X: failure states.
(Note: This figure has already been published in Jilek et al. [186].)

Once the character layer recognizes a word, e.g. `w6`, a new word node processor in the second layer is spawned (see upper left part of the figure; purple color). If layer 1 then reports the next word `w7` (highlighted in green), processor 1 goes one step further in the graph now having a traversed path containing both words. Additionally, another processor is spawned, starting directly with `w7`. Drawing all start nodes as a single node would yield the image of a rake. Metaphorically speaking, spawning another processor is like adding another tine to the rake. Traversals in the word layer are only possible if the next detected word is a successor of the current one within any term of the FST, which, for example, is not the case when processor 2 tries to handle `w9`, or processor 4 tries to start directly with `w9`. The latter means that there is no term in the FST starting with the word `w9`. These two processors are then in a failure state as indicated by an "X". If there was a matching term in their traversed path, it is collected to be later processed by the voter. If that is not the case, the failed processors may be removed from the rake. The second case in which processors are removed, whether they are in a failure state or not, is after an explicit signal from the first layer, e.g. when reaching the end of a file or sentence. Spawning additional processors to evaluate different possibilities in parallel especially originated from the latter. Consider the case of interpreting a dot: It could either indicate the end of a sentence (*"Today, I met my Prof."*), or an abbreviation (*"Prof. Smith was also there."*). Thus, there is a forking in the second layer to evaluate both possibilities separately. In theory, this could lead to endless forking, which is prevented by processors reaching failure states (i.e. given word sequences not matching any term) followed by their removal. The basic steps of the multi-layer FST are given as pseudocode in Algorithm 6.1.

Algorithm 6.1 Basic Steps of the Multi-layer FST in Pseudocode.
(Note: This pseudocode has already been published in Jilek et al. [186].)

input: text to process (*text*)
output: found entities (*foundEntities*)

```
foundEntities ← { }
collectedTerms ← { }
c ← first character of text
w ← c
while c not equals EOF (end of file or text snippet) do
    if c is whitespace character then
        if w matches in character layer then
            add new word node processor (in word layer)
            for all word node processors p do
                process w with p (may either lead to word match or failure state in p)
            end for
        end if
        collectedTerms ← collectedTerms ∪ collect matching terms from word layer
        remove word node processors in failure state (word layer)
        w ← ∅
    else
        w ← w + c
    end if
    c ← read next character of text (character layer)
end while
collectedTerms ← collectedTerms ∪ collect matching terms from word layer
foundEntities ← do voting on collectedTerms (word layer)
return foundEntities
```

Real-time Capability. Reading an input text of length n characterwise yields a basic runtime complexity of $\mathcal{O}(n)$. The same is true for processing n characters in the first layer (at most n transitions having a constant amount of operations; no backtracking needed). The processing of a character may lead to the detection of a new word, which then triggers transitions in the word layer. The number of these transitions depends on the number p of processors. p does not depend on n, but on the vocabulary, i.e. all words fed to the FST, especially $w_{\max}$, the maximum number of words in all multi-word terms. Although $p_{\max}$ is constant for a given vocabulary, it may still be very large in worst case, which is considered in the following.

In worst case, a term consisting of $w_{\max}$ words is read, whereas each sub-term also exists in the vocabulary. Moreover, for each of these sub-terms, there is an additional variant ending with a dot. This leads to forking after every word and a total amount of $p_{\max} = \sum_{i=1}^{w_{\max}} 2^i$ processors before the first one of them fails and is removed. In practical scenarios however, $p \ll p_{\max}$ can be assumed since the vocabulary is only a tiny fraction of the power set of its words. As a consequence, processors fail very fast due to given word combinations not matching any term in the FST. Considering an additional constant amount of $c > 0$ operations per processor in each transition of the second layer yields an upper limit of $c \cdot p_{\max} \cdot n$. Since n is thus only multiplied with constants, the overall runtime complexity remains $\mathcal{O}(n)$. Although the second layer's overhead is noticeable in practice, the overall runtime complexity is still linear and benefits our system's applicability in scenarios of real-time processing.

Inflection Tolerance. To accept different lexical variations of terms, e.g. induced by inflection, information is utilized that comes from connected KGs and other language information sources. With regard to the latter, a lemmatization table from the *German part-of-speech dictionary* [Sw91] of *LanguageTool* [Sw69] was used. As an example, all entries for the German adjective *künstlich* [*artificial*] are shown in Table 6.1. They contain the inflected form, its lemma as well as declension information like word class, case, number, gender, etc. Using *JWKTL* [Sw115], further language information was extracted from the *German Wikitionary* [Sw123], a free wiki-based dictionary. Although these sources provided a good foundation to start with, lots of words were still not covered by any of them, especially compound words like *Forschungszentrum* [*research center*]. To counter this, an additional heuristic was implemented that uses longest suffix matching to decompound words and use the inflected forms of the last part (if available). In the case of *Forschungszentrum* not being found in the initial database, the heuristic would first look for the word *orschungszentrum* (which fails), then *rschungszentrum* (fails again), *schungszentrum* (fails), etc., until finally finding *zentrum* and using its inflected forms, i.e. *zentrum*, *zentrums* and *zentren*. The matching part of the original word is then replaced with these inflected forms as shown in Figure 6.7. The heuristic expects a parameter indicating the minimum length of the remaining suffix to get more meaningful results. For example, in the case of *Malerei* [*painting*] (not a compound word), the algorithm would add the inflected forms of *Ei* [*egg*], which would not be meaningful. A minimum suffix length of five is currently used and has also been used in the evaluation. With the heuristic, the tool is able to handle yet unknown words to a certain extent without user interaction.

In this regard, remember the aforementioned 576 variations of the term DFKI. As also mentioned, most of these variants are grammatically not correct. Since yet unknown words, especially compound words, should still be handled by the system while keeping the user

inflected form	**lemma**	**linguistic information**
künstlich	künstlich	ADJ:PRD:GRU
künstliche	künstlich	ADJ:AKK:PLU:FEM:GRU:SOL
künstlichem	künstlich	ADJ:DAT:SIN:MAS:GRU:SOL
künstlichen	künstlich	ADJ:AKK:PLU:FEM:GRU:DEF
künstlicher	künstlich	ADJ:DAT:SIN:FEM:GRU:SOL
künstlichere	künstlich	ADJ:AKK:PLU:FEM:KOM:SOL
künstlicherem	künstlich	ADJ:DAT:SIN:MAS:KOM:SOL
künstlicheren	künstlich	ADJ:AKK:PLU:FEM:KOM:DEF
künstlicherer	künstlich	ADJ:DAT:SIN:FEM:KOM:SOL
künstlicheres	künstlich	ADJ:AKK:SIN:NEU:KOM:IND
künstliches	künstlich	ADJ:AKK:SIN:NEU:GRU:IND
künstlichste	künstlich	ADJ:AKK:PLU:FEM:SUP:SOL
künstlichstem	künstlich	ADJ:DAT:SIN:MAS:SUP:SOL
künstlichsten	künstlich	ADJ:AKK:PLU:FEM:SUP:DEF
künstlichster	künstlich	ADJ:DAT:SIN:FEM:SUP:SOL
künstlichstes	künstlich	ADJ:AKK:SIN:NEU:SUP:IND

Table 6.1: Lemmatization Table for the German Adjective "künstlich" ["artificial"].
Linguistic information: *[word class] : [case] : [number] : [gender] : [comparison level] : [usage].*
Abbreviations: *ADJ: adjective / AKK: accusative / DAT: dative / DEF: definite / FEM: feminine / GRU: basic form / KOM: comparative / MAS: masculine / NEU: neuter / PLU: plural / PRD: predicative / SIN: singular / SOL: solitary / SUP: superlative* [224].
(Source: Naber et al., *German part-of-speech dictionary* [Sw91].)

interaction as low as possible (i.e. no feedback asked), IT-RTC-OBNER accepts all variations obtained as the Cartesian product of all inflected forms of each of a term's words. This follows the assumption that grammatically wrong variants do rarely occur in given texts and if they do, users will agree with the entity being recognized despite the misspelling. Nevertheless, the question remains whether this decision considerably increases the recognizer's false positive rate. This is addressed in the evaluation section (6.2.3).

To avoid actually harmful false positives of incorrectly inflected variants, additional ontological information like the type of an entity is exploited. For example, the name of a person tolerates far less variants than the name of a topic. Basically, only a genitive case "s" at the end is allowed as a variation as shown in Figure 6.8. As mentioned before, the overall Multi-FST is a combination of several FSTs with different configurations. Besides the ones with higher and lower tolerance (examples shown in Figure 6.7 and 6.8), there is also the option to deal with acronyms for which only exact matches with uppercase-only characters are accepted.

To further avoid non-meaningful variants, only adjective and noun information from the lemmatization table is used, which reduces ambiguities when not having thorough NLP information (e.g. cases like the German noun *Herzog* [*duke*] and the verb *herzog* [*pull near* or *drag along*]). This is an acceptable compromise since labels more often contain nouns and adjectives than verbs.

Further Multi-FST Details. As discussed earlier, when processing input text, the different FSTs of the Multi-FST operate in parallel. In the end, a voter receives, assesses, filters and finally returns their results. Additionally, each FST has its own internal voter, which assesses all results simultaneously present in a processing rake (see Fig. 6.9). In the current implementation, these voters follow a strategy of only keeping the longest match, e.g. if the term *personal information management* is found in the text, the also matching terms of *personal information* and *information management* would be discarded.

In summary, IT-RTC-OBNER combines four ideas:

- the well-known Trie-based string matching algorithm by Aho and Corasick [9],
- finite state cascades as proposed by Abney [6],
- explicitly listing all inflected forms of an entity label as proposed by Savary and Piskorski [310] and
- using ontological background information like entity types to decide which inflection tolerance strategies to apply.

For now and especially for the evaluation presented in the next section (6.2.3), the Multi-FST consists of two multi-layer FSTs with a high and low tolerance configuration, respectively. This configuration is referred to as *IT-RTC-OBNER Prototype I*. However, further improvements have been made after the evaluation taking its results into account. Thus, an *IT-RTC-OBNER Prototype II* is presented in the section after next (6.2.4).

6.2.3 IT-RTC-OBNER Prototype I Evaluation

The two man objectives of this evaluation were to find out 1) how fast *IT-RTC-OBNER Prototype I* performs in practice and 2) whether the above design decisions (e.g. accepting various inflected forms despite being grammatically not correct) would lead to a considerable increase in false positives.

Evaluation Setting. The author therefore looked for large amounts of German natural language texts (prose) written by different people to test the approach. The German Wikipedia [398] met this requirement but lacked ground truth data for the inflected forms present in these texts. There was, however, the idea to only look at the *wikilinks* (see Fig. 6.5, top section, blue words) and take them as a silver standard: A human has annotated terms in the text (often in inflected form) with the label of their respective Wikipedia article (typically in basic form). Figure 6.5 also shows that users themselves decide which terms they annotate: There are lots of entities (highlighted in green), which are not annotated although there is a Wikipedia article for them. This is especially true for self-references, e.g. the term *Aussagenlogik* [*Propositional calculus*] is not annotated in "its own" article (i.e. the one about *Aussagenlogik*). Recognizers fed with such terms, would nevertheless find them, which has to be considered when measuring precision. Regardless of possible shortcuts, annotations are structured as follows: the term appearing in the text and the name of the Wikipedia article it refers to (in the following also shortly called *the link*) are written in double brackets separated by a pipe symbol, e.g. `[[Häuser|Haus]]` [*houses* | *house*] – the plural form of *house* appears in the text, whereas the article is labeled with the singular form. Since inflection usually just changes one to four characters, the *Levenshtein distance* (LD) [221] between term and link can help identifying samples that can be used to evaluate the approach. Note that independent term-link-combinations like [hometown | Eton] or adjective-noun-combinations like [entscheidbar | Entscheidbarkeit] [*decidable* | *decidability*] are undesirably also covered by such an LD-based heuristic. On the other hand, this evaluation approach offers millions of inflection samples – tests were run on 3.9 million articles having 50.4 million annotations. A German Wikipedia dump file [Sw122] was downloaded and 3.9 million article names were taken as a basis for feeding the recognizers. Disambiguation information in brackets like in *"Berlin (Russland)"* (a village in Russia sharing its name with the German capital) were removed – this raises disambiguation issues as discussed later. Number-, symbol- and single-character-only labels were also removed since they were not relevant for the investigation. As ontological background information, type information coming from *DBpedia* was used [Sw35], which was available for about 0.5 million entities. For types like person, city, film, etc., a low tolerance strategy was applied (i.e. genitive case "s" was the only accepted variation), whereas all other ones were treated with higher tolerance.

Evaluated Named Entity Recognizers. As mentioned in the previous section, the Prototype I of IT-RTC-OBNER consists of two multi-layer FSTs with a high and low tolerance configuration, respectively. Since both apply lemmatization, Prototype I is abbreviated by *LemMLFST* (Lemmatizing Multi-layer FST) in the following. LemMLFST was evaluated against three baseline methods: The first one, *StemSLFST* (Stemming Single-layer FST) uses LemMLFST's character layer combined with the *Lucene German Stemmer* [Sw14], which is based on Caumanns [55]. The other methods were the previously mentioned ones by Dlugolinskỳ et al. [93], who made two of their gazetteers available online [Sw93]: one based on *hash-map multi-way trees (HMT)* and the other based on *first child-next sibling binary trees (CST)*. Both produce the same results in terms of found named entities, but differ in memory consumption and runtime performance. After filtering and editing as mentioned in the previous paragraph, a bit more than 3.3 million article names of the German Wikipedia were left and fed to all four recognizers. HMT and CST take this input without further changes. StemSLFST splits each term into words and reassembles it after stemming them. Then it adds the altered term to its FST. LemMLFST does the same but instead of stemming the words, it looks up (or tries to infer) their inflected forms. Completely filled, the inner high-tolerance FST contained 8.5 million character nodes and 3.5 million word nodes, the low-tolerance part kept 1 million and 0.4 million nodes, respectively.

Results

All computations were performed on an Intel Core i7-4910MQ 2.9 GHz PC having 16 GiB of RAM and running on Windows 7 64-bit [Sw83].

Recall. In the context of NER, Tjong Kim Sang [375] defined recall to be "the percentage of named entities present in the corpus that are found by the system":

$$\text{recall}_{\text{NER}} := \frac{\text{number of named entities present in the corpus that are found}}{\text{number of named entities present in the corpus}}$$

All recognizers reached recall values slightly below or above 70%. Figure 6.10 shows the results itemized by Levenshtein Distance (LD) [221]. If term and link match exactly (`LD=0`, which is the case for 69% of all annotations), all recognition rates were above 92%. Note: Errors in the dump and imperfect parsing caused a slight decrease – 100% actually were expected. In LD ranges of `LD=1` to `LD=4` (11% of all annotations), HMT/CST's recall was close to 0%, whereas LemMLFST still had rates of 79%, 66%, 36% and 8%, respectively. StemSLFST even had slightly higher rates. Reaching recall near 100% should not be expected since not all variations are caused by inflection and their number decreases with increasing LD. For higher LD values (`LD>4`, 21% of all annotations), all recognition rates were close to 0%.

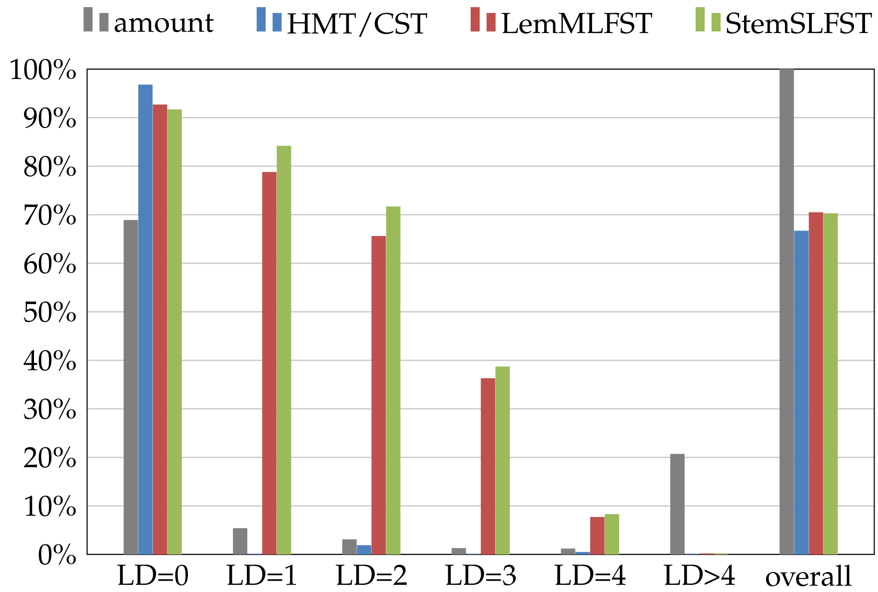


Figure 6.10: IT-RTC-OBNER Prototype I: Recall.
Comparing *LemMLFST* to *StemSLFST* and *HMT/CST* [93, 269] with regard to recall itemized by Levenshtein Distance (LD) [221] of term and link.
(Note: A variant of this figure has been published in Jilek et al. [186].)

Precision. As in the case of recall, the definition of precision by Tjong Kim Sang [375] in the context of NER is used:

$$\text{precision}_{\text{NER}} := \frac{\text{number of named entities found that are correct}}{\text{number of named entities found}}$$

The problem of how to measure precision adequately in the given scenario has already been mentioned. As a solution, the author decided to calculate multiple values: P_{O} measures precision only for terms *overlapping* with annotation positions, because only there "ground truth" data is available. As shown in Figure 6.11, some found terms (purple highlighting) were not exactly matching the actual annotation (blue word, highlighted in green as the only true positives). If terms were overlapping with the annotation, they were counted as a false

Figure 6.11: IT-RTC-OBNER Prototype I: Illustrating Different Precision Values.
(Note: This figure has already been published in Jilek et al. [186].)

positive. P_{A} counts *all* terms that were not exactly matching as a false positive, especially also the non-overlapping ones (red highlighting). Since disambiguation was out of scope and there were labels belonging to more than 1000 instances (e.g. *Jewish cemetery*), it makes a large difference whether or not additionally counting more than 1000 false positives for each true positive in a text. Thus, P^*_{O} and P^*_{A} were introduced, which count multiple entities having the same label only once. P^*_{O} was 79% for HMT/CST and 80% for LemMLFST, while StemSLFST only reached 71%. Figure 6.12 depicts P^*_{O} itemized by the number of words a

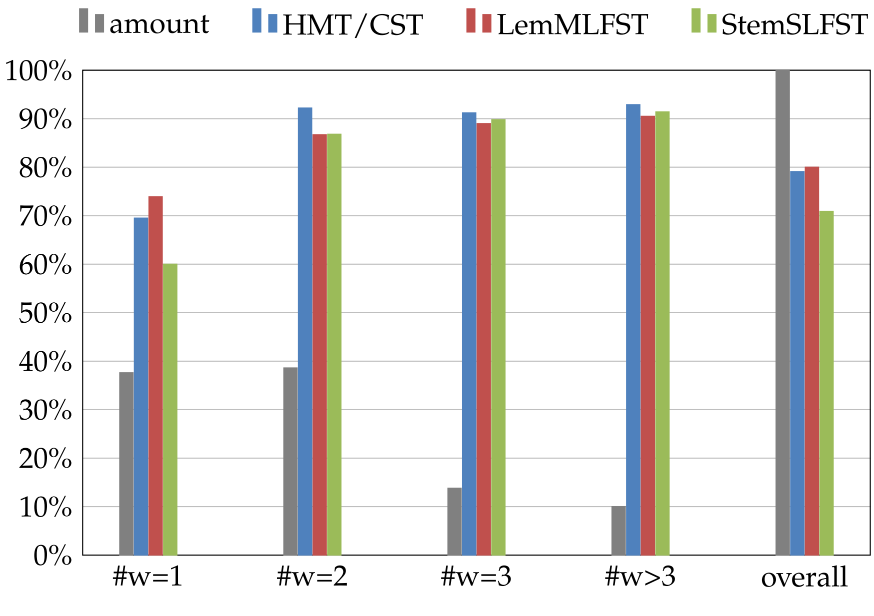


Figure 6.12: IT-RTC-OBNER Prototype I: Precision P^*_{O}.
Comparing *LemMLFST* to *StemSLFST* and *HMT/CST* [93, 269] with regard to precision P^*_{O} itemized by the terms' number of words (#w).
(Note: A variant of this figure has been published in Jilek et al. [186].)

term consists of. For multi-word terms, all approaches achieved values between 87% and 92%. There is a remarkable difference for single word terms: Here, stemming seems to be too rough causing terms to lose their specifity and StemSLFST to lose 14% compared to LemMLFST, which performed best having 74%.

The other overall precision values P_O, P_A and P^*_A are shown in Figure 6.13. They were far

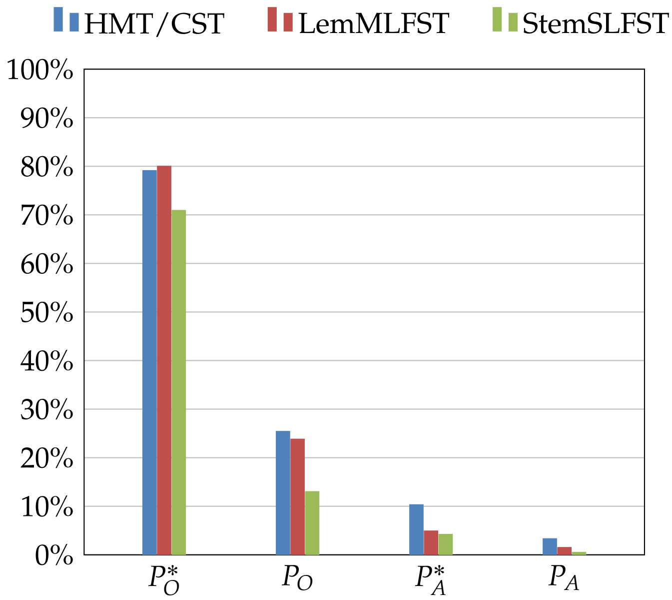
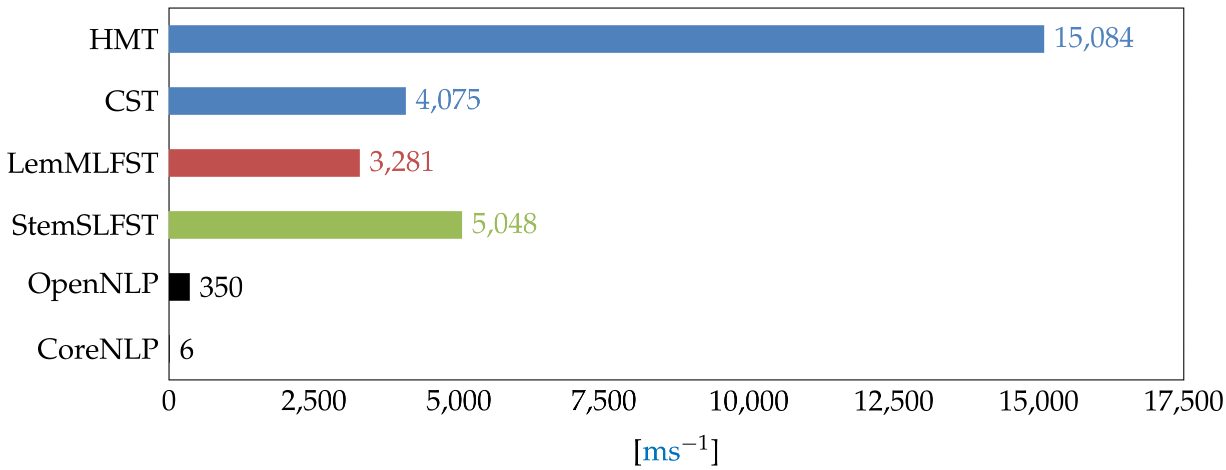


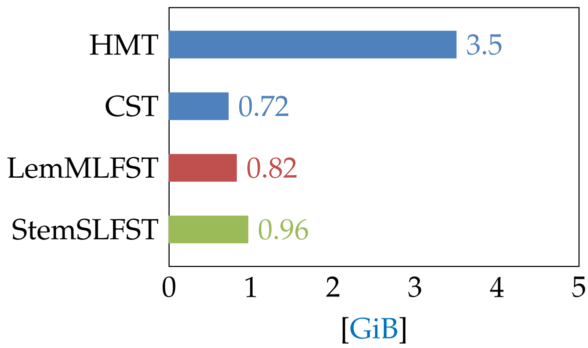
Figure 6.13: IT-RTC-OBNER Prototype I: Precision Values P^*_O, P_O, P^*_A and P_A.
Comparing *LemMLFST* to *StemSLFST* and *HMT/CST* [93, 269] with regard to precision values P^*_O, P_O, P^*_A and P_A.
(Note: A variant of this figure has been published in Jilek et al. [186].)

lower than P^*_O due to the aforementioned reasons. However, in a short experiment, in which students annotated some randomly chosen articles manually, values for P^*_A that were similar to P^*_O above could be observed. There is thus a slight indication that P^*_A heavily underestimates LemMLFST's precision. With regard to one of the evaluation's main questions, the false positive rate of LemMLFST was not considerably higher – in some cases even lower – than with the other recognizers.

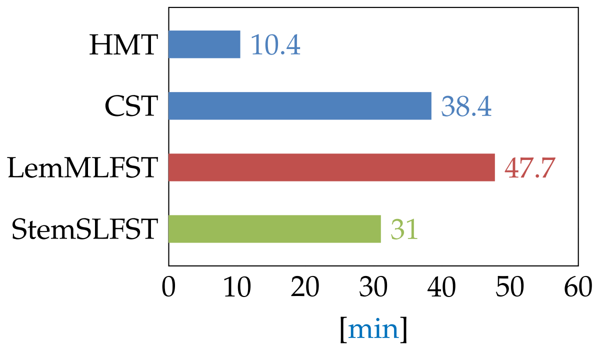
Memory and Runtime Performance. As shown in Figure 6.14c, HMT only needed 10.4 min for processing all 3.9 million articles (9.4 billion characters), while the others needed 31.0 to 47.7 min. Figure 6.14b shows that HMT trades memory efficiency for speed since it was the only recognizer passing the 1 GiB mark by needing 3.5 GiB. The others needed 0.72 to 0.96 GiB. Regarding runtime performance in terms of characters per millisecond, LemMLFST and StemSLFST processed between 3281 and 5048 characters/ms and are thus comparably fast as CST – see Figure 6.14a. HMT is about three times faster at the expense of memory consumption. All tested recognizers are by orders of magnitude faster than basic NLP pipelines: *Apache OpenNLP* [Sw16] and *Stanford CoreNLP* [Sw113] were tested using a basic pipeline consisting only of a tokenizer, sentence splitter and part-of-speech (POS) tagger. Although no NER-specific analyzers like noun chunkers or type classifiers were added yet, their processing time was already out of our targeted range. Running the basic pipeline on all 3.9 million articles would presumably have taken about 18 days in the case of CoreNLP, for example.



(a) processed characters per millisecond



(b) memory usage after feeding all German Wikipedia labels [Sw122]



(c) runtime to process all German Wikipedia articles [Sw122]

Figure 6.14: IT-RTC-OBNER Prototype I: Memory and Runtime Performance.
Comparing *LemMLFST* to *StemSLFST*, *HMT/CST* [93, 269], *Apache OpenNLP* [Sw16] and *Stanford CoreNLP* [Sw113] with regard to runtime and memory performance. Used hardware: Intel Core i7-4910MQ 2.9 GHz PC, 16 GiB RAM, Windows 7 64-bit [Sw83].
(Note: Variants of these figures have been published in Jilek et al. [186].)

Conclusion. IT-RTC-OBNER Prototype I (LemMLFST) was shown to be comparably fast as available high speed methods while outperforming them in the recognition of terms that lexically vary slightly as induced by inflection. In particular, the false positive rate of LemMLFST was not considerably higher – in some cases even lower – than the one of other recognizers. As a consequence, it was possible to narrow the quality gap to more sophisticated but also much slower NLP pipelines a bit more without losing real-time capable runtime performance. However, while LemMLFST outperformed StemSLFST by 14% in terms of precision for single-word terms, StemSLFST was slightly faster and better in both, precision and recall, for multi-word terms. Thus, a new configuration of IT-RTC-OBNER's Multi-FST has been created after the evaluation to combine the best of both approaches. This configuration is presented in the next section.

6.2.4 IT-RTC-OBNER Prototype II with Post-Evaluation Improvements

Given the results of the evaluation of Prototype I (see last section's conclusion), a new configuration has been created. It uses three types of FSTs:

- **Basic FST**: A single-layer FST only using the character layer depicted in Figures 6.7 and 6.8. It only detects exact matches of fed labels and its only configuration option is enabling or disabling case-sensitivity.
- **Lemmatizing FST**: If a term is added to this FST, it looks up or tries to infer the term's inflected forms and also adds those. For multi-word terms, it would typically be used as a multi-layer FST to keep memory consumption low. This version is identical to LemMLFST used in the evaluation. However, for single-word terms, it may also be used with just a single layer (the character layer).
- **Stemming FST**: This FST is basically the same as the lemmatizing one but uses stemming instead of lemmatization. In the evaluation, it was used with a single layer and called StemSLFST. For multi-word terms, it could also be used with multiple layers, but the memory saving is far less considerable as in the lemmatization case since the stemming reduces several variants to their identical stem.

Figure 6.15 shows the process of adding a new term to the newly configured Multi-FST illustrating which term characteristics are checked in order to find the appropriate FST the term is then assigned to. In a first step, it is checked whether the term is an acronym. If that is the case, it is assigned to a *Basic FST* applying an exact match strategy (additionally checking for uppercase-only characters). If the term is not an acronym, its language is detected, for example using the *Tika Language Detector* [Sw20]. Currently, only German and English are supported, and all FSTs except for the previously mentioned one (for acronyms) exist in a German and English version. All steps described in the following are thus performed for both variants according to the detected language. The next step is checking whether the thing the term is a label of is associated with high or low tolerance. In case of the latter, the term is assigned to a *Basic FST* having low tolerance as depicted in Figure 6.8. High-tolerance terms are either added to a *Lemmatizing FST* (see Figure 6.7) if they are single-worded or to a *Stemming FST* in the case of being a multi-word term.

This configuration has been used in the cSpaces multi-month user study as well as 24/7 in CoMem (including installations at the customer) for more than three years now.

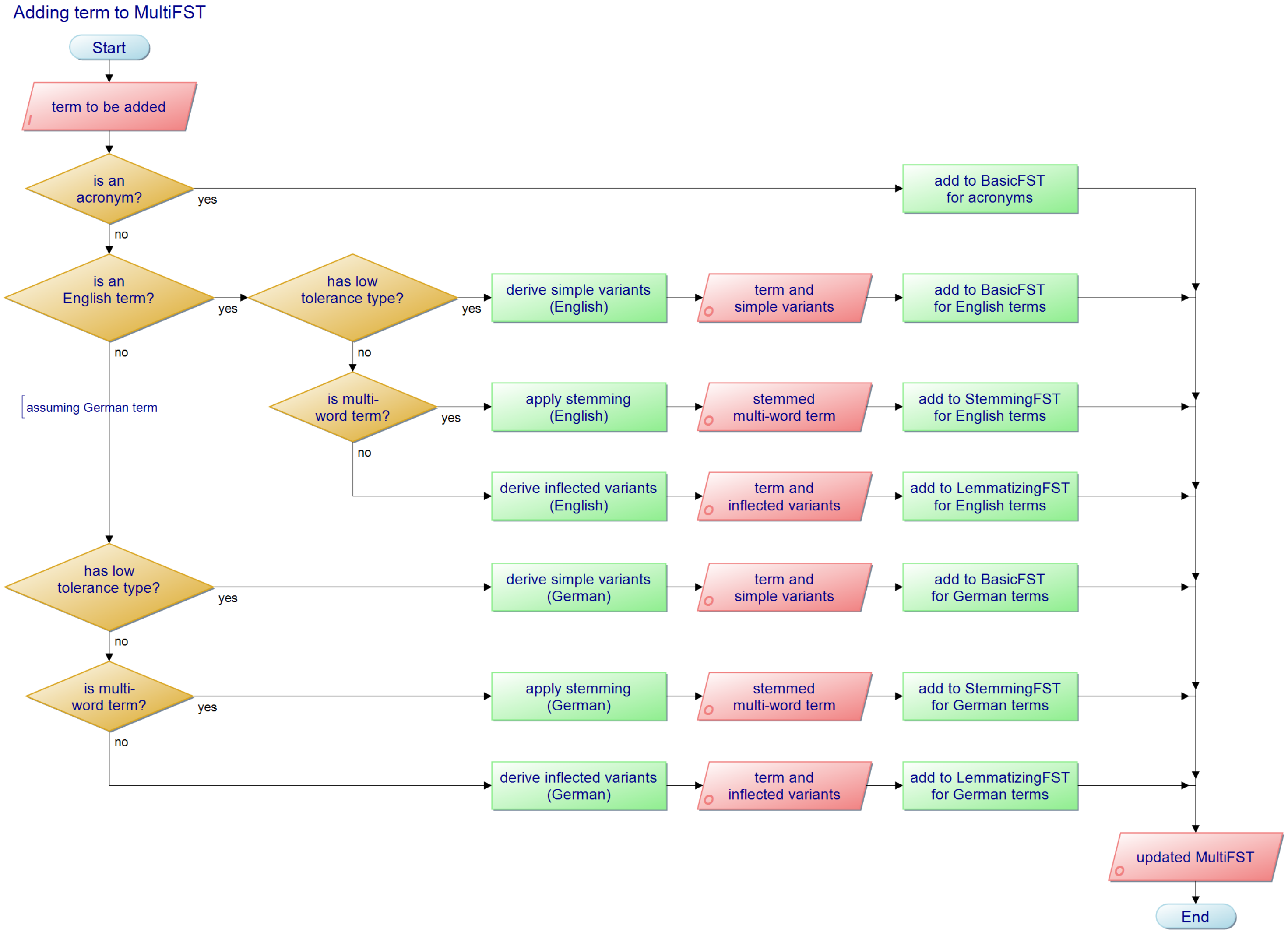


Figure 6.15: IT-RTC-OBNER Prototype II: Improved Multi-FST Configuration.
Characteristics to be checked when adding a new term in order to assign it to the appropriate FST: term being acronym, detected language (currently German and English are supported), term associated with high or low inflection tolerance and term being a single- or multi-word term.

6.3 Context Elicitation by User Activity Stream Processing

After evidence collection and information extraction, the next task to be performed (step 3 of the cSpaces interaction cycle) is the actual context elicitation, i.e. the interpretation of evidence snippets, in particular text and named entities mentioned in the text, to find contexts they may refer to or to decide whether a new context needs to be created (e.g. since the similarity to all existing ones is too low).

A continuous stream of evidence snippets is analyzed and every evidence snippet (basically mentioning an application and an information item) is assigned to a number of plausible contexts (each having an individual confidence score). The interpretation of currently observed evidence snippets may change after more subsequent evidence is collected. In Section 3.3.1, the example of two different car manufacturers was mentioned: tracked user evidence snippets could indicate that the user is definitely in a automotive context. It is unclear, however, to which of the two customers, or maybe even a third new one, the current activities are related. After observing a few more evidences, the situation gets more clear. Wrong decisions by the system could be reverted as mentioned with *context retraction*. However, a certain sensor and interpretation gap typically remains. Users should thus help the system by explicitly selecting in which context they are currently working or wish to work next. This is modeled by the user's *context awareness* [343]: It equals one directly after selecting a context and decreases towards zero with time passing by. Nevertheless, the system cannot fully rely on users since they may forget to signal a switch to another context or they may consider switching to be too expensive since they intend to just "shortly do something else", i.e. perform a so-called *tiny task* [343], before continuing in their actual context. That is why cSpaces needs to interpret the giving evidences with regard to the current context awareness and consider the possibilities of tiny tasks or forgotten context switches. Thus, associating context association statements with a certain confidence (see *ATIC characteristics*, Sec. 5.2.1), backtracking and (automatically or manually) revising system decisions is needed. A system that does not need to rely on user feedback at all and still finds "the right" contexts would be desirable but is out of this thesis' scope.

A prototypical implementation is presented in Sections 6.3.1 and 6.3.2 after shortly addressing related work in the following.

Related Work. Regarding related work, several approaches have been mentioned in Section 4.1:

- *TaskTracer* [97, 368, 350, 349, 347],
- solutions of the *APOSDLE* project like the *User Interaction Context Ontology (UICO)* [282] and works on user activity tracking and task detection [284, 282, 281, 283, 90],
- works by Gyllstrom et al. on activity tracking and task detection [150, 151],
- the *ACTIVE* project's work on context/task mining approaches [363, 362] based on their context model [111],
- works on context/task recognition [210, 209, 214, 301, 298, 297] in the *SWELL* project,

- *DCON*, *PiMx(T)* and *(X-)iDeTaCt* [311, 5, 4, 3] by the group of Abela, Staff, Handschuh, Scerri et al.,
- works by Schmidt, Stoitsev, Mühlhäuser et al. on task detection and mining [319, 318] and activity-centric document clustering [345] and
- *UserContext* [Sw106], a tool collection by DFKI and research partners, for example containing the *User Observation Hub* [Sw108] and *DragonTalk* [Sw107].

The approaches above typically **differ from cSpaces** by using a different context/task model, tracking different activities or using different sensors, seeing contexts/tasks more static (in cSpaces, associated tags and contained items as well as the contexts themselves can be added/created or deleted at any time) or looking at a fewer number of contexts (in the evaluation).

These differences vary from approach to approach with some dimensions being closer to cSpaces than others. They are, however, rather minor in comparison to the **main difference** of cSpaces to all approaches above, which is its characteristic of being a self-organizing and especially forgetful information system. Each context could "sink away" into the forgotten area or it could be brought back up by its memory buoyancy rising again. When assigning activities to these contexts, their buoyancy needs to be taken into account. In addition, cSpaces are seen as an explicit interaction element that should be used actively and constantly. Discarding this explicit feedback coming from an essential feature of the system for a comparison with other systems was not considered an option, especially since the forgetting aspect hinders a meaningful comparison anyway.

Thus, in order to cope with the scenario of self-organization and especially forgetting as well as having explicit user feedback as another important input source, an own approach has been developed in this PhD project. In the following, the approach is referred to as the *Stream Context Miner* in order to set it apart from the retrospective context miner, called the *Contextifier*, presented in Section 5.3.3. The Stream Context Miner was developed as part of the master thesis by Jessica Chwalek [65]. In her project, two solution methods were investigated: a mostly exact one (if a branching heuristics works well) and a more randomized one. Both are presented in Section 6.3.2. However, to keep the amount of work appropriate for a master thesis, she could use a pre-analyzer developed by the author that is introduced first.

6.3.1 User Activity Stream Pre-analyzer

In the development and evaluation scenario of the Stream Context Miner, each context (space) had a number of tags (*context is about*), contained elements (*context contains*) and a history list of activities (*context's activity history*). The former are depicted in Figure 8.25 (highlightings F and G), while the activity history is shown in Figure 8.15 (highlighted section 3), for example. The main purpose of the pre-analyzer was to provide a rough estimation of plausible context candidates for each observed evidence snippet in only a few milliseconds. It was assumed that the main application would be able to invest more time and effort into analysis. For example, the pre-analyzer only uses bit vectors, simply indicating whether a word or concept occurs or not, whereas the main application can operate on float values thus obtaining a much more fine-grained analysis result.

In a similar, text-focused scenario, Schmidt et al. [319] reported good results using the **Vector Space Model** [295], which was in line with the author's own experience, for example with *PIMO Diary* [184] (see Sec. 8.4.1). cSpaces indexes every thing of the user's PIMO as well as every word it ever processes with any information item. Each context space S therefore basically has "isAbout" (A) and "contained items" (I) vectors: $V_A(S)$ and $V_I(S)$. For each of them, there is actually one for the textual content (T) and one for the concepts (C), i.e. the named entities found in the text. This results in four vectors: $V_A^T(S)$, $V_I^T(S)$, $V_A^C(S)$ and $V_I^C(S)$.

For the scenario, evidence snippets were unified to have a title (L) and a body (B), just like a website. For a file, these were the full path as the title and its content as the body. Mails had the subject as a title and also their content as the body. Again, there were separate vectors, $V_L^T(E)$ and $V_B^T(E)$ for text and $V_L^C(E)$ and $V_B^C(E)$ for the concepts.

Based on these vectors, the pre-analyzer calculated the similarity of a context space S and an evidence snippet E as follows:

$$\begin{aligned}\mathrm{sim}(S,E) = 10 \cdot [\mathrm{sim}(V_A^C(S), V_L^C(E)) + \mathrm{sim}(V_A^C(S), V_B^C(E))] + 5 \cdot [\mathrm{sim}(V_I^C(S), V_L^C(E)) + \mathrm{sim}(V_I^C(S), V_B^C(E))] \\ + \mathrm{sim}(V_A^T(S), V_L^T(E)) + \mathrm{sim}(V_A^T(S), V_B^T(E)) + \mathrm{sim}(V_I^T(S), V_L^T(E)) + \mathrm{sim}(V_I^T(S), V_B^T(E))\end{aligned}$$

with a well-known similarity measure derived from the *Vector Space Model* [295], i.e. the dot product of both vectors divided by their norms:

$$\mathrm{sim}(V,W) = \frac{V \cdot W}{||V|| \cdot ||W||} = \frac{\sum_{i=1}^{n} v_i \cdot w_i}{\sqrt{\sum_{i=1}^{n} (v_i)^2} \cdot \sqrt{\sum_{i=1}^{n} (w_i)^2}}$$

As shown in the formula, the context's *isAbout* concepts were considered most important: their similarity to the evidence snippet's concept vectors for title and body was weighted with a factor of 10. The *contained items* concepts were the second most important characteristic: their similarity to the evidence snippet's concept vectors for title and body were weighted by 5. All other similarities were based on text instead of concepts and added with a weight of 1. This follows experiences made with *PIMO Diary* [184], a retrospective context clustering application (see Sec. 8.4.1). Note that these weights were not optimized, they simply were an educated guess. Before putting the pre-analyzer to permanent productive use, the author would thus spend additional effort on its further elaboration and optimization. In particular, testing in Big Personal Data scenarios is still missing.

With the pre-analyzer as a support tool, the Stream Context Miner could perform the main task as discussed in the next section.

6.3.2 Context Miner for User Activity Streams

Problem and Approach. The problem of mapping evidences to plausible contexts can be modeled as a tree as shown in Figure 6.16.

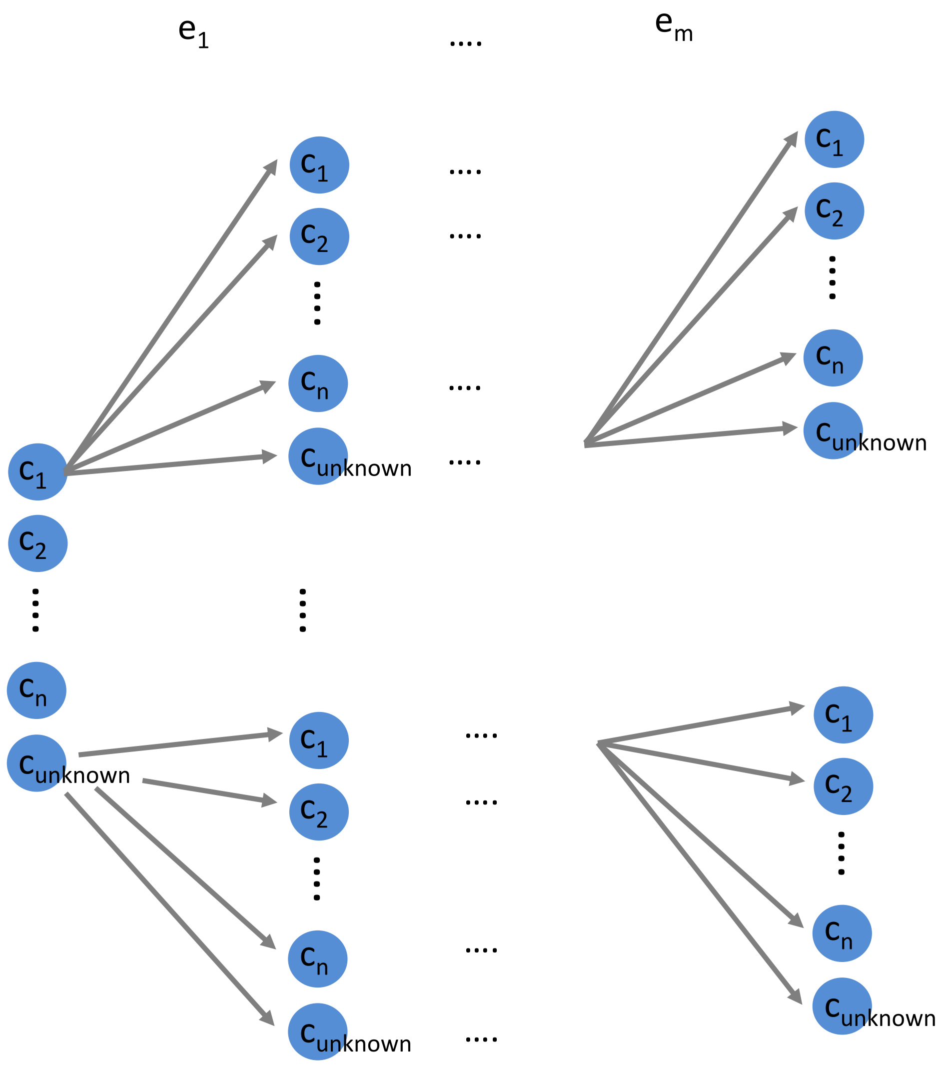


Figure 6.16: Stream Context Miner: Mapping Problem Modeled as a Tree.
Notations: c_i: *contexts* / e_j: *evidences*.
(Note: This figure was created based on supplementary material of Chwalek [65].)

In each state (drawn as columns), the user is in one of their n contexts (or an unknown/new one). After the processing of the next evidence snippet e_i, they are again in one of these n contexts or an unknown/new one. For example, consider a user being in context c_1. If the next evidence snippet e_1 suggests that a context switch to c_2 has happened, the path through the emerging tree would be $c_1 - e_1 \rightarrow c_2$. If the next snippet e_2 still suggests c_2 as the current context, the path is $c_1 - e_1 \rightarrow c_2 - e_2 \rightarrow c_2$ and so forth. Each edge is associated with a certain weight stating how plausible following that path would be given the previous and the next context. For example, switching contexts within only a few seconds or even less would result in a very expensive path through the tree since such a situation is typically not or very rarely seen in real-world scenarios. An example of such a scenario could be a person clicking several emails of their inbox, each being associated with a different context. Still, a dwell time of less than a second suggests that a person did not really perceive much of the actual content. Too sudden "jumps" between different contexts is just one example. More heuristics were applied to associate rather "natural" flows of events with lower weights and typically more rare ones with higher weights. Weight estimations were influenced by experiences of the student, the author, other colleagues at DFKI SDS as well as papers like González and Mark [137]. Additionally, events like creating a new context, adding/removing a tag or information item to/from a context or clicking one of its elements could be used as implicit feedback greatly helping in pruning the mapping tree. Typically there were only very few or just one path left to follow for such a given evidence snippet.

With such a model, the evidence to context mapping of an observed event stream can then be seen and solved as a **shortest path problem** [321] or a *top-k shortest path problem* if not only the best solution should be considered but other similarly good ones as well.

From the above model, one may easily derive that such a tree grows exponentially. Therefore, the approach by Chwalek [65] takes countermeasures: A **first solution method** basically tries to find an exact solution but uses a heuristic to dynamically prune the ever-expanding tree. It is therefore called **Dynamic Tree Algorithm (DTA)** in the following. One design decision was, if the user at some point confirms a certain context mapping, it should not be reverted by the system. Additionally, if longer breaks (phases of user absence) occur, reverting old mappings should also not be performed. The Stream Context Miner is meant to be integrated into a real-time user support system, reverting old decisions possibly hours, days or weeks ago is not its purpose. It is more focused on the current situation, taking event stream windows of only a few minutes into account. Thus, after a certain time previous mapping results become "frozen" and the system may rebase itself on last constellation of possible contexts starting a new, empty observation window waiting for further user evidence snippets to arrive. Since Jessica's bachelor thesis [63] was the extension of *Memory Buoyancy* (see Sec. 7.2.2), she could use this experience to additionally establish a similar mechanism in the Stream Context Miner. A list was kept showing which contexts were stimulated recently and which ones started "sinking away". This *memory list* influenced the ranking of plausible context candidates and thus weight estimations, branching and pruning decisions.

A **second solution method** is based on **Genetic Algorithms (GA)**. Its path evaluation is basically the same as the one of DTA. However, instead of constructing the tree step by step, it rather randomizes solutions in alternating phases of exploration and intensification typically known from genetic algorithm approaches. Only the fittest individuals of a generation enter the current population, new generations are created by cross-overs, mutations etc. and in the end, the top-k best individuals (solutions) are used for further calculations or the actual mapping.

Evaluation. To evaluate both solution methods, data of the first few days of the cSpaces multi-month user study was used. In this early phase, five participants had started using the app in their daily work. Statistics about the number of created contexts, tracked events, etc. can be found in Table 6.2.

Participants could give explicit feedback about each of their activities by removing items from the context's activity history signaling that this event did not belong to the current context. If they did not remove an activity, it was assumed that it actually belonged to that particular context. The event to context mapping obtained that way could be used as ground truth for the evaluation. Note that all solution methods did not have access to event snippets involving a context's activity history, those events were filtered out. Participants submitted their collected data after anonymization as described in Sec. 6.1.2. Using the ATIC feature of *time-dependency* (see Sec. 5.2.1), all KGs could be set back in time and be queried event-by-event. The Stream Context Miner's evaluation could thus be performed as if events would happen in real-time. The performance of DTA, GA and also the pre-analyzer introduced in the previous section (6.3.1) was assessed by accuracy, which is defined as follows:

$$\text{accuracy} := \frac{\text{number of correct predictions}}{\text{total number of predictions}}$$

In the following, especially *top-k accuracy* is used, which states whether the correct prediction is among the top-k of suggested candidates.

All computations were performed by Jessica Chwalek on her private computer, a standard notebook without any special or particularly powerful hardware to mention.

The **mean processing time per evidence** (overall) was 0.66 ms for the pre-analyzer and 16.53 and 14.97 ms for GA and DTA, respectively. This especially means that one of the master thesis' secondary goals, having (near-)real-time capable solution algorithms, has been fulfilled.

With regard to **top-1 accuracy**, the pre-analyzer has the highest overall performance of 0.676. The main algorithms achieve slightly lower results. However, the advanced calculations by these methods pay off when looking at the **top-2 accuracy**: Here, they outperform the pre-analyzer, in some cases by large amounts if looking at the scores of participants P2 and P3, for example: 0.553 (Pre) vs. 0.821 (GA). As discussed earlier, the pre-analyzer heavily relies on text and concept annotations. P2 and P3 having the lowest mean number of "contained items" thus resulted in poor results. The main algorithms performed better taking more aspects (as described before) into account. Overall scores were 0.85 (GTA) vs. 0.78 (Pre). To put it in other words: using GTA, for example, the "right" context was among the top-2 solutions in 85% of cases. In the GUI, the system could thus show an option offering the user to switch to the second best candidate thus correcting a possibly wrong suggestion with a single click.

The number of times each method voted for an unknown context is, in the case of GA/DTA, close to the number of contexts each participant has created. Further explore the "unknown" classification as an indicator to create a new context is thus a potential topic to investigate in future work.

In Chwalek [65], several views to monitor the algorithms and their performance were created. There is the *Step by Step Solution Browsing View* (Figure 6.17), the *Algorithm Comparison View* (6.18), and the *Processing Time View* (6.19).

		P1	P2	P3	P4	P5	Total
num. of evidences		1068	507	466	194	241	2476
num. of contexts		11	9	7	8	8	43
num. of contexts' *is-about* tags	min	0	0	1	0	1	0
	max	6	6	6	6	6	6
	sum	40	25	19	6	13	103
	mean	3.64	2.78	2.71	0.75	1.63	2.40
num. of items *contained-in* contexts	min	0	0	0	1	2	0
	max	28	7	6	11	8	28
	sum	59	23	16	35	33	166
	mean	5.36	2.56	2.29	4.38	4.13	3.86
mean processing time per evidence [ms]	Pre	0.36	0.40	0.25	1.50	0.79	0.66
	GA	28.17	18.27	15.52	10.64	10.05	16.53
	DTA	38.89	6.67	3.71	8.60	16.98	14.97
top-1 accuracy	Pre	0.824	**0.475**	**0.461**	0.777	**0.759**	**0.676**
	GA	0.781	0.422	0.418	0.649	0.627	0.579
	DTA	**0.809**	0.456	0.416	**0.778**	0.680	0.628
top-2 accuracy	Pre	0.869	0.553	0.794	0.787	0.817	0.780
	GA	**0.941**	**0.821**	**0.884**	0.799	0.805	**0.850**
	DTA	0.868	0.554	0.511	**0.804**	**0.880**	0.723
num. of *unknown* classifications	Pre	40	4	26	1	1	72
	GA/DTA	15	9	8	0	3	35

Table 6.2: Stream Context Miner: Computational Results as Reported in Chwalek [65] and Complemented by the Author (last row).
Abbreviations: *num. : number / Pre: Pre-analyzer / GA: Genetic Algorithm / DTA: Dynamic Tree Algorithm.*

Conclusion. The Stream Context Miner can be used by cSpaces to map observed evidence snippets to plausible context candidates and, as a consequence, adjust its support measures accordingly. In the study, the "right" context was among the top-2 predictions in up to 85% of cases. However, there is still potential for improvement. Additionally, exploring whether the "unknown" classification can be used as an indicator for creating a new context (i.e. context spawning, see Sec. 3.3.1) is a topic to investigate in future work.

The evaluation has also been a first demonstration of ATIC capabilities (see Sec. 5.2): By setting the KG back to arbitrary points in time, the Stream Context Miner could be further improved (algorithm variants tested, bugs fixed, etc.) without redoing the study. Additionally, the whole scenario of each user could be processed as if it was live.

As mentioned earlier, Jessica also worked on Memory Buoyancy, which is a form of information value assessment (step 4 of the cSpaces interaction cycle) and the next chapter's topic.

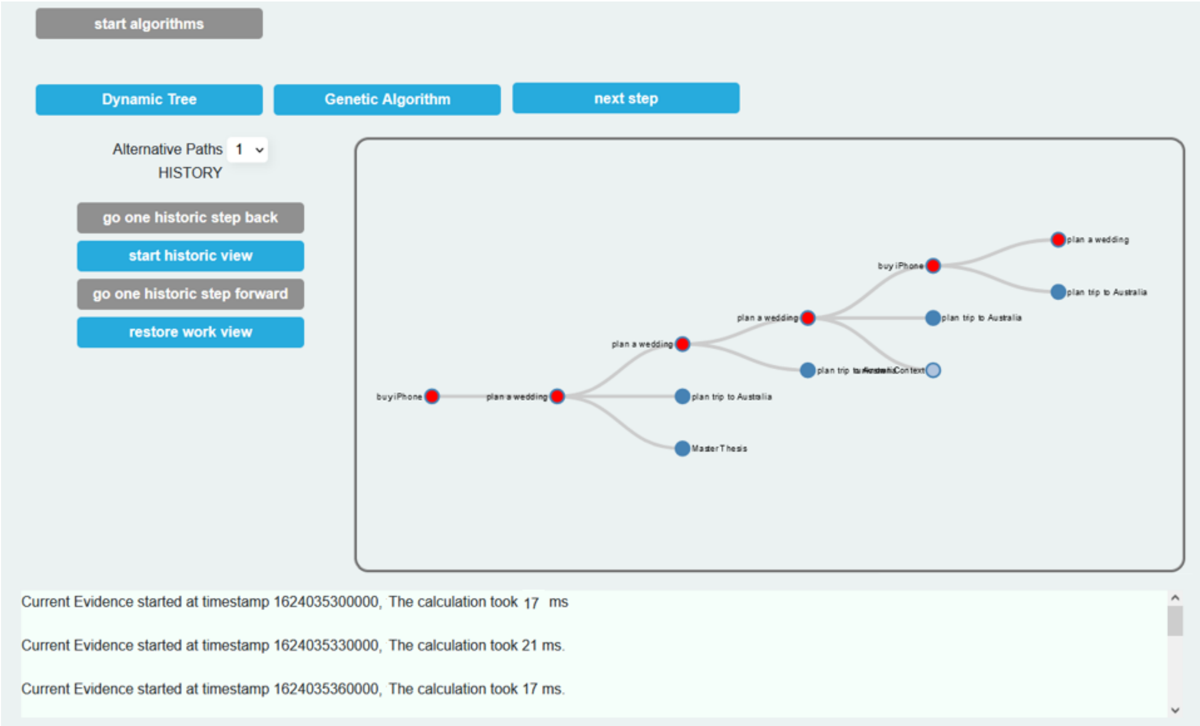


Figure 6.17: Stream Context Miner: Step by Step Solution View Realized in Chwalek [65]. This view allows for expanding the events to context mapping tree step by step. (Image source: Chwalek [65], whitespace compacted for better readability.)

Dynamic Tree Accuracy Value Top1: 62.5% Accuracy Value Top2: 75%				Genetic Algorithm Accuracy Value Top1: 20% Accuracy Value Top2: 20%	
Top 3	Scores	Ground Truth	Evidence Id	Scores	Top 3
1. 465 2. 465 3. 465	465: 1	465	20210811233245370	465: 0.71428573 460: 0.2857143	1. 465 2. 460 3. 465
1. 465 2. 465 3. 465	465: 1	465	20210811233253252	465: 1	1. 465 2. 465 3. 465
1. 465 2. 465 3. 465	465: 1	465	20210811233346730	465: 0.85714287 464: 0.14285715	1. 465 2. 465 3. 464
1. 465 2. 465 3. 465	465: 1	465	20210811233417336	465: 1	1. 465 2. 465 3. 465

Figure 6.18: Stream Context Miner: Algorithm Comparison View Realized in Chwalek [65]. This view allows for comparing DTA (left) to GA (right) with the given evidence snippet's ID as well as the ground truth (stated as the context's KG ID, e.g. 465) in the middle. (Image source: Chwalek [65], image cropped and whitespace compacted for better readability.)

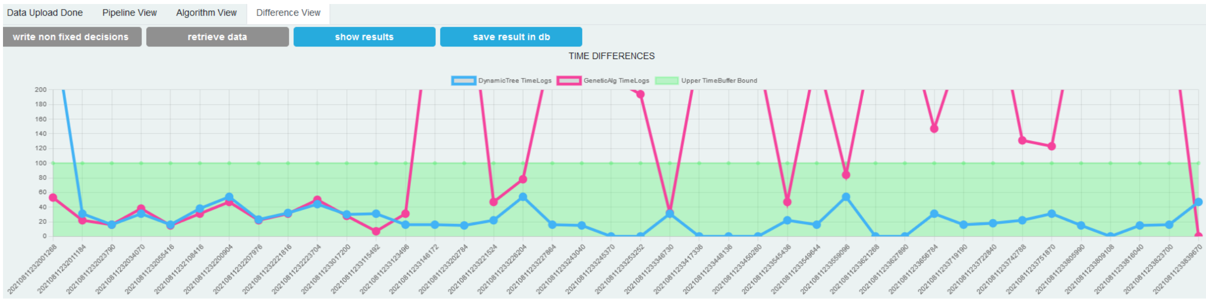


Figure 6.19: Stream Context Miner: Processing Time View Realized in Chwalek [65].
Colors: *blue graph: dynamic tree algorithm / red graph: genetic algorithm / green area: 100 ms range*.
(Image source: Chwalek [65]. Note: The depicted processing times are not those of the final version, whose results were actually lower [i.e. faster processing].)

7 Context-sensitive Information Value Assessment

This chapter is about *Information Value Assessment (IVA)*, step 4 of the cSpaces interaction cycle. Apart from giving an overview of existing IVA approaches (Sec. 7.1), *Memory Buoyancy* (7.2), a measure for short- and medium-term IVA as well as the *Preservation Value* (7.3), its long-term counterpart, are introduced.

Note: Parts of this chapter have already been published [182, 240, 245, 246, 407].

Memory Buoyancy is a form of *Information Value Assessment (IVA)*, to the author's best knowledge, the only one designed and implemented according to findings of cognitive psychology about human memory and cognition. It especially takes a wide range of evidence into account, e.g. a person's files, mails, bookmarks, calendar entries, tasks, etc. and, on a higher level, episodes/contexts of a person's life (as far as they are represented in a machine-understandable way). However, there are also other approaches mentioned in literature. A selection of them is briefly presented in the following.

7.1 Existing Information Value Assessment (IVA) Approaches

Wijnhoven, Amrit et al. investigated identifying and dealing with *information waste*, for example in the file system [397] or on the internet [11].

Hasan and Burns [156] speak of *waste data*: they analyzed the impact of such data on computing environments and proposed solutions on how to cope with it.

Turczyk et al. [379, 382, 381, 380] investigated *file valuation* in the area of *Information Lifecycle Management (ILM)*, which seeks to store files on different storage systems according to their (business) value.

Gyllstrom and Pedersen [149] proposed *LostRank*, an approach to estimate which documents are most likely to be lost for the user (e.g. important but not recently used) and are thus harder to re-find.

Sappelli et al. [300] presented an approach for email importance estimation. Their model could predict which emails will be replied to with a 72% accuracy. They had doubts whether this value would be high enough for a standalone application but it may nevertheless serve as an email importance indicator.

Milo [256] discussed "the logical, algorithmic, and methodological foundations required for the systematic disposal of large-scale data, for constraints enforcement and for the development of applications over the retained information".

Referring to the mostly synonymously used term of *data value assessment*, Attard, Brennan et al. provided, for example, a *Data Value Vocabulary* [17, 19] and an overview of further approaches [47], while also discussing related topics like *data value chains* [46] or *value-driven data governance* [18].

Some of the approaches above are more business-centric, others aim at more personal use cases. Although applications in business use cases are possible and investigated, the focus of *Memory Buoyancy* and *Preservation Value* so far has been a bit more on individual-centric use cases. They especially were designed and implemented according to findings of cognitive psychology about human memory and cognition. Both are introduced more thoroughly in the remainder of this chapter.

7.2 Memory Buoyancy (MB) – Short- and Medium-term IVA

Note: Section 7.2 is in large part a revised version of Jilek et al. [182, Sec. 3 and 4].

Section 3.1 introduced the idea of *Managed Forgetting* with Preservation Value and especially Memory Buoyancy (MB) as two of its cornerstones. The introduction also mentioned the high-level idea of MB, which is items losing relevance gradually "sink away", while those that are important (again) are pushed closer to the user by their higher buoyancy (see Fig. 3.1).

This section consists of three parts: First, in Section 7.2.1, an initial version of MB (also referred to as *MB Prototype I*) is presented especially covering general design principles. Its major drawback was the missing support for different user contexts. Thus, a revised version, *Prototype II*, is introduced in Section 7.2.2. Last, Section 7.2.3 discussed further ideas for a third prototype, especially with regard to (big) personal data processing and corporate scenarios.

7.2.1 MB Prototype I: Memory Buoyancy in CoMem

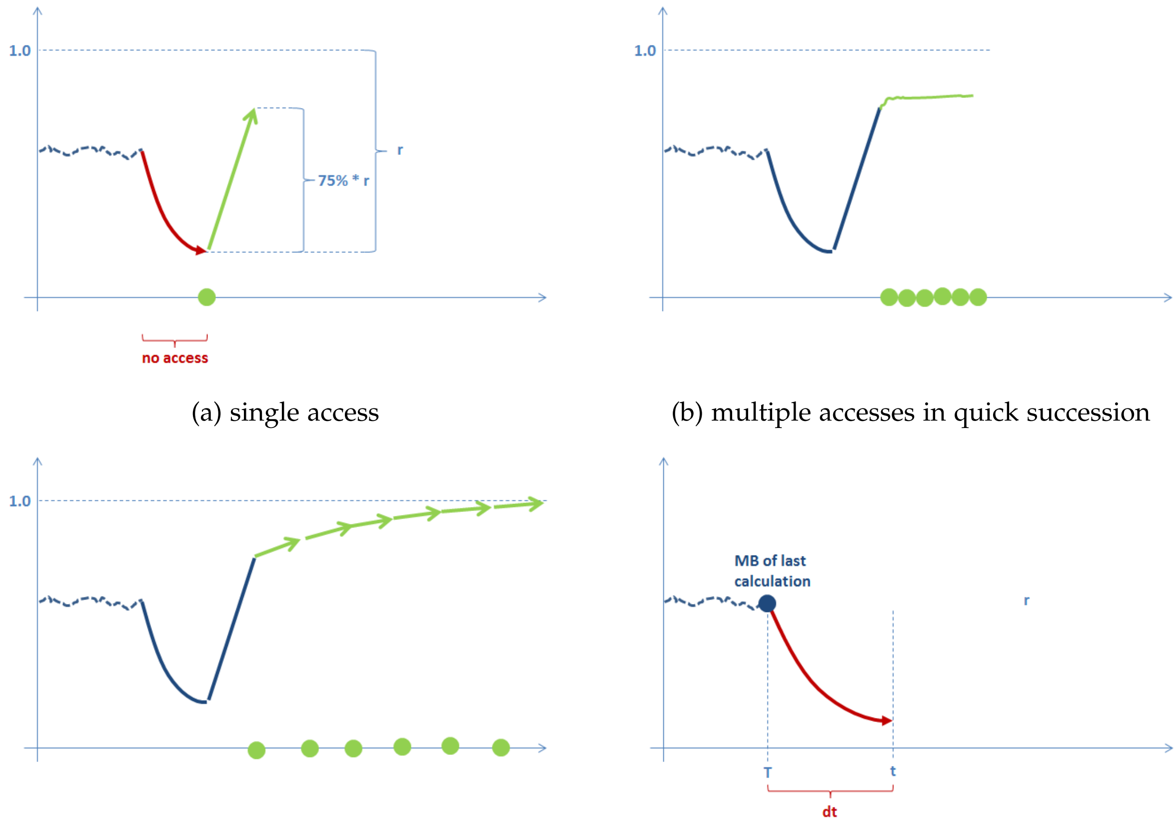
An initial version of MB was realized in CoMem led and mainly developed by Sven Schwarz with only minor contributions by the author. Its calculations evolved from insights of Tran et al. (2016) [376] and finally follows design principles presented in the following. It has been evaluated in a dedicated study presented in Tran et al. [376] and in 24/7 usage in the DFKI SDS department for more than five years. The basic principles were inspired by human brain activity applied to the user's mental model as represented in a semantic graph (PIMO) and discussions with the team of Prof. Robert Logie[1], who presented their insights in Logie et al. [230].

Memory Buoyancy Design Principles. Every thing in a user's PIMO has its own MB value.

- MB is updated every time a thing is stimulated (which in worst case could be after each click of a user).
- The **strength of a stimulation** depends on:
  - user interaction (e.g. viewing, modifying, annotating),

[1]Human Cognitive Neuroscience Group, Psychology, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Scotland (https://www.ed.ac.uk/ppls/psychology/research/areas/human-cognitive-neuroscience)

  - the thing itself (e.g. mails are assumed to be more ephemeral than a presentation) and
  - its connections in the semantic network (PIMO).
- MB calculation has to cope with **activity bursts as well as erratic accesses**.
  - MB values are normalized ($0 \leqslant MB \leqslant 1$).
  - Single access of an item should not directly lead to an MB of 1.0. See Figure 7.1a: user access is depicted as green dots on the x-axis and the resource's MB value is shown on the y-axis. A single access after some time does not lead to a value of 1.0 but rather 75% of the difference between the resource's former MB value and 1.0.
  - Multiple accesses in quick succession (e.g. every minute) are treated reluctantly, i.e. the individual events are rather treated as a single one, so the MB value does not quickly escalate to 1.0 (see Figure 7.1b).
  - Over day, multiple accesses will saturate against 1.0 (see Figure 7.1c).



(a) single access

(b) multiple accesses in quick succession

(c) multiple accesses every day

(d) static and dynamic parts

Figure 7.1: Memory Buoyancy Prototype I: Selected Design Principles.
(Note: These are figures drawn by Sven Schwarz already published in Jilek et al. [182].)

Forgetting is realized by orienting on human brain activity on the mental model as mentioned above.

- MB drops for things that are not stimulated:
  - first, there is a steep decline, then
  - a long-tail of slow decline.
- MB increases for things that are stimulated.
- MB decreases slower for things that are repeatedly stimulated over time (learning effect).
- Associations to accessed things are stimulated as well.

Rules and heuristics were added to deal with requirements of **various domains** like PIM:

- upcoming events should stimulate connected things,
- finished items (tasks, events) shall decrease faster unless other indicators speak against this,
- times with low user interaction should not lead to massive decay in MB

MB utilizes the semantic network (PIMO): a form of **spreading activation** [70] is performed (see Fig. 7.2) using additional heuristics for

- types (e.g. emails are faster forgotten than persons), or
- the number of relations that connect things.
- Additionally, relation types could also be considered (e.g. a project potentially has many members but just one project leader).

Figure 7.2: Memory Buoyancy Prototype I: Spreading Activation.
The thing directly impacted by a stimulation is highlighted in dark red. The activation then spreads with a bit of decay towards its two neighbors (slightly lighter red) and from there to two more neighbors each (pink). Again, a bit of decay is induced resulting in the activation to stop after this step. Thus, three nodes of this example network (upper right) were not activated.

In addition to these conceptional design principles, there are also **technical requirements**:

- MB is **time-dependent** which poses a challenge for calculation as MB values change permanently just as time passes, i.e. decay (see Figure 7.1d).

  - Conceptually splitting MB value in two parts: a static (time-independent) and a dynamic (time-dependent) part.
  - Database stores most recently calculated MB value and calculation timestamp.
  - Decay happens at retrieval time: value stored in database is "decayed" according to time difference between retrieval time and last calculation time.
- MB values must **always be up-to-date**.
  - Use an online calculation (instead of offline-like overnight calculation).
  - Accessing PIMO things must trigger re-calculation of MB involving their semantic neighborhood to also identify hot regions.
- Keep **high performance** despite additional MB calculation.
  - Requesting MB values is done very often, therefore, retrieval must be fast (processing high-volume access event streams is not feasible at runtime).
  - Minimize database updates: updating MB values for all things in the PIMO is not feasible.
  - Incremental updates: only re-calculate the most minimal set of PIMO things.

The core calculation of MB in CoMem comprises about 1000 lines of *Java* [Sw95] code. There are various methods taking type-specific aspects (as described above) into account, for example with regard to events. Without going into details, an interesting value to be mentioned here is the **decay** since it is used in many of these methods. If a resource is not stimulated anymore, its old MB value is multiplied by a time-dependent decay factor to obtain the new value. After multiple iterations of tweaking over the years, MB is currently calculated as follows:

$$b_{t_{n+1}} = b_{t_n} \cdot \frac{2 \cdot 10^9}{2 \cdot 10^9 + (t_{n+1} - t_n)}$$

Note that the times t_{n+1} and t_n are given in milliseconds "since the standard base time known as 'the epoch', namely January 1, 1970, 00:00:00 GMT" [276]. The previous MB value b_n as well as the timestamp of its calculation t_n can be found in the database (for each resource individually). After the calculation, this database entry is replaced by the new MB value $b_{t_{n+1}}$ and its timestamp t_{n+1}. The decay factor (see fraction in the formula) makes MB drop about 25% after a week of no stimulation, about 40% after two weeks and about 55% after four weeks.

Context-insensitivity Problem. As mentioned in the beginning of this section, the major drawback of MB Prototype I was the missing support for different user contexts. This problem is illustrated in the following using an example: Figure 7.3 (left-hand side) shows the context of a *Trip to Rome in July 2018*. Its elements are ordered by descending buoyancy. Looking at these elements from low to high MB, the context's "golden thread" could be as follows: The person whose contexts and PIMO are depicted here, shortly called "the owner" in the following, and a colleague, Peter Stainer, planned to meet in Mannheim and then travel on to Rome. An email

Figure 7.3: Memory Buoyancy Prototype I: Context-insensitivity Problem.
(Note: This figure has already been published in Jilek et al. [182].)

was written asking in which hotel to stay. This email was answered a bit later. Then, the owner booked a ticket with Deutsche Bahn, checked two different hotels and presumably booked one of them. At some point, the owner switches to another context to plan a different meeting (middle section of the figure). By chance, Peter Stainer is also an attendee of that meeting, the location is Mannheim and they again travel by train (Deutsche Bahn). Thus, elements also present in the first context (see red highlighting) are now stimulated. When later revisiting the *Trip to Rome* context (right-hand side of the figure), a context-free MB (like the one of MB Prototype I) would destroy the context's "golden thread" by re-ordering the elements although there was no action in that particular context that would justify this.

To tackle this problem, a revised version, MB Prototype II, has been developed and is presented in the next section.

7.2.2 MB Prototype II: Context-sensitive Memory Buoyancy

The major enhancement to be incorporated into a successor of MB Prototype I was taking user context into account. This especially also included context switches: As an addition to the aforementioned gradually changing MB, another value applicable in scenarios of sudden changes (like context switches) was required. With this goal, the current section addresses research question RQ6.

Local Memory Buoyancy. For the above reasons, a *Local MB* representing a resource's current value for a specific user in a specific context was introduced. Stated more formally, this results in:

$$f_{\mathrm{MB}_{local}}: (\text{ resource, user, context }) \rightarrow \mathrm{MB}_{local}$$

Figure 7.4 (top section, gray shapes) shows different contexts of two users (green person icon: user 1, red person icon: user 2). As an example, one may see that the website of Deutsche

Bahn is relevant for user 2 in the context of a Mannheim meeting as well as the Managed Forgetting project. Its MB varies in both contexts: 0.68 in the first and 0.38 in the second. The same website is also relevant for user 1 in the context of a Trip to Rome. user 1 is also involved in the Managed Forgetting context, a shared context, but for them the Deutsche Bahn website is not relevant there, indicated by an MB value of zero. This is an example to illustrate that the very same thing can be very important in one context while being totally irrelevant in another. Second, a certain thing can belong to a certain context (like the Managed Forgetting project), whereas its relevancy in that context varies from user to user (here: 0.00 for user 1 vs. 0.38 for user 2).

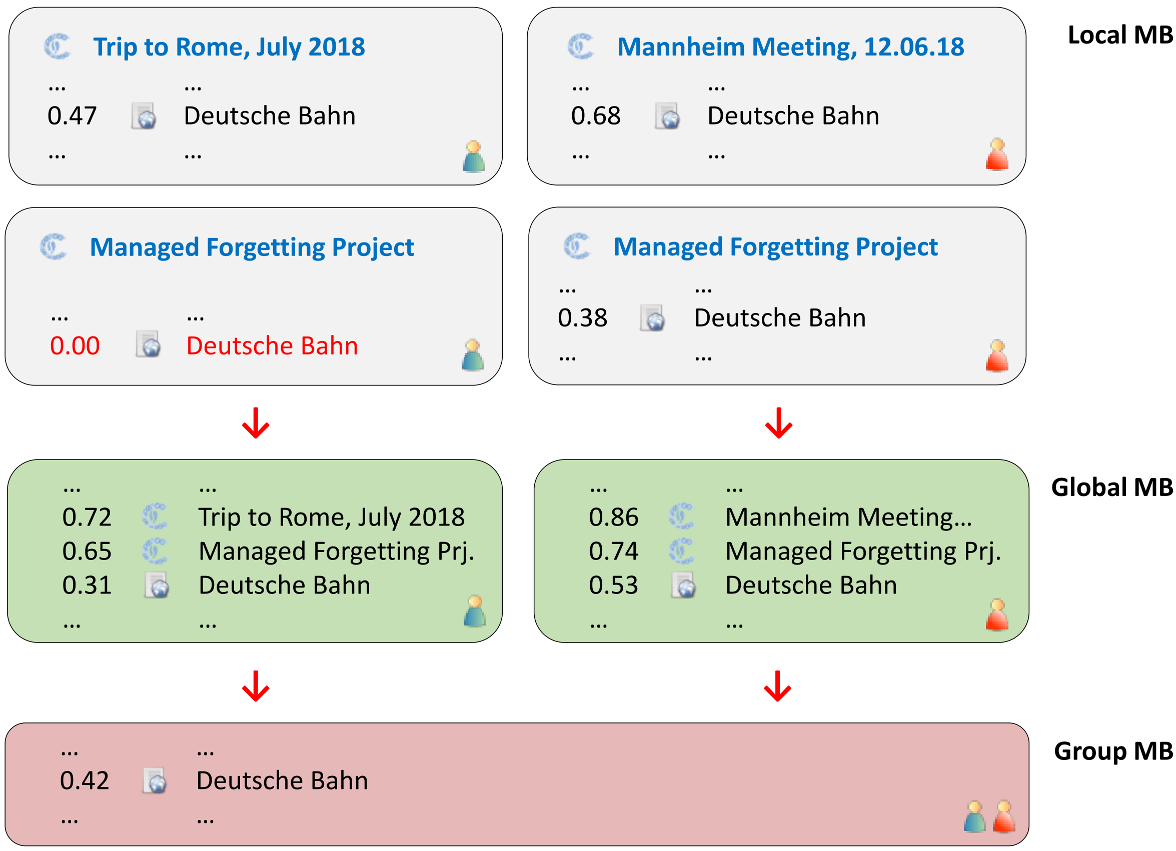


Figure 7.4: Memory Buoyancy Prototype II: Local, Global and Group Memory Buoyancy. (Note: This figure has already been published in Jilek et al. [182].)

Global Memory Buoyancy. The example showed that MB of a certain resource may vary from one context to another. For several applications (e.g. synchronization, archival), it is helpful to be able to give a summarizing statement about a resource's relevance. Thus, *Global MB* was introduced as an additional value providing a context-free relevancy information of a resource for a certain user:

$$f_{\mathrm{MB}_{global}}: (\text{ resource, user }) \rightarrow \mathrm{MB}_{global}$$

It is very similar to the initial, "classic" MB used in Prototype I and addresses less abrupt changes. In Figure 7.4, the green shapes in the middle section contain Global MB values. Note that each contexts also has a Global MB value. Thus, when leaving a certain context, the Local MB values are frozen and will not change anymore until the context is revisited and therefore active again. Nevertheless, the MB of the context as a whole, i.e. its global MB, may rise or drop. In a first implementation of MB Prototype II, the Global and Local MB values are calculated independently of each other. An idea for future would be however, to investigate the possibility of deriving the Global MB from all local values instead of assessing all user evidences multiple times (once for each MB type). The same is true for the Group MB, which is introduced next.

Group Memory Buoyancy. To especially address corporate memory scenarios, a *Group MB* value was additionally introduced, which represents a resource's relevance for all users of a group (red shape at the bottom of Figure 7.4):

$$f_{\mathrm{MB}_{group}}\colon \text{resource} \rightarrow \mathrm{MB}_{group}$$

Experimental Prototype. A prototypical implementation has been developed as part of the bachelor thesis by Jessica Chwalek [63]. Since all newly introduced MB values (also referred to as *advanced MB* values) are still experimental, the prototype also has an experimentation GUI depicted in Figure 7.5.

As an evaluation in Chwalek [63], typical knowledge work scenarios were modeled: a single person working on a task, a group of people collaboratively working on a task, whereas in some scenarios only some members of the group were active while others were just readers, a situation shortly before or after an event, etc.

In the first section of the experimentation app, one can choose a specific scenario. Each scenario was represented by a specific list of events, i.e. a generated list of user and system events, serving as input to the MB calculation (together with the semantic graph, PIMO, as the second input source). Such an event consists of the user's ID and context, which action was performed (e.g. accessing a document, adding a bookmark, etc.), which items of the semantic graph are affected as well as a timestamp. The lower half of the figure shows the different MB values: the initial/classic MB value (of Prototype I) as well as the new advanced MB values, whereas implementing Group MB was not part of the bachelor thesis. In each case, the left table contains the MB values before processing the current event (highlighted in green in the event log), while the right one shows the values after it. It is therefore possible to see how each individual event changes the MB values of different resources related to this event. Additionally, MB values can be viewed for each user of a certain context individually.

As a baseline for the evaluation the initial/classic MB algorithm implemented as part of CoMem was used. Using a pre-version of the ATIC KG Store (see Sec. 5.2), running both algorithms in the application developed by Chwalek [63] was possible. Although only a pre-version was used, this may still serve as an example demonstrating ATIC KG Store's compatibility to CoMem. The experimenters could thus verify that undesired effects originating from the context-insensitivity of the initial MB version were not present anymore with the new MB values.

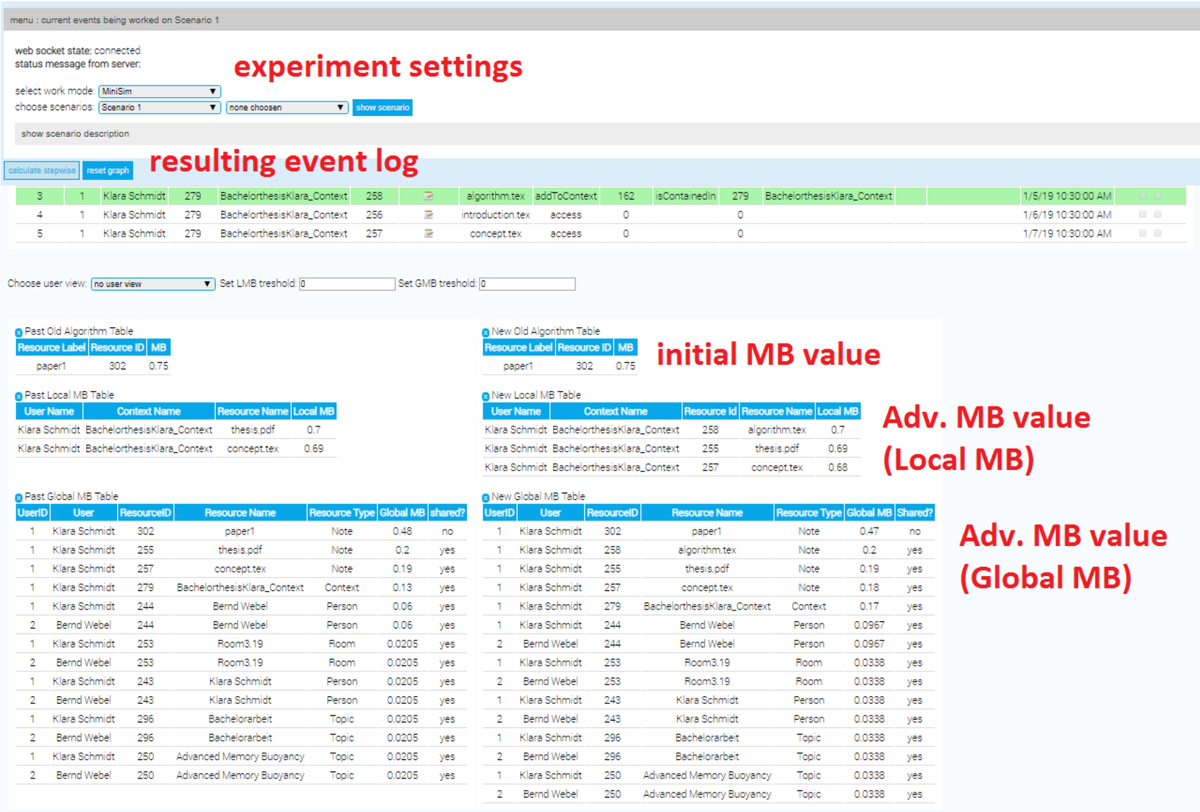


Figure 7.5: Memory Buoyancy Prototype II: Experimentation GUI Realized in Chwalek [63]. (Note: This screenshot has already been published in Jilek et al. [182].)

7.2.3 Ideas for Big Personal Data Processing and MB in Corporate Scenarios

This section names a few ideas potentially addressed and incorporated into a third MB prototype.

Ideas regarding **Big Personal Data processing**:

- **Optimizations with regard to Global MB calculation**: calculating *Global MB* based on *Local MB* values would reduce effort. Since activities are typically performed in a context, Local MB calculation (for the resource in that context) should have already been made, which can be used for further processing.

- **Explore potential for simplification of calculations**: Currently, MB updates are triggered by events, resources being queried, etc. "Classic" and Global MB both have the problem that MB drops by the passing of time, even if a user possibly has not been on their computer at all (no activities, e.g. due to being in hospital for weeks after an accident). For future work, the author therefore had the idea of incorporating a kind of "MB heartbeat", i.e. tracking how many actions have occurred in a certain time interval and additionally tracking which resources have been used in that interval. By time progressing, such activity vectors could later be condensed: one for the very recent past (a few days), one for a bit longer (possibly a few weeks), maybe a month, quarter and year, and finally a total sum. Currently, there is a static and a dynamic part of MB calculation. This could be enhanced to also take the "heartbeat" and related activity vectors into account possibly reducing the amount of active MB calculations in the background.

Ideas for **MB in corporate scenarios**:

- **Legal regulations**: More than in private scenarios, enforcing compliance with legal regulations and laws is important in corporate settings. Certain information items need to be deleted after certain periods. This could be incorporated into the MB calculation.
- **Item pinning**: There are items which should more or less "never" sink into the forgotten area, although they are not frequently used. MB Prototypes above applied this strategy based on item types, e.g. contracts. In corporate settings, enviaM [109], for example, has use cases, in which basic data about properties (of land) should never sink away, although possibly rarely used. Users could be enabled to also manually "pin" an item thus preventing – or at least greatly slowing down – the "sinking away" process. If this feature is overused, however, the system would, in the end, still have to find ways of deciding which pinned items to show or hide, e.g. in short item lists. (Item pinning is also a meaningful feature in non-corporate scenarios.)
- **Historic lanes**: The SDS research department typically updates its research road map every year. Thus, there is presentation file for each year, which is slightly different the last year's. A historic lane of such a presentation file has emerged. MB could thus give the highest score to the most recent version and treat the other ones quite different from other files that are not so strongly connected to a certain version, e.g. they may sink away faster since they are outdated and were replaced.
- **MB boostrapping**: Section 5.3 presented tools to bootstrap context spaces. One next step could be to also set initial MB values for these context spaces and their contained items. In corporate scenarios, such a MB bootstrapping would likely be highly influenced by processes in the company determining general importance and estimated frequency of use of certain items.

In the next section, Memory Buoyancy's counterpart, the *Preservation Value*, is presented. Instead of a short-/medium-term perspective, it tries to assess information value from a long-term perspective.

7.3 Preservation Value – Long-term IVA

Analogous to *Memory Buoyancy* trying to identify relevant items for the current situation (short-term) or the foreseeable future (medium-term), the *Preservation Value* has the goal of identifying information items worth preserving for long times like decades, possibly even longer. Note: Very-long-term preservation (e.g. hundreds of years) raises several additional problems like service agreements. Which company actually offers a storage space *guaranteed* to be available for hundred or more years? Since this is typically longer than a user's life: how is inheritance handled? etc. This topic has only been briefly addressed in the *ForgetIT* project [120] and is also out of scope of this thesis. However, trying to identify which items would be candidates to be preserved for longer, possibly put into such a very-long-term archive, is a topic of this thesis. From a knowledge management perspective, one could metaphorically speak of "finding the gold nuggets" in a user's information sphere.

Two Preservation Prototypes haven been envisioned and realized in this PhD project. The first one with a reduced scope of only dealing with users' photo collections in CoMem serves as a first evaluation of the Preservation Value and is introduced in the remainder of this chapter. Prototype II addresses general preservation in CoMem (i.e. not just photos but possibly all items) and is presented in Section 8.6.1 as an example of support measures by permanent reorganization.

7.3.1 Preservation Prototype I: Photo Management in CoMem

Colleagues of the ForgetIT project [120] conducted a survey on personal preservation of photos [401]. The following paragraph summarizes the results of this study [401] as well as further remarks of Maus et al. [245]. In the study, four personas representing users' attitudes towards personal preservation strategies could be identified. Those personas were defined along two key preservation dimensions they labeled "Loss" (worried about losing important photos) and "Generations" (importance of preserving important photos for future generations). For each dimension, two habits could be clustered: With regard to "Loss", one group (*Safety in redundancy*, L_1) tended to have redundant copies of their photos on different storage units in order to guard against the failure of one or more of such units. The strategies of a second group (*File and forget*, L_2) were using secure storage, automated backups and printing on paper. Regarding "Generations", one group (*Curators*, G_1) was more likely to curate their photos using both manual (e.g. assigning keywords) and automatic options (e.g. face recognition). A second group (*Filing first*, G_2) almost exclusively relied on files and folders. With these four dimensions, personas were identified as follows:

- **Safe Curator** (L_1/G_1),
- **Safe Filer** (L_1/G_2),
- **File & Forget Curator** (L_2/G_1) and
- (Filers that) **File & Forget** (L_2/G_2).

For each persona, a preservation strategy has been defined. To match the personas' preferences and assumed behavior, different weights were assigned to policies and rules of six different dimensions of Preservation Value assessment:

- **Investment**:
  - number of annotations (things having a higher number of annotations are more likely to be preserved)
  - wiki text length (the longer a thing's wiki text, the more likely it will be preserved)
  - usage (the more frequently a resource is modified, the more likely it will be preserved)
- **Gravity**:

  – connectivity (the more a thing is related to other things, the more likely it will be preserved)
  – type-based heuristic (certain things such as contracts are more likely to be preserved)
  – important projects (the higher the number of persons involved in a project, the more likely it will be preserved)
  – closeness to important things (things related to tasks or events are more likely to be preserved)

- **Social Graph**:
  – important persons (the higher the number of projects a person is involved in, the more likely it will be preserved)
  – PIMO (CoMem) user on photo (photos containing PIMO [CoMem] users are more likely to be preserved)

- **Popularity**:
  – image rating (the higher an image's rating, the more likely it will be preserved)
  – number of views (the more a thing is accessed/viewed, the more likely it will be preserved)

- **Coverage**:
  – cover photo collections (at least one photo of each photo collection should be preserved)

- **Quality**:
  – image quality (high quality images are more likely to be preserved)

With a GUI in CoMem, depicted in Figure 7.6, users may choose one of these personas, which applies a corresponding preset of values. In addition, they may further fine-tune the settings by selecting or de-selecting aspects. Which aspects are set by default for each persona is shown in Table 7.1. (Note that there is a preset called *Basic Curator* that has been additionally used in the evaluation – see next section.) The actual Preservation Value for each resource is then calculated as a weighted score of all six dimensions.

Technical Realization. The system performs IVA based on the aforementioned six preservation dimensions. It thus creates its suggestions completely based on metadata. During the development, the author first investigated how interaction profiles with photos in CoMem looked like for colleagues at DFKI SDS. These profiles turned out to look quite different matching various of the identified personas. Next, a **generator for photo metadata** was implemented taking aspects like number of clicks, length of comments, number of tags, frequency of photo revisitation, etc. into account. For each of the aspects, value ranges and simple distributions (like uniform, linearily increasing or decreasing) could be chosen. The generator performed well enough to prevent colleagues and the author from being able to distinguish between artificial usage profiles and real ones anymore. Then, the preservation system could be further

	File & Forget	File & Forget Curator	Basic Curator	Safe Curator	Safe Filer
Investment					
number of annotations	☑	☑	☑	☑	☑
wiki text length	☑	☑	☑	☑	☑
usage	☑	☑	☑	☑	☑
Gravity					
connectivity	☑	☑	☑	☑	☑
type-based heuristic	☑	☑	☑	☑	☑
important projects	☑	☑	☑	☑	☑
closeness to important things	☑	☑	☑	☑	☑
Social Graph					
important persons	☑	☑	☑	☑	☑
PIMO (CoMem) user on photo	☑				☑
Popularity					
image rating	☑	☑	☑	☑	☑
number of views	☑				☑
Coverage					
cover photo collections				☑	☑
Quality					
image quality	☑	☑			☑

Table 7.1: Preservation: Strategy Presets in CoMem.
The table shows the four personas identified in Wolters et al. [401] and their default values in the six dimensions of Preservation Value assessment. There is another preset called *Basic Curator*, which has been used in the evaluation (see Sec. 7.3.2).

tweaked by additionally generating a large variety of profiles, many of them not observed in CoMem, yet. For example, this could be persons commenting a lot but only short comments, persons almost not commenting anything but if they do, comments are very long, persons that have a preferred set of photos while not revisiting most others and so forth.

As mentioned before, the Preservation Value is a **weighted sum of all indicators from the six dimensions**. The sum is basically calculated by applying an approach previously used in similar scenarios by Schwarz [343] based on the **Dempster-Shafer Theory of Evidence** [(343) referring to (138, 403)]. According to this approach, two evidence scores v and w ($v, w \in [0,1]$) are added as follows:

$$v \oplus w := 1 - (1 - v) \cdot (1 - w)$$

For example, the $\oplus$-sum of 0.6 and 0.7 is 0.88. More detailed explanations can be found in Section 8.6.1 presenting *Preservation Prototype II*, an advanced version of Prototype I taking more types of information items into account. Both prototypes are purely rule-based, while the first one was used in the photo management study presented next and the second one for CoMem in general.

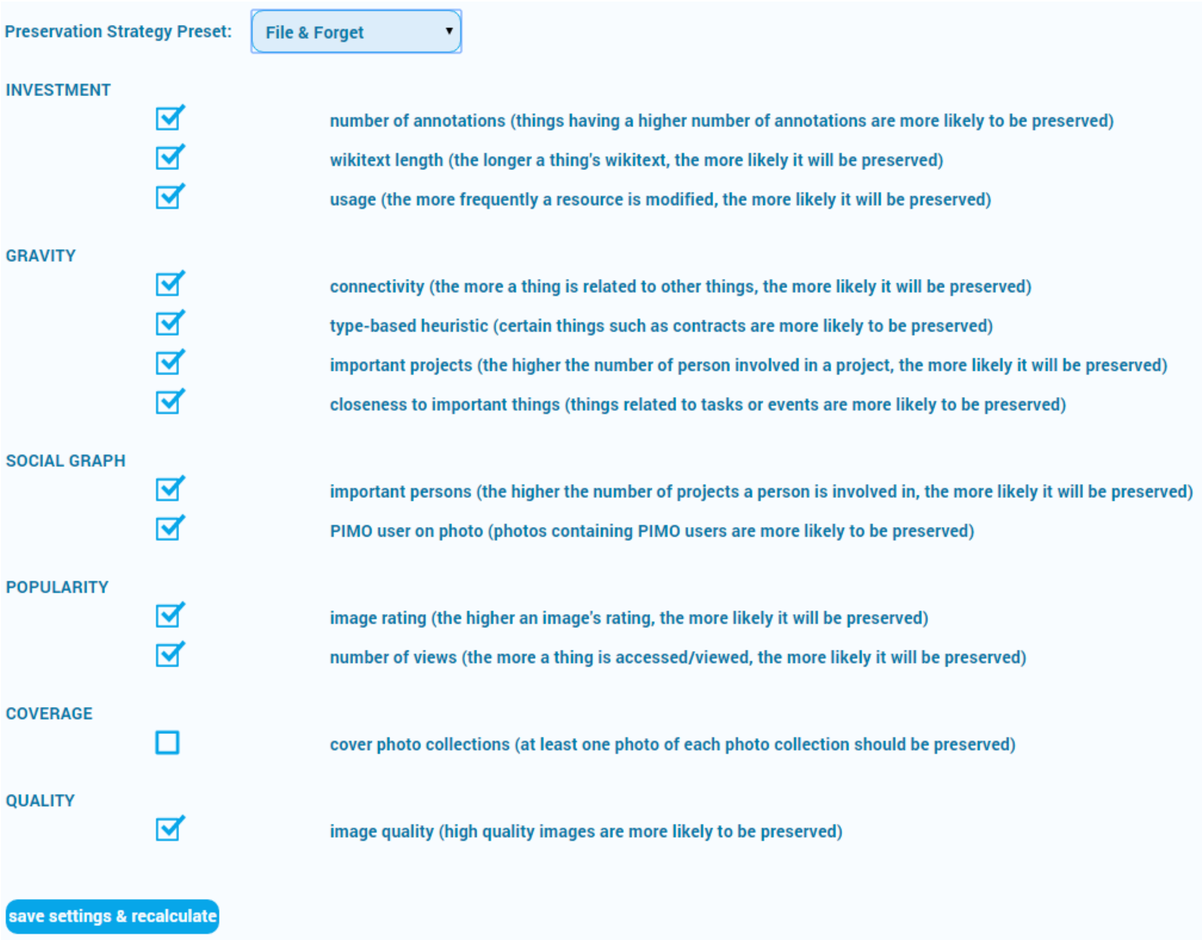


Figure 7.6: Preservation: Settings GUI in CoMem.
This screenshot's text is identical to the list of six dimensions given in Section 7.3.1.
(Note: Variants of this screenshot have been published multiple times, e.g. [240, 244, 245].)

7.3.2 Preservation Prototype I Evaluation

Evaluation Setting. The *Preservation Prototype I* was evaluated in a user study involving ten participants using CoMem to organize their photo collections (see CoMem's photo management features in Sec. 2.3.3). Since CoMem is available as a cloud service, they had access to the system 24/7 for several weeks. The system was manually bootstrapped by interviewing the participants, so their PIMO could already be pre-filled with topics, person names, locations, etc. they knew and were probably interested in using when managing their photos with CoMem. The initial interview was also used to determine the best fitting persona for the participant by asking about their attitudes and habits towards photo preservation.

Evaluation Procedure. After using CoMem for several weeks, participants were invited for a final interview. First, they were shown a special **Time Capsule Feedback GUI** (developed by Sven Schwarz), basically a vertical split-screen showing photos (sorted by collection) to be preserved on the left and ones not to be preserved on the right. With a single click, users could move photos from one side to the other as shown in Figure 7.7. Then, the final interview with each participant was conducted.

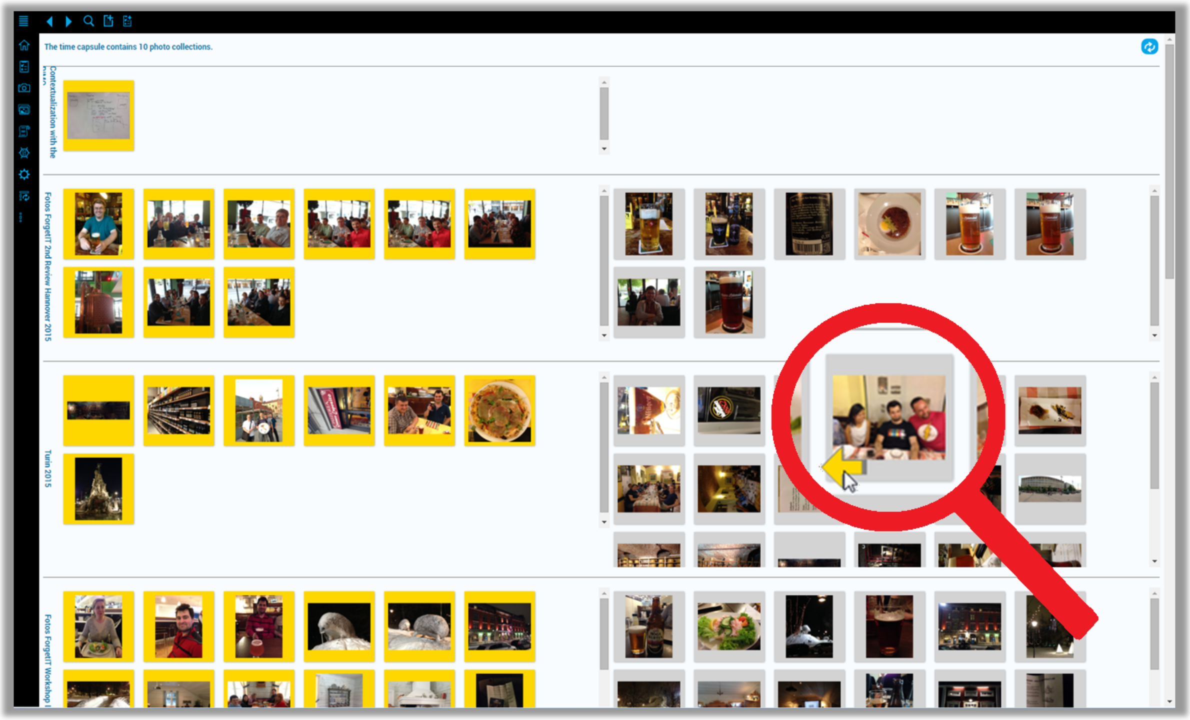


Figure 7.7: Preservation Prototype I: Time Capsule Feedback GUI in CoMem.
(Note: A variant of this screenshot has been published in Maus et al. [245].)

Part.	Preservation strategy	archived	arv←rem	arv→rem	agreed
P1	File & Forget	76%	4	2	76%
P2	Safe Curator	71%	1	1	93%
P3	Basic Curator	83%	4	0	94%
P4	Basic Curator	61%	0	2	93%
P5	File & Forget	73%	1	0	93%
P6	Safe Curator	83%	1	2	87%
P7	Basic Curator	76%	0	2	92%
P8	Safe Curator	55%	1	1	96%
P9	File & Forget Curator	56%	1	2	96%
P10	File & Forget	60%	0*	0*	100%*

Table 7.2: Preservation Prototype I: Results of User Study.
Abbreviations: *arv: archived / rem: removed*.
*) Participant P10 disagreed with a large portion of the system's suggestions and thus omitted giving feedback per image.

Results. Results can be found in Table 7.2 and Figure 7.8. The table shows for each participant the selected preservation strategy, the total amount of photos that were preserved, the number of photos that were moved by the participant from non-preserved to preserved (arv←rem) and vice versa (arv→rem) and the rate of agreement to the system's suggestions. Figure 7.8 additionally depicts the archival and agreement rates graphically. The system selected between 55% and 83% of users' photos to be preserved (mean: 72%, median: 73%). For participants P2 to P9, the agreement to the system's suggestions was between 87% and 96%. Participants P1 and P10 stated in post-experiment interviews to not be convinced of the system resulting in a lower agreement of P1 and no time capsule feedback at all for P10.

Further details about the study can be found in Maus et al. [246].

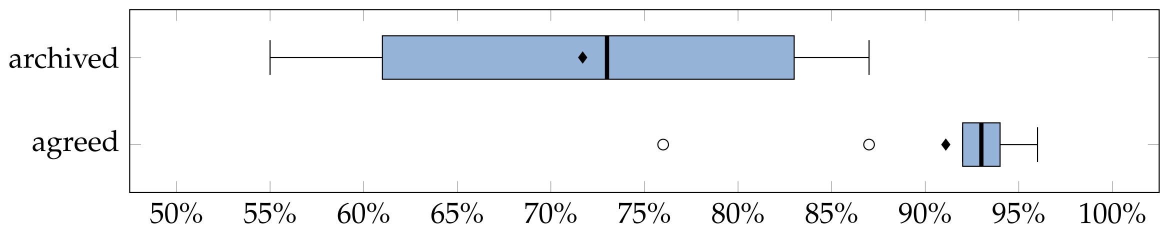


Figure 7.8: Preservation Prototype I: Results of User Study.
(n=9; one participant not considered: participant P10 disagreed with a large portion of the system's suggestions and thus omitted giving feedback per image; details on box plot semantics in Sec. B.2)

Conclusion. In summary, one can say that the presented Preservation Value assessment for photo collections gave helpful suggestions on what users should preserve for a longer time. Eight out of ten participants had agreement rates of 87% to 96% with an overall median and mean value of 93% and 91%, respectively (participant P10 not considered due to the lack of detailed feedback).

The *Preservation Prototype I*, which focused on photo management, has been enhanced to also work with all other types of information items available in CoMem. The resulting *Preservation Prototype II* as well as further insights about preservation value calculations are presented in Section 8.6.1.

This chapter has introduced the IVA approaches of Memory Buoyancy with a short- and medium-term perspective and the Preservation Value as its long-term counterpart in more detail. Both values are core components of support measures discussed in the next chapter.

8 Self-organizing Context Spaces (cSpaces)

Part II's last chapter is mainly about cSpaces' actual support measures, steps 5 and 6 of the interaction cycle, which are *self-(re)organization measures* and *UI updates* (including general aspects of *HFCI*). They are the last steps before the cycle repeats itself.

The chapter first discusses evaluation aspects (Sec. 8.1) like challenges in the research area of IM and KW support in general (8.1.1) and the chosen strategy, methods and tools in particular (8.1.2). Next, the prototypical implementation of cSpaces with its different variants and user interfaces is presented in Section 8.2.

The main part of the chapter covers all experiments and studies conducted with variants of cSpaces. Section 8.3 discusses working *with* and *in* (Self-organizing) Context Spaces in general and measured benefits in particular. The rest of this part is basically sorted by the different types of support measures: condensation and summarization (8.4), temporal reorganization, fading out and resurfacing (8.5) and permanent reorganization (8.6).

Although HFCI aspects are also part of other sections, the last part of the chapter particularly focuses on them: Searching (8.7) and trust (8.8) in such highly autonomous, forgetting-enabled information systems (FIS) like cSpaces are discussed and various prototypes are presented.

> Note: Parts of this chapter have already been published [126, 182–185, 187, 240] or are in preparation to be published [309, 383].

8.1 cSpaces Evaluation

This section gives an overview on evaluation challenges in the research area of IM and KW support and discusses the author's chosen strategy, methods and tools to cope with them.

8.1.1 Evaluation Challenges in the Research Area of IM and KW Support

Overview of Challenges. In the research area of IM and KW support, there a several challenges with regard to evaluation, which mostly originate from users having subjective views on their data [88, 28] and privacy issues (see Sec. 6.1.2). Both aspects strongly restrict evaluation scenarios and are – complemented by copyright issues of involved material – presumably the reasons why there are still no public PIM datasets available, even after decades of research – see next paragraph.

With cSpaces being an self-(re)organizing and especially forgetting-enabled system, there is the additional challenge that several support measures need some time to have passed. For example, the process of things slowly fading out of memory happens steadily over longer periods of time. This either rules out short-term laboratory studies (e.g. of 30 to 90 minutes) or forces bringing out effects artificially.

PIM Dataset Problem. For decades, there has not been a publicly available PIM dataset. To the author's best knowledge, Abela et al. (2015) [5] is the most recent paper mentioning the plan to release a dataset "in the near future", which is meanwhile already years ago. Gonçalves [136] even argues that if such a dataset was available, it would still lack the semantic information to really make use of the data, e.g. whether a term is the name of a project or whether a mentioned person is a co-worker or spouse, etc. Other approaches like Kim and Croft [204, 205] created *pseudo desktop collections* for their experiments on information retrieval. These collections neglect important sources like bookmarks or calendar events as well as structures like the file folder hierarchy, which also carry a lot of semantics. There is also the well-known *Enron Email Dataset* [Sw33] but its focus on emails only covers a fraction of PIM activities and data.

Since 2014, there is the semi-public **SWELL Knowledge Work Dataset for Stress and User Modeling Research** [Sw67] published by members of the SWELL project [371] (see Sec. 4.1.3). Some of its files are freely available, while others need a granted permission request (justified by the authors due to the dataset's size of about 7 GiB) [213]. The dataset was provided to members of DFKI SDS. It contains collected data (including *uLog* [Sw94] activity logs) of 25 participants performing typical knowledge work activities like writing reports, making presentations, reading emails and searching. Participants worked for about three hours in total on up to six different tasks (depending on how fast they were working). Three of them were opinion tasks (stress at work, healthy living and privacy on the internet), and three other ones required more information seeking (tourist attractions in Perth, a roadtrip to the USA and the life of Napoleon) [299]. Kraaij et al. note that "due to copyrights not all original documents and images are included in the dataset" [Sw67]. File histories that list accessed files and websites contain on average 426 entries per participant (min.: 87, max.: 611, total: 10,656). In these files, websites are given as a URL and a title. The same is true for local files, whereas for some of them no title is given. Each document written by participants is available in its final state, which amounts to eight files on average per participant (min.: 4, max.: 13, total: 201).

While the dataset is a welcome progress in the area of PIM research (and colleagues and the author are grateful to have been granted access to it), it still lacks too much to be used in cSpaces scenarios:

- While there is a lot of information about performed activities, cSpaces would be completely blind for the actual items mentioned in these logs and especially their content: the content of websites and local files is missing completely except for those files that have been written by the participants themselves, which is just a fraction of all items mentioned in the file histories (about 201 of 10,656). Intermediate states of these self-written documents are not given. It is thus not possible to synchronize the writing progress with logged events.
  A colleague of the author, Mahta Bakhshizadeh, tried to reconstruct the websites' content either by accessing the sites directly or by using Internet Archive's *Wayback Machine*[1].

[1] https://web.archive.org/

Many pages were not available anymore and for the ones available, one cannot be sure whether the obtained content was the one actually seen by participants. Still available pages may have changed in the meantime and archived copies of the Wayback Machine are typically available for a time before and after the experiment, leaving it unclear whether there were additional changes for which no crawled version is available.

- The aforementioned semantic information, whose general lack was brought up by Gonçalves [136], is also missing here. There is no KG available relating mentioned items. Since the study mainly was about topics for which a lot of public information is available, such KGs could be constructed post hoc but whether they really reflect the views and vocabulary of the participants, i.e. their PIMO, is doubtful.

- In addition, contexts are not seen as explicit interaction elements actively used by the participants, thus another essential source of information for cSpaces (like added tags, items, descriptions, etc.) is missing.

- The experiment's duration of about three hours is typically too short for forgetting effects and measures to happen or be triggered naturally, and artificially forcing them is not possible post hoc.

- Last, even if all this information was available, ground truth data regarding cSpaces support measures would be missing.

In summary, there is no dataset available that could be used for cSpaces scenarios. Thus, the author had to collect own datasets or find other solutions as described in the next section.

8.1.2 Evaluation Strategy, Methods and Tools

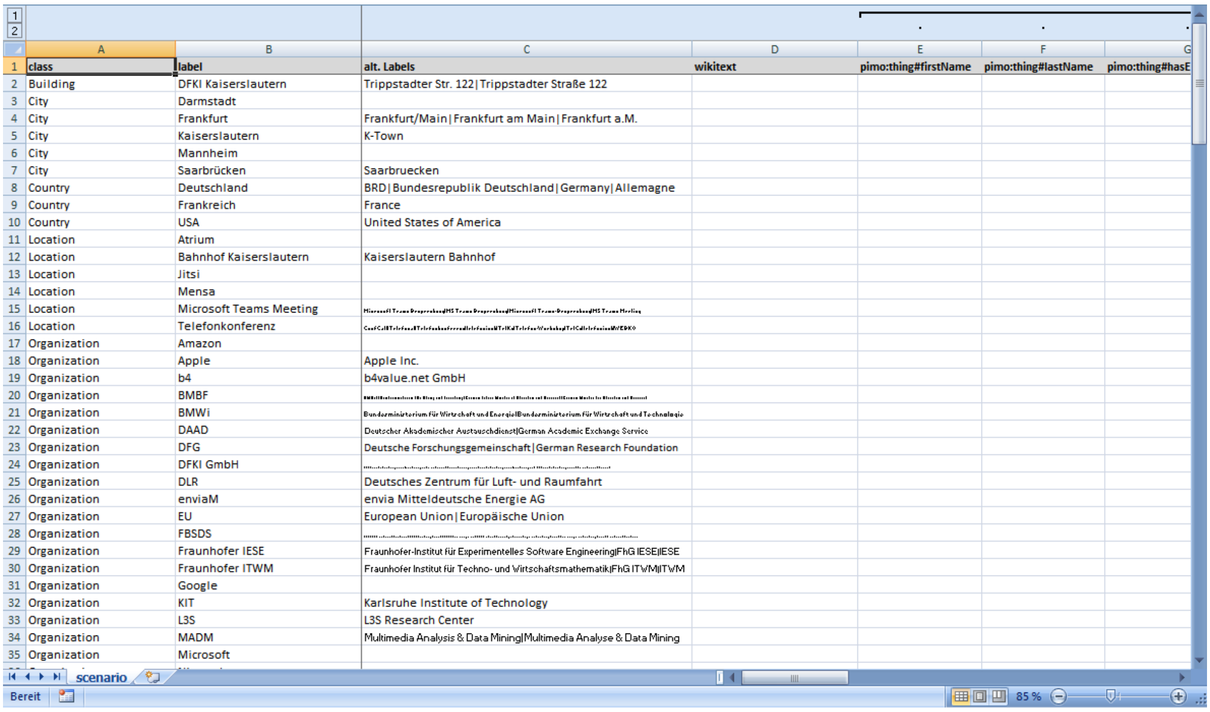
The previous section named four major evaluation challenges in cSpaces scenarios. They are repeated in the following together with ideas on how to tackle them:

- **Subjectiveness**: How could participants judge whether cSpaces support measures, especially reorganization and forgetting, are correct and helpful if they never knew the involved information items? As a consequence, participants should mostly work on their own data. Bootstrapping measures as presented in Section 5.3 can be used to pre-fill a participant's PIMO before starting the evaluation. These measures are, for example, complemented by tools to manual bootstrap certain scenarios using a simple spreadsheet editor as shown in Figure 8.1. As in the case of the evaluation of the *Preservation Prototype I* (see Sec. 7.3.2), the experimenter (or the participants) may enter relevant persons, locations, topics, etc. (see the figure's description for detailed information), so cSpaces already has some background information to work with and does not suffer from a cold start. Experiments then reach interesting situations which are the actual target of the investigations faster.

- **Privacy issues**: Solutions on how to tackle privacy issues have been presented in Section 6.1.2. The main measures were informing participants as explained in detail by Koldijk et al. [211]. Regarding longer studies, participants should be allowed to disable observation temporarily, and sensitive data should only be stored on their local machine without any server/cloud uploads. Data may be manually send to the experimenter after anonymization.

- **Several support measures needing certain amounts of time to have passed**: Experiments of only 30 to 90 minutes are too short to get participants or the system in a situation, in which several of the support measures would actually be applied. Although some aspects can be simulated (e.g. seeing unknown data for the first time is similar to seeing your own data again after a very long time), testing a long-term measure by testing its modified variant for short-term experiments remains at least difficult due to possibly falsifying assumptions and modifications to be made. As a consequence, the author pursues a two-fold strategy of short-term and medium/long-term experiments dependent on each feature.

- **Missing PIM datasets**: Own datasets are collected instead.



	A	B	C	D	E	F	G
1	class	label	alt. Labels	wikitext	pimo:thing#firstName	pimo:thing#lastName	pimo:thing#hasE
2	Building	DFKI Kaiserslautern	Trippstadter Str. 122|Trippstadter Straße 122				
3	City	Darmstadt					
4	City	Frankfurt	Frankfurt/Main|Frankfurt am Main|Frankfurt a.M.				
5	City	Kaiserslautern	K-Town				
6	City	Mannheim					
7	City	Saarbrücken	Saarbruecken				
8	Country	Deutschland	BRD|Bundesrepublik Deutschland|Germany|Allemagne				
9	Country	Frankreich	France				
10	Country	USA	United States of America				
11	Location	Atrium					
12	Location	Bahnhof Kaiserslautern	Kaiserslautern Bahnhof				
13	Location	Jitsi					
14	Location	Mensa					
15	Location	Microsoft Teams Meeting	[illegible]				
16	Location	Telefonkonferenz	[illegible]				
17	Organization	Amazon					
18	Organization	Apple	Apple Inc.				
19	Organization	b4	b4value.net GmbH				
20	Organization	BMBF	[illegible]				
21	Organization	BMWi	Bundesministerium für Wirtschaft und Energie|Bundesministerium für Wirtschaft und Technologie				
22	Organization	DAAD	Deutscher Akademischer Austauschdienst|German Academic Exchange Service				
23	Organization	DFG	Deutsche Forschungsgemeinschaft|German Research Foundation				
24	Organization	DFKI GmbH	[illegible]				
25	Organization	DLR	Deutsches Zentrum für Luft- und Raumfahrt				
26	Organization	enviaM	envia Mitteldeutsche Energie AG				
27	Organization	EU	European Union|Europäische Union				
28	Organization	FBSDS	[illegible]				
29	Organization	Fraunhofer IESE	Fraunhofer-Institut für Experimentelles Software Engineering|FhG IESE|IESE				
30	Organization	Fraunhofer ITWM	Fraunhofer Institut für Techno- und Wirtschaftsmathematik|FhG ITWM|ITWM				
31	Organization	Google					
32	Organization	KIT	Karlsruhe Institute of Technology				
33	Organization	L3S	L3S Research Center				
34	Organization	MADM	Multimedia Analysis & Data Mining|Multimedia Analyse & Data Mining				
35	Organization	Microsoft					

Figure 8.1: Experiment Tooling: Manual Scenario Bootstrapping with Spreadsheets. Columns: *class, [preferred] label, alternative labels [separated by "|"], further object and literal properties like wiki text, first-name, last-name, has-Location, is-Member-of, etc.* The preferred label is assumed to be unique. If users would like to state a relation to another thing, they may thus refer to each thing by their preferred label (e.g. *PersonZ –isMemberOf→ CompanyXY*). The system later generates an actual URI for each thing. By default, the PIMO Thing Ontology is assumed (see Sec. A.1). For classes of other ontologies, stating the full URI is required.

Multi-lane Evaluation Strategy. In summary, cSpaces' evaluation comprises several lanes:

- First, **short-term studies** are conducted, preferably under controlled laboratory conditions and supported by colleagues of cognitive psychology. Such studies typically involve a comparably large number of participants but the exposure to the system is quite short, e.g. 30 to 90 minutes. As a consequence, some aspects need to be "artificially" simulated or enforced to happen during the experiment. Gathered data may be quantitative or qualitative. Although conducted in a less controlled setting, the previously mentioned *Contextifier* (see Sec. 5.3.3) is an example of this lane. Participants used the app on their own data for a few hours and gave qualitative feedback. Another example is the *Stream Context Miner* (see Sec. 6.3), which participants used for a few days. After submitting their collected data, quantitative results could be obtained.

- A second lane covers less controlled experiments but therefore span **larger periods of time like weeks or months (sometimes even years)**. Mainly for cost reasons, less participants are typically involved but effects to be observed occur naturally due to the longer times of exposure to the system. The already mentioned cSpaces multi-month user study has been conducted with colleagues of DFKI SDS. Another example is the evaluation of the *Preservation Prototype I* that lasted several weeks (see Sec. 7.3.2).

- Another option is to provide research and demo material to audiences, let them perform a thought experiment like assuming a certain role in a hypothetical situation and perform an **inquiry** on a given topic.

- One could also follow a **purely data-driven approach by using existing datasets** as in the case of IT-RTC-OBNER (see Sec. 6.2.3). However, given the reasons stated in the previous section, this an unlikely option.

- A topic for future work is using **simulation and additional dataset generators** for PIM-related evaluations. An example of such a generator was given in the context of *Preservation Prototype I* (see Sec. 7.3.1). Instead of generating the actual information items, metadata is created, e.g. things mentioned in each information item, how things relate to each other (i.e. connections in the KG), etc. Together with Markus Schröder and Michael Schulze, other generators unrelated to cSpaces have already been created [327, 335, 340].

Experimentation Framework and Tools. The collaboration with cognitive psychologists and ergonomists has already been mentioned multiple times so far. To ease setting up and fine-tuning experiments, the author has created an experimentation framework. Basically, it adds experiment guiding features to existing cSpaces prototypes. The framework typically involves a web browser-based dashboard to display instruction texts and collect feedback. However, other widgets are possible as well. For example, Figure 8.16 shows the cSpaces sidebar displaying an extra tutorial section tailored for the experiment (green widget, top right, highlighting B). Every phase of an experiment is called a *scene*. Examples are an initial personal data questionnaire, an instruction text, a form or survey to enter results of a task, post-experiment questionnaires, etc. – see Figure 8.2 as an example of such a survey. After performing a certain task or the expiration of a timer, the system transitions to the next scene.

Using this framework, colleagues and the author could easily fine-tune or re-adjust (reduce or extend) experiments without recompiling or redeploying code. A typical scenario demanding for a bit of fine-tuning is misunderstood instruction text in the pre-tests of an experiment. Such rephrasing and other customizations could also be performed by non-technical team members since the configuration was realized using well-known spreadsheet editors like *Microsoft Excel* [Sw79] or rather intuitive languages like *YAML* [25, 404]. Regarding the latter, Listing 8.1 shows the example configuration of a survey. The resulting GUI is depicted in Figure 8.2.

```
---
id: 300
instructions: 'Im Folgenden finden Sie eine Reihe von Aussagen.
  Bitte geben Sie für jede der Aussagen an, wie gut sie auf Sie persönlich zutrifft. Vielen Dank!'
numValues: 5
items:
- id: 1
  question: "Ich liebe es, neue elektronische Geräte zu besitzen."
  minName: "trifft voll zu"
  maxName: "trifft gar nicht zu"
  allNames: ["trifft voll zu", "trifft eher zu", "teils/teils", "trifft eher nicht zu", "trifft gar nicht zu"]
- id: 2
  question: "Elektronische Geräte machen krank."
  minName: "trifft voll zu"
  maxName: "trifft gar nicht zu"
  allNames: ["trifft voll zu", "trifft eher zu", "teils/teils", "trifft eher nicht zu", "trifft gar nicht zu"]
- id: 3
  question: "Ich gehe gern in den Fachhandel für elektronische Geräte."
  minName: "trifft voll zu"
  maxName: "trifft gar nicht zu"
  allNames: ["trifft voll zu", "trifft eher zu", "teils/teils", "trifft eher nicht zu", "trifft gar nicht zu"]
```

Listing 8.1: Experiment Tooling: Survey Configuration with YAML [25, 404].
(The resulting survey is depicted in Figure 8.2.)

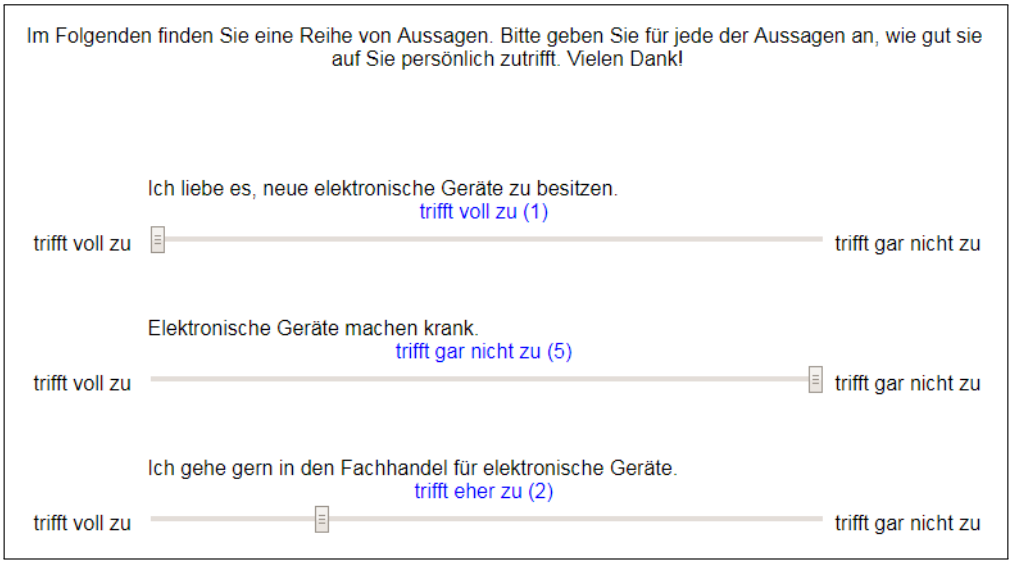
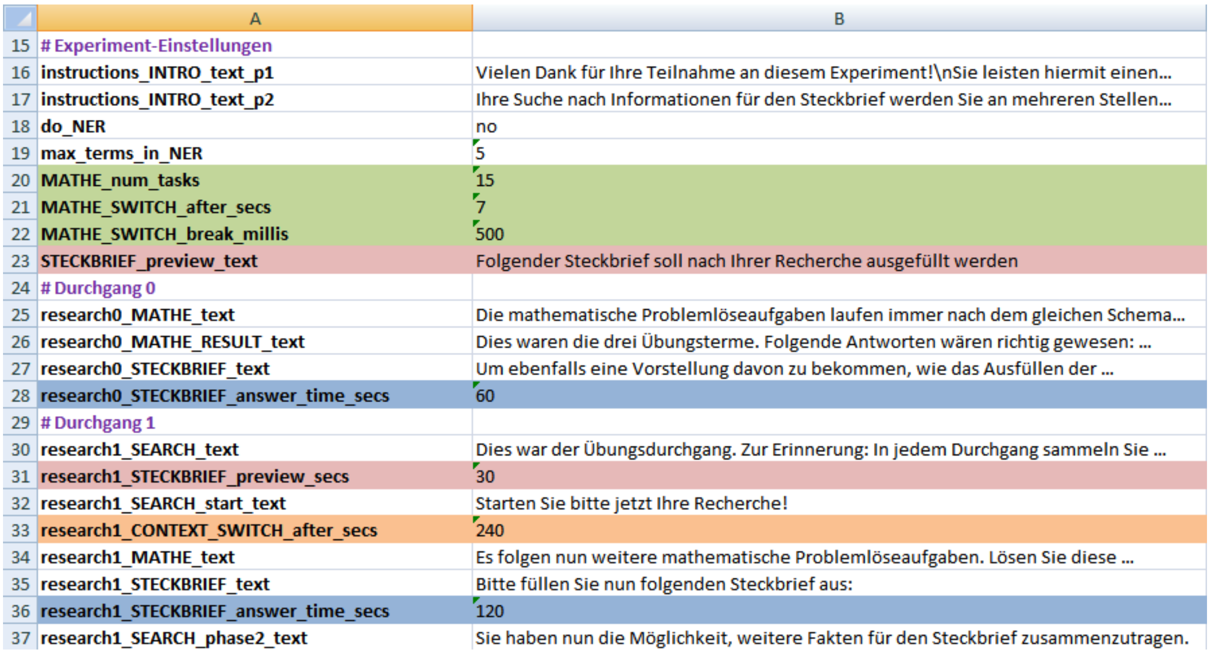


Figure 8.2: Experiment Tooling: Survey GUI.
(The depicted survey's configuration is shown in Listing 8.1.)

If the used Likert scale [226] should, for example, be changed from five points to seven points, one would just increase the number in line 5 and add two additional item texts in lines 11, 16 and 21 (copy-and-pasting lines is advised). Other configurations like the aforementioned rephrasing of instruction texts, enabling or disabling functionality (e.g. perform NER and show results), regulating amounts like the number of items asked in a test or distraction task or setting processing and transition times can be done using spreadsheets as shown in Figure 8.3.



	A	B
15	# Experiment-Einstellungen	
16	instructions_INTRO_text_p1	Vielen Dank für Ihre Teilnahme an diesem Experiment!\nSie leisten hiermit einen...
17	instructions_INTRO_text_p2	Ihre Suche nach Informationen für den Steckbrief werden Sie an mehreren Stellen...
18	do_NER	no
19	max_terms_in_NER	5
20	MATHE_num_tasks	15
21	MATHE_SWITCH_after_secs	7
22	MATHE_SWITCH_break_millis	500
23	STECKBRIEF_preview_text	Folgender Steckbrief soll nach Ihrer Recherche ausgefüllt werden
24	# Durchgang 0	
25	research0_MATHE_text	Die mathematische Problemlöseaufgaben laufen immer nach dem gleichen Schema...
26	research0_MATHE_RESULT_text	Dies waren die drei Übungsterme. Folgende Antworten wären richtig gewesen: ...
27	research0_STECKBRIEF_text	Um ebenfalls eine Vorstellung davon zu bekommen, wie das Ausfüllen der ...
28	research0_STECKBRIEF_answer_time_secs	60
29	# Durchgang 1	
30	research1_SEARCH_text	Dies war der Übungsdurchgang. Zur Erinnerung: In jedem Durchgang sammeln Sie ...
31	research1_STECKBRIEF_preview_secs	30
32	research1_SEARCH_start_text	Starten Sie bitte jetzt Ihre Recherche!
33	research1_CONTEXT_SWITCH_after_secs	240
34	research1_MATHE_text	Es folgen nun weitere mathematische Problemlöseaufgaben. Lösen Sie diese ...
35	research1_STECKBRIEF_text	Bitte füllen Sie nun folgenden Steckbrief aus:
36	research1_STECKBRIEF_answer_time_secs	120
37	research1_SEARCH_phase2_text	Sie haben nun die Möglichkeit, weitere Fakten für den Steckbrief zusammenzutragen.

Figure 8.3: Experiment Tooling: Experiment Configuration with Spreadsheets.
Depicted configuration types: instruction texts (lines 16-17, 23, 25-27, 30, 32, 34-35 and 37) / enable or disable functionality (line 18) / regulate amounts (lines 19-20) / processing times (lines 21, 28, 33 and 36) / transition times (lines 22 and 31). Lines starting with "#" are comments ignored by the system.

The framework was successfully used in four studies (related publications: [126, 183, 293, 337]). All these studies as well as further ones discussed in the remainder of this chapter used slightly different variants of cSpaces. More details about these proof-of-concept implementations are given in the next section.

8.2 cSpaces Proof-of-concept Implementations

cSpaces is – unfortunately – not yet a comprehensive application comprising all findings and components developed in this PhD project. Several aspects were investigated in parallel, e.g. as part of student projects or theses or as smaller explorative experiments by the author. Some were later integrated into the next cSpaces prototype, some were also integrated in CoMem and are thus in productive use (including industry). For others, integration is still a future task. Thus, all screenshots of cSpaces may look slightly different but core aspects like context selection, tags, etc. are typically recognizable. This section gives more details about different proof-of-concept implementations of cSpaces including further conceptual as well as technical aspects. It especially addresses research question RQ4.

8.2.1 Technical Summary of cSpaces

Like CoMem, cSpaces' back-end is written in *Java* [Sw95], while its front-end uses web technologies like *HTML5* [275].

Its GUI architecture follows the *MVVM pattern* [139] as depicted in Figure 8.4. As in the case of CoMem, front-end applications typically call back-end methods by means of *JSON-RPC* [194, 195] in a REST-like way [118] (drawn as blue arrows in the figure). In addition, *WebSockets* [116] are used to inform the front-end about events that happened in the back-end (drawn as green arrows). Although the whole communication could have been realized using WebSockets, they are only used unidirectional (back-end to front-end). This is a deliberate design decision to stay mostly compatible to CoMem. As defined in the MVVM pattern, there is the core *model* in the backend (left-hand side of the figure), typically used for back-end (business) logic and data. The front-end has an own model, the *view model*, containing UI-related parts of the model (possibly converted to suit presentation needs) as well as the presentation logic. Last, there is a "simple" data binding (purple color) between the view model and the actual *view*. Back- and front-end development can thus be easily separated. For example, the cSpaces sidebar and dashboard access the same model but basically have different view models and views (some components are shared though). There is an additional modularization step: Beside the core model, there are smaller models for sub-components like the *Annotator* (responsible for tagging), the *Notifier* (informs about updates or events, etc.), the *Browsed Item Info* (basic information about the currently browsed item, found named entities, ...), etc. Thus, if, for example, students would like to contribute functionality to cSpaces, they could develop one "row" in this figure from left to right: their component would need its own model, view model and view, whereas the overall model, view model and view can be accessed by all sub-components. They are therefore drawn as part of a roof over all sub-components in the figure (also see black vertical arrows which represent intra-component communication).

Regarding the back-end, cSpaces uses the introduced *ATIC KG Store* (see Sec. 5.2) as its core component for storage, analytics, etc. Its basic PIM ontology is an export [Sw110] from CoMem using *PimoScript* [Sw109]. Remember the one from CoMem emerged from the *PIMO Ontology* [305–307]. Further technical hints and details have already been given in previous chapters:

- overall idea of working *with* and *in* context spaces with a few technical hints (Sec. 3.2),
- self-(re)organization measures (Sec. 3.3.1),
- nucleus and flavors, overall technical scenario, user interface types and comparison with a traditional Semantic Desktop (Sec. 3.3.3),
- context model (Sec. 5.1),
- *ATIC KG Store and Search* and technical details of CoMem mostly also applicable to cSpaces (Sec. 5.2),

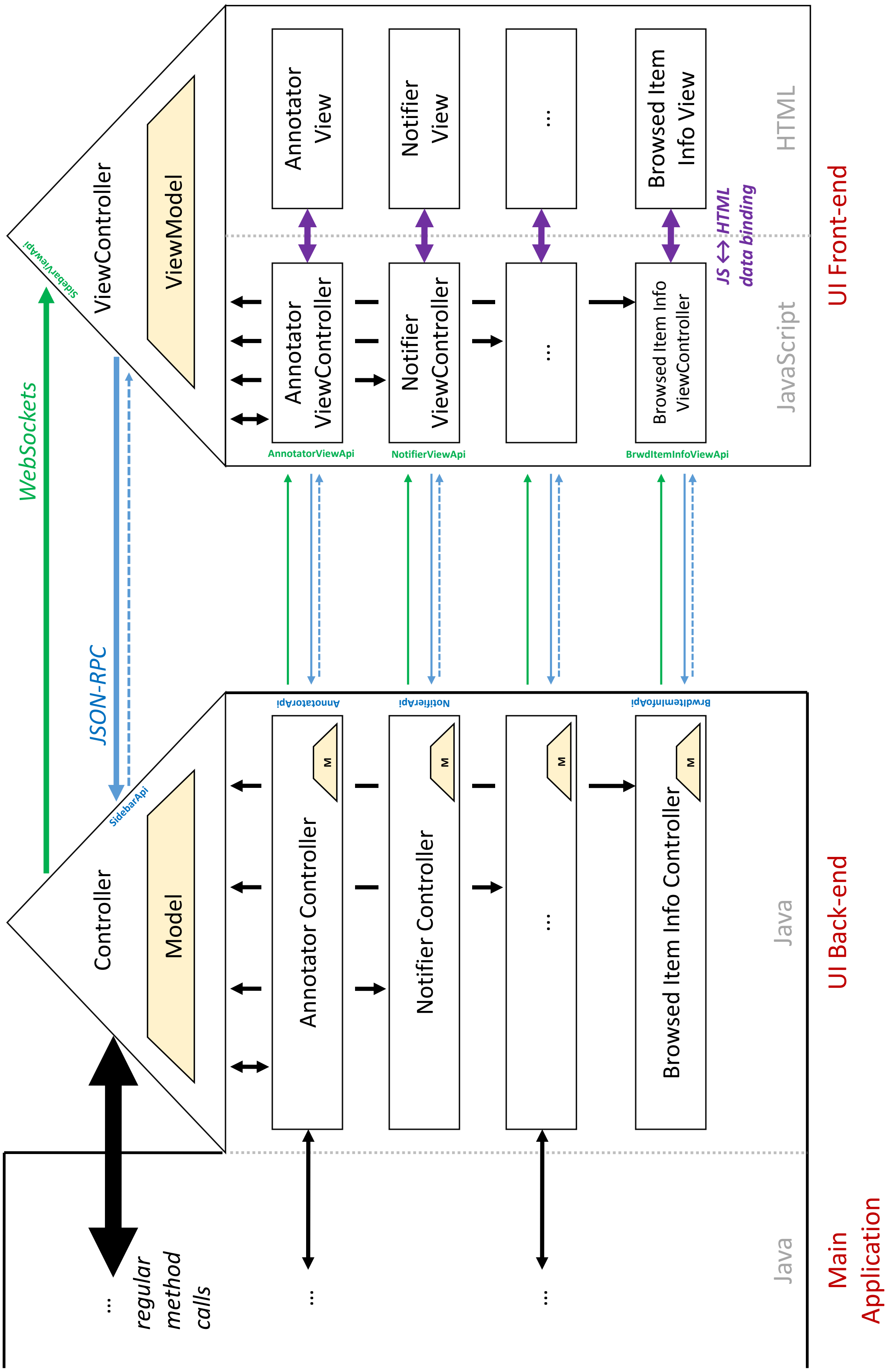


Figure 8.4: cSpaces GUI Architecture.
The cSpaces GUI is basically structured according to the *Model–View–ViewModel (MVVM) pattern* [139].

- *PIM Crawler*, *Semantifier* and *Contextifier* as bootstrapping tools (Sec. 5.3),
- *Plug-outs* and *AEBI* for user activity tracking as well as remarks on privacy preservation measures (Sec. 6.1),
- *IT-RTC-OBNER* as cSpaces' information extraction approach (Sec. 6.2),
- context elicitation using the *Stream Context Miner* (Sec. 6.3),
- *Memory Buoyancy* (Sec. 7.2) and *Preservation Value* (Sec. 7.3) as two variants of information value assessment.

8.2.2 cSpaces User Interfaces

Recapitulation of cSpaces User Interfaces. As mentioned in Section 3.3.3, cSpaces can have different user interfaces like a sidebar, a dashboard (and the idea of a "sashboard" mechanism to allow a dynamic transition between the two) and transparent injections into existing systems. As also said, these interfaces come with different levels of familiarity and are trade-offs between displaying/offering functionality and occupying/obscuring space on screen. For a more detailed comparison, please see the mentioned Section 3.3.3. In the following, the more general remarks of that section are complemented by details regarding their implementation, especially in the case of the transparent injections as the most complex UI of the three.

Context Dashboards. The first sketch drawn in this PhD project was the one of a dashboard (see Fig. 8.5). It was meant as a vision or mock-up guiding further ideas and development.

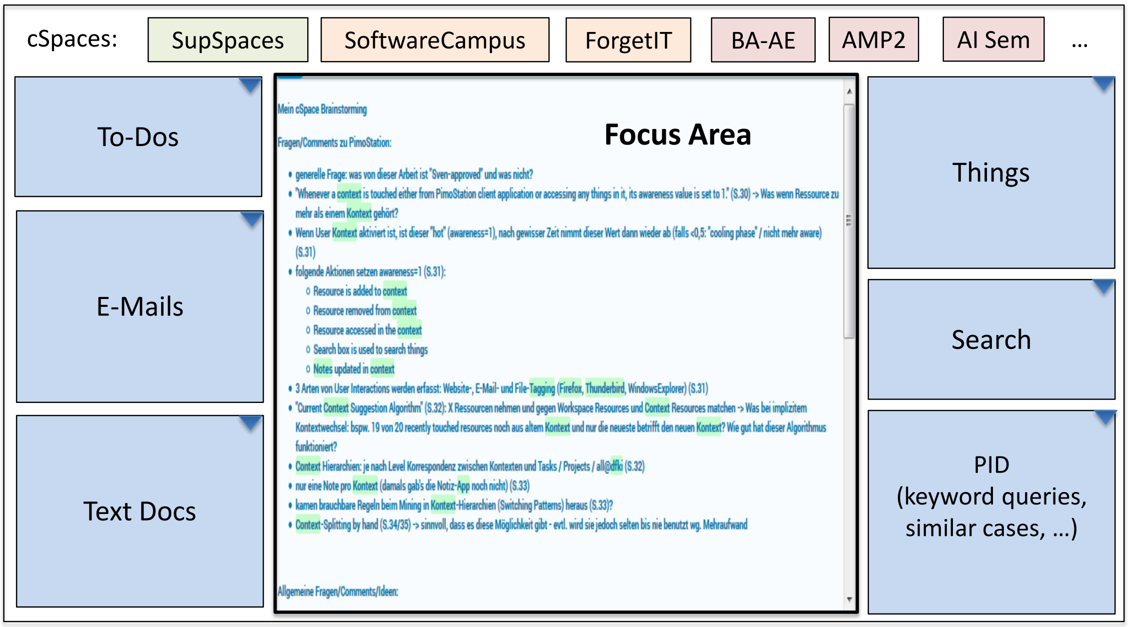


Figure 8.5: Guiding Context Dashboard Mock-up for this PhD Project.
It consists of three main parts: a focus area in its center, widgets for documents, to-dos, etc. on both sides and a context selection at the top.

Its core idea was having a *focus area* and additional widgets containing further contextual information around it. At the top, other contexts could be selected. According to these guiding ideas, a dashboard for the property management department of *enviaM* [109] was created. Figure 9.1 shows a first prototypical implementation by Sven Schwarz having the property (i.e. a piece of land) in its focus area and additional information like related files or mails, information of the land registry office, etc. in widgets around it. A second example of such a context dashboard was developed as part of the *supSpaces* project [370]. It is depicted in Figure 9.2. Here, the focus area is not a tile in the center but rather the whole middle column of the GUI. The scenario is about IT support tickets coming in (A), an information extraction task is performed leading to named entities being displayed, a clerk can confirm them or add additional tags (C) and take notes (B). According to this analysis older tickets (cases) are searched for a solution. Either such solutions or at least relevant resources are presented (D). The content of each shown entity can directly be browsed on the right (E). Later versions also allowed for keyword based search using an early version of CoMem/ATIC search (see Sec. 5.2.3).

These dashboards were the only ones related to this PhD project. They were both developed as part of CoMem. For the given business scenarios, CoMem was a better suited choice and a dedicated cSpaces app did not exist at the time anyway. Both dashboards are discussed in a bit more detail in Sections 9.1 and 9.2.1, respectively.

One outcome of the supSpaces project was that a sidebar or a "sashboard" mechanism (later called this way by the author) may have been helpful – presumably much more helpful than a dashboard alone. This was not the only reason why the author started to mainly focus on sidebar UIs after the supSpaces project. Another one was that the sidebar better resembles the idea of a helping companion or "information butler" [85, 86] *next to* the user – literally next to the windows the user is currently working in. And with the sashboard idea, there was still the possibility to enlarge sidebars to dashboard if necessary.

Context Sidebar. Analogous to the previously discussed dashboards, there is also a vision or guiding mock-up for cSpaces sidebars. It is depicted in Figure 8.6. Typically its top-most section (see highlighting A) is the context selection as well as basic information about the context like tags or contained items. Additionally, users may take notes for each contained item. Next, there is a short history of recent activities performed in the current context (B). Typically, each entry shows the title of the involved information item, e.g. a browsed website, a note taken by the user, etc. Clicking the green plus would add the item to the current context. The last browsed item is shown in another section (C) to allow for easy tagging. Results of NER are shown as *suggested topics*, which the user may confirm to make them *annotated topcis* (or *tags*) of the browsed item. Tagging automatically adds the item to the current context and also triggers its "rebirth" in the user's PIMO. Further widgets are, for example, PID for the current context or a selected thing (D) or special widgets like virtual file system controls (E). Apart from the first widget (context selection and basic information about the selected one), the order of widgets is not fixed.

Each cSpaces-related experiment (discussed in the remainder of this chapter) had its own combination of these widgets. Figure 6.2 showed an early sidebar prototype supporting web and email browsing. This sidebar version was tailored to particularly work with CoMem and is nowadays known as the **PIMO (CoMem) Desktop Application**. Rudolf Koch has meanwhile taken over further development and maintenance, in particular with regard to industry applications.

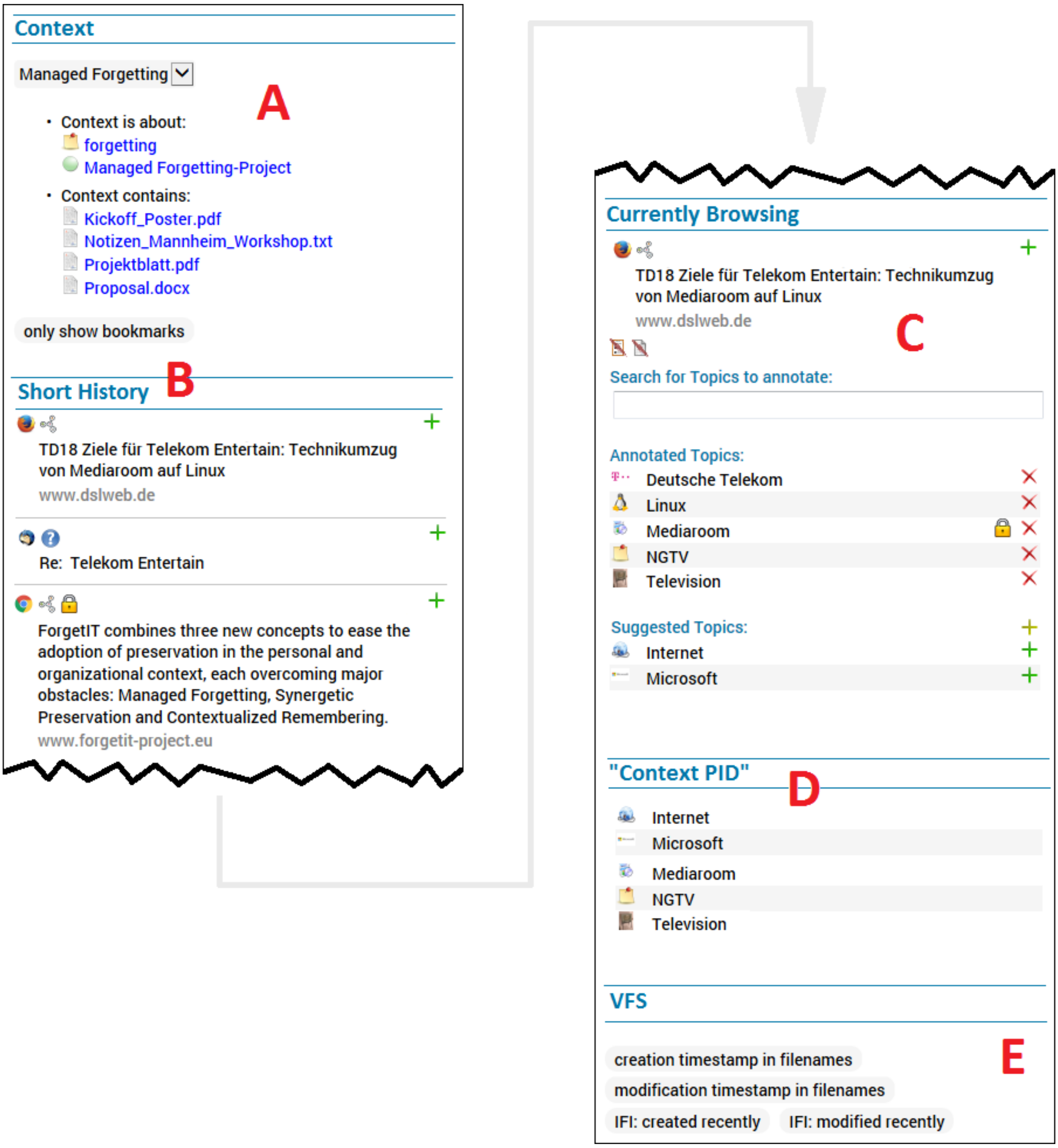


Figure 8.6: Guiding Context Sidebar Mock-up for this PhD Project.
Highlighted sections: *A: context selection and the current context's tags and content / B: history of activities in the selected context / C: details about currently browsed item and possibilities to tag the item / D: context-sensitive PID / E: further widgets, e.g. for quickly adjusting views like the Virtual File System (VFS).*
(Note: Mock-up split for the sake of better readability.)

Context Injections – A Near-transparent Semantic Desktop. One major goal of this thesis is to enable users to actually work *with* and *in* their different context spaces. Achieving this goal is mainly a technical challenge: Writing plug-ins or modifying all possible applications (including those yet to come in the future) to respect users' contexts is only a theoretical option. The same is true for rewriting core parts of operating systems to inject contexts (very expensive or proprietary software). However, a realistic option is utilizing standard protocols and the like. They are supported by all major operating systems today, thus allowing cSpaces to be integrated on a very low level without rewriting any operating system code. As shown in Figure 8.7, different protocols and standards are used for the integration:

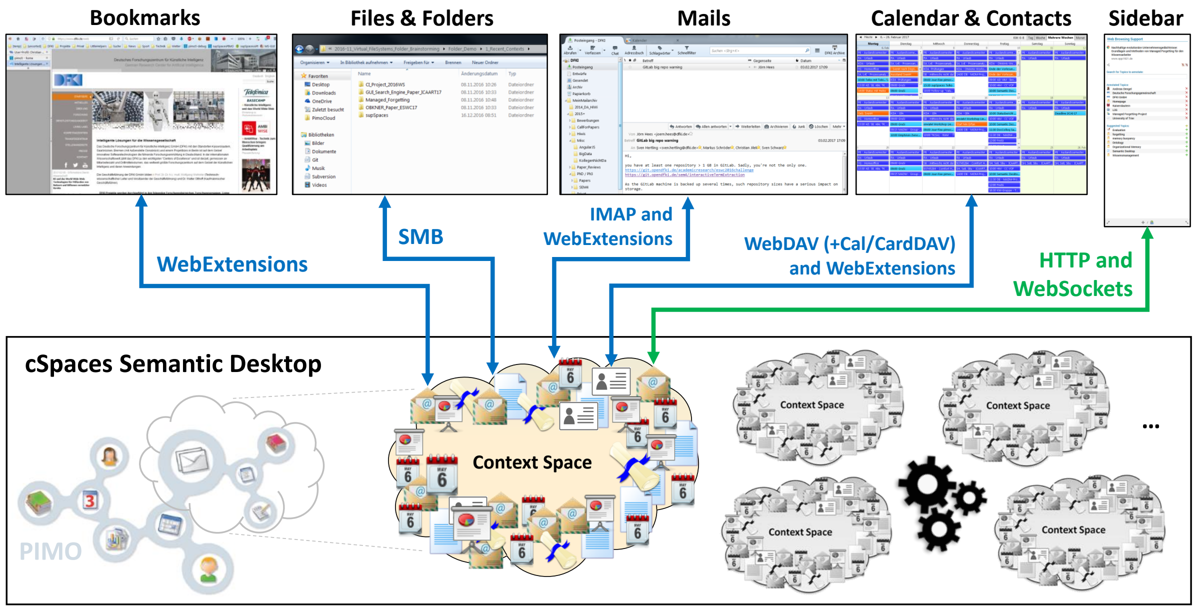


Figure 8.7: Transparent Injection: Realized Using Standard Protocols and the like.
(Note: An earlier version of this figure has been published in Jilek et al. [187].)

- web browser integration using *WebExtensions* [261],
- file system integration using the *Server Message Block (SMB) protocol* [161, 254],
- mail client integration using the *Internet Message Access Protocol (IMAP)* [71, 248] and *WebExtensions* [261]
- calendar and address book integration using *WebDAV* [134, 99] and its extensions *CalDAV* [75] and *CardDAV* [74] as well as *WebExtensions* [261],

Although WebExtensions are not a standard protocol, they are widely available in today's web browsers providing cross-browser functionality and an integration level similar to having an underlying protocol. Utilizing only standard protocols (and the like) has certain limitations due to their rather basic, low-level character. Some activities like tagging or writing a note or comment about a resource can become inconvenient or non-intuitive. To avoid this, the sidebar (or a similar widget) is used as a complement and single interaction point for using advanced features. As mentioned earlier in this section, it is connected via the *HyperText Transfer Protocol (HTTP)* [33, 117] and *WebSockets* [116]. With these *context injections*, applications operate on information items, which are actually handled[2] by the ATIC KG and transparently provided by cSpaces. An example is illustrated in Figure 8.8: A user clicks on *MyContexts* (topmost window). Another window opens up showing their different contexts: in this small example, there are three of them, which are also highlighted in the (symbolic) KG on the left. The user clicks on the *Managed Forgetting* context (middle window) and the context's content comes up in another file browser window (bottom right).

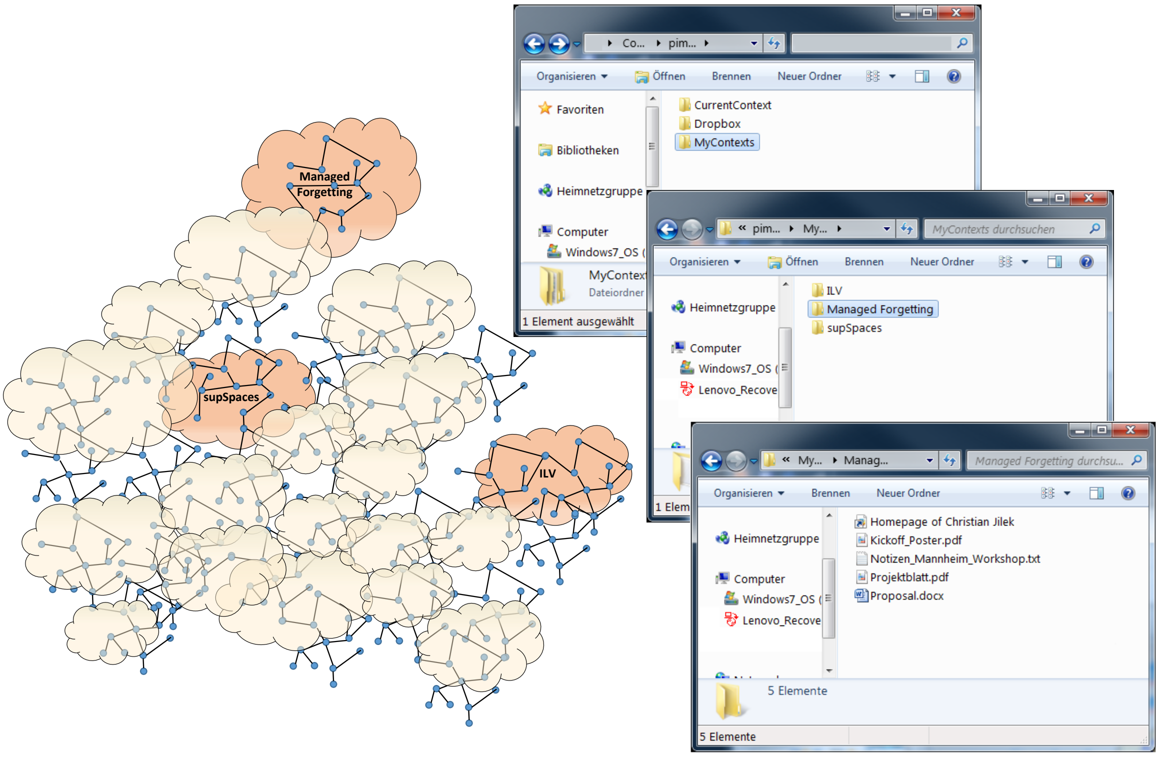


Figure 8.8: Transparent Injection: Traversing cSpaces Using the File System.

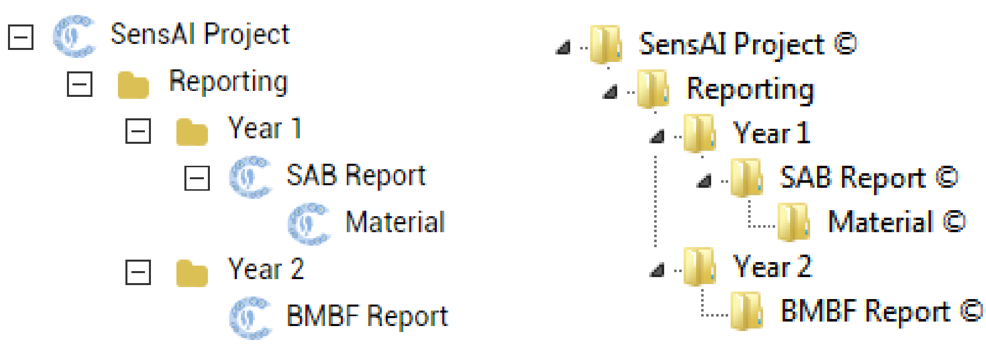
Since related approaches (see Sec. 4.3, especially *Semantic File Systems*) typically only cover one of these systems, e.g. the file system, and particularly do not address user contexts, cSpaces is, to the author's best knowledge, the most transparent and comprehensive version of a Semantic Desktop so far.

However, there are also challenges with such a low-level approach. In Section 5.2.1, for example, the issue of Microsoft Word [Sw84] creating temporary files when opening a document [252] was mentioned. Some editors refuse to display a virtual text document if the byte order mark [385] is missing, etc. Additional code to handle app-specific behavior possibly has to be written. Another tricky problem is a conceptual one emerging from the interplay between context spaces and folders enabled by these injections.

[2]They may be stored directly in the ATIC KG or in arbitrary locations and are registered in the ATIC KG. In both cases, the ATIC KG serves as a proxy or mediator providing the information item.

8.2.3 cSpaces Hierarchies and Interplay with Folders

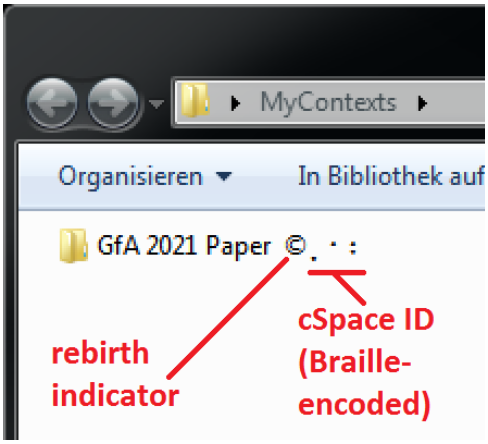
Context spaces can be hierarchical, i.e. each of them can have sub- or super-contexts. The above possibility of injecting them into, for example, the file system may raise the question **what actually is a context and what is "just" a regular folder**. This is as intended since the transition between those entities should be fluent by design (e.g. see remarks on context flavors in Sec. 3.3.3). Imagine the case of a user creating a new context for a presentation and dropping a template file into this new context. Next, a sub-folder called *material* is created. This could either be a "just a folder" of items the user needs to paste into the presentation. On the other hand, it could also be associated with the task of getting more photos for the presentation ready until the end of next week. As mentioned in earlier chapters, making this folder a "task-ish" context should be easily possible. In this case, the user would click on the folder, the sidebar would realize the click and show the folder as well as some basic information about it. Next, the user would change the folder's type from "native (file) folder" to a context with the additional flavor of a task. This triggers the "rebirth" of the folder in the user's PIMO making it a context space. Other actions that trigger "rebirth" are, for example, adding tags, notes, comments, etc. Using the injections, folders that are actually contexts could (automatically by the system) be set apart from regular (native) ones by adding an additional symbol to their name, e.g. the copyright symbol, which looks similar to the cSpaces logo. The symbol is deliberately used as a postfix (as opposed to a prefix) to not distract users and alter folder sorting. Figure 8.9a, a partial screenshot of the cSpaces sidebar, shows an example of such an interplay of context spaces and folders. The top-level context of this example is called *SensAI Project*. There is a sub-folder called *Reporting*, which could also be a context but apparently the user has not "rebirthed" it, yet. It contains two more folders, *Year 1* and *Year 2*. The former contains a context called *SAB Report* with an additional sub-context labeled *Material* that could have emerged exactly like the one of the example given before. If configured to use a copyright symbol as a rebirth indicator, this context and folder hierarchy would appear in the file explorer as depicted in Figure 8.9b.



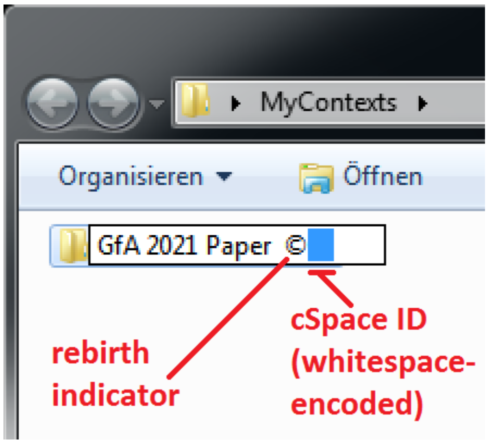
(a) view in the sidebar (b) view in the file explorer

Figure 8.9: Transparent Injection: cSpaces Hierarchies and Interplay with Folders.

A last exemplary challenge to be addressed here involves the **identification of individual cSpaces**. Since the system performs reorganization measures like splitting or merging (see Sec. 3.3.1), the path to a context in the file system, for example, may thus be different before and after a support measures has been performed. This is not a problem when users navigate their injected contexts and folders. However, if applications saved the old path, for example as part of a recently used documents list, cSpaces needs to map the old path to the new one. Typically, requests to the virtual file system can come in high frequency. To avoid performance issues caused by such path reconstructions, it would be helpful if the ID of each cSpace could be encoded into its name, for example. This should preferably be unnoticeable for the user or at least as less distracting as possible. Solutions could involve using Braille [384] symbols as an alphabet to encode the ID and append it to the cSpace's name. Using only a subset with less than three dots already allows for a base37 encoding. Alternatively, an alphabet only consisting of different white space characters could be used, possibly allowing for a base16 encoding [191, 192]. Cases using both alphabets are depicted in Figure 8.10. While white space characters are less distracting than Braille symbols, they could cause problems on some operating systems. Finding a problem-free set of characters is a task for future work.



(a) 3-digit Braille-encoded cSpace ID



(b) 4-digit whitespace-encoded cSpace ID

Figure 8.10: Transparent Injection: IDs of cSpaces Encoded into Folder Names.
The IDs should be encoded as unobtrusive as possible: Braille [384] (possibly a subset using only a few dots) (8.10a) or whitespace characters (8.10b) serve this purpose. Additionally, rebirth may be indicated by another character. For example, the *copyright* symbol looks similar to the context icon used in cSpaces as shown in Figure 8.9a.

The examples of this and the last section were meant to give some impressions of at first hidden problems that emerged when following such a low-level injection approach. Some of the author's implementations are still in an early state, which is why only the web browser injection – as the most stable one – made it into the cSpaces multi-month user study.

Having enabled users to actually work *with* and *in* their different context spaces and having applications respect them, cSpaces can then perform support measures like temporal hiding or reorganization as discussed in the main part of this chapter, the sections after next. First, the aspect of working *with* and *in* context spaces is discussed in more detail in the following.

8.3 Working with and in cSpaces

One major idea of this thesis is to make contexts an explicit interaction element, allowing users to actually work *with* (i.e. a "tangible" object similar to a folder) and *in* them (immersion). This section addresses this aspect in general and measured benefits in particular.

8.3.1 ConTextMarker – Context-sensitive Highlighting and Note Taking

In the following, a tool for context-sensitive highlighting and note taking is presented. (Since it has been used in several studies described next, having its introduction first made sense, although it is the least important topic of Section 8.3.)

Working in context spaces will lead to users possibly viewing the same resource in different contexts. If they wish to highlight certain text passages, it would be helpful if these highlightings were saved in a context-sensitive way. For example, think of a reference book, legal text, etc.: in different contexts, different passages might be relevant.

The developed tool is called *ConTextMarker*, a combination of the terms *context* and *textmarker*, the German term for highlighter. Since cSpaces' front-end components are realized using web technologies, this additional component is also implemented that way. It is basically a web browser add-on based on *WebExtensions* [261].

Note: Although the current focus is on web documents, tools like *pdf2htmlEX* [Sw120] make it possible to "convert PDF to HTML without losing text or format". And since the vast majority of documents can be converted to PDF, this enables the possibility of converting more or less any information item used in PIM to an HTML version whose individual passages can be context-sensitively highlighted.

Three consecutively developed prototypes of the ConTextMarker are presented after a brief overview on related work.

Related Work. The following are approaches coming close to the idea of the ConTextMarker:

- *Annozilla* [Sw124] (2001), an implementation for Mozilla of a W3C LEAD (Live Early Adoption and Demonstration) called *Annotea*, which is/was meant to "[enhance] collaboration via shared metadata based Web annotations, bookmarks, and their combinations".

- *Annotizer* [197] (2006), which allows for asynchronous text annotations for web content with a central database architecture. The paper also gives an overview on other related approaches of the past.

- *dokieli* [Sw31] (2013), which implements the W3C *Web Annotation Data Model* [296].

However, to the author's best knowledge, there is no tool available allowing for *context-sensitive* highlighting in documents, in particular web documents.

ConTextMarker Prototype I. Figure 8.11 shows a screenshot of the first prototype of the ConTextMarker. The user selects a text passage (see A, although still selected, not highlighted, at the time) and clicks on the highlighter button (B). As a consequence, the text is highlighted (A) (in yellow) and added as a note to the currently browsed website in the PIMO (C).

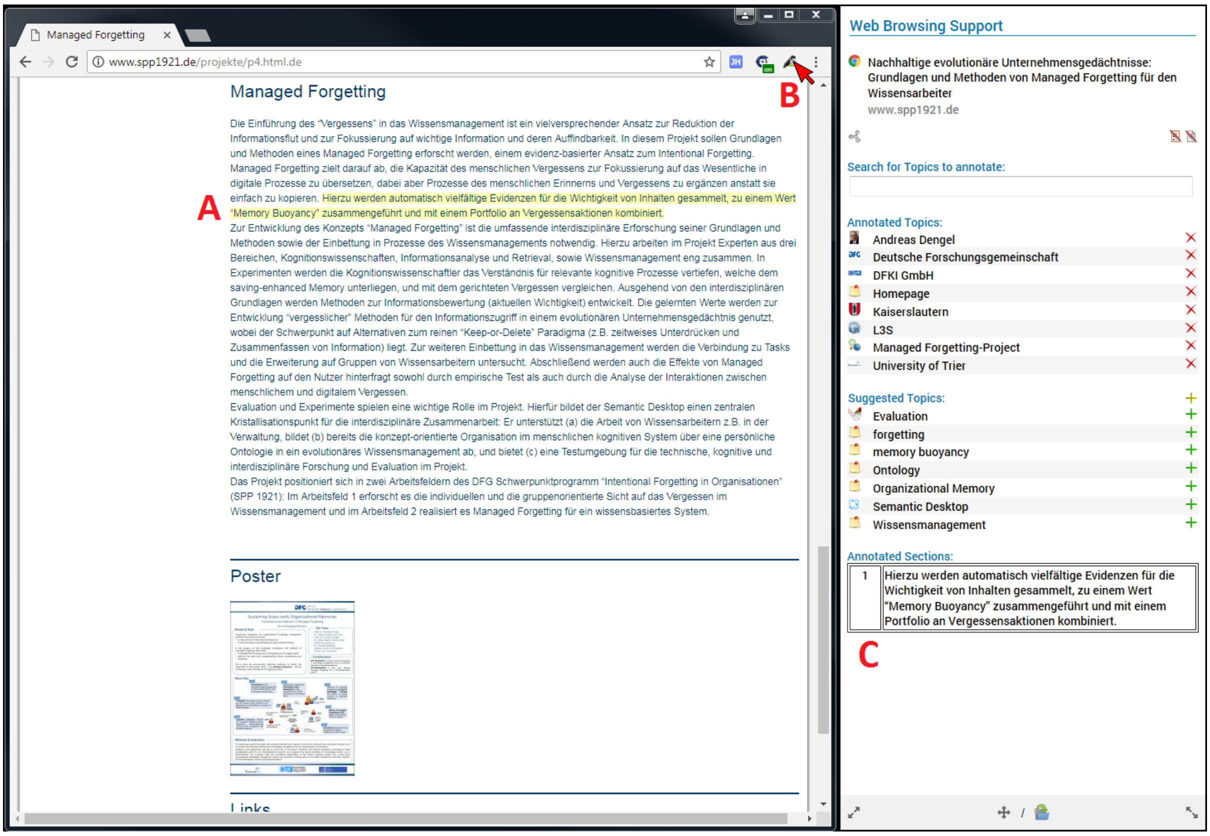
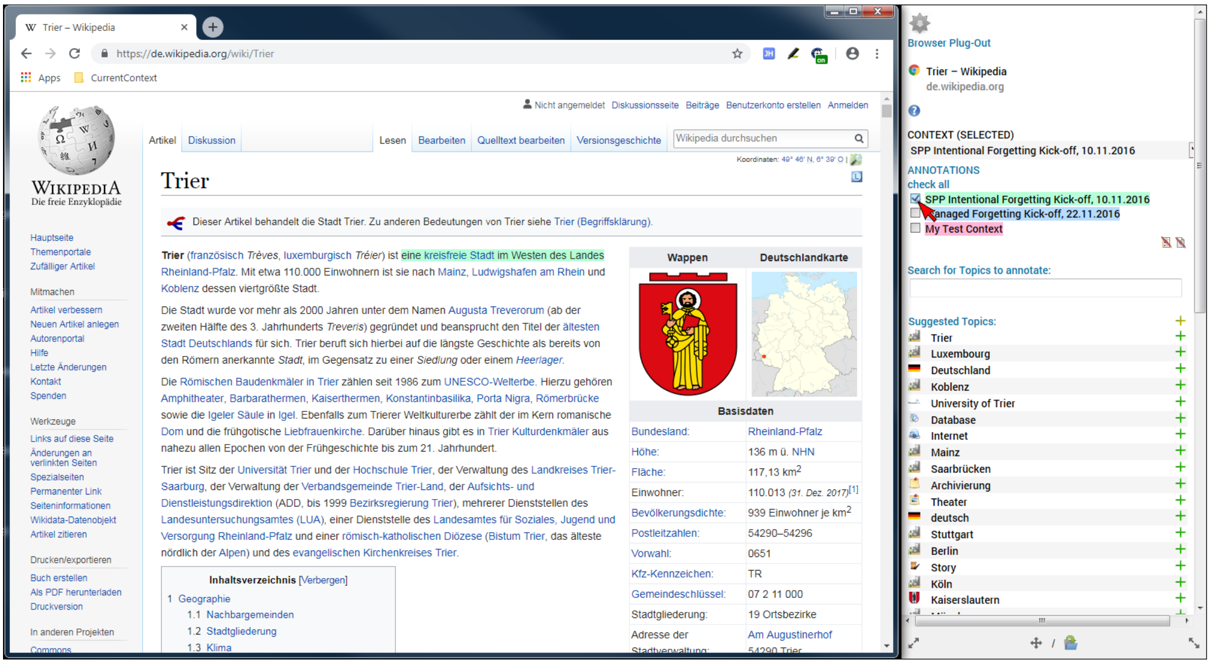


Figure 8.11: ConTextMarker Prototype I.
Highlightings: *A: the previously selected, now highlighted text / B: the highlighter button (as a web browser add-on) / C: the highlighted text as a note added to the browsed website in the* PIMO.

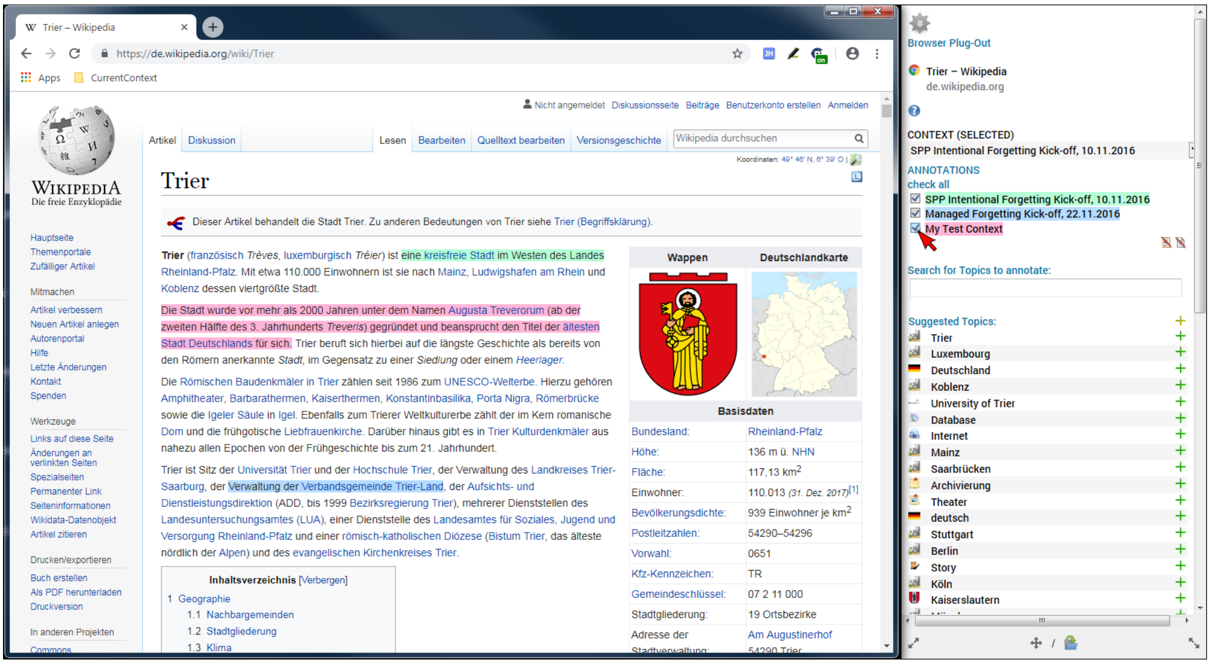
A slightly improved version is depicted in Figure 8.15. Again, users may select a text passage (A, not highlighted at the time) and click the highlighter button (B). The text is highlighted (A) and added as a note to the website in the context (visible between highlightings E and F).

Prototype I has been used in various studies presented in the remainder of Section 8.3.

ConTextMarker Prototype II. A second prototype allows for displaying all highlightings of a website regardless of the context the user is currently in. As shown in Figure 8.12, the user is currently in the *SPP Intentional Forgetting* context. But they may also select highlightings of other contexts to be displayed (see upper right part of the figure). In Figure 8.12a, only highlightings of the current context are enabled, whereas in Figure 8.12b, the highlightings of the other two contexts are shown as well. Each context is associated with a different color. While the focus of Prototype I was more on turning selected text passages into notes in the contexts (highlightings disappeared upon reloading or revisiting the webpage, notes did remain though), this second prototype is able to re-highlight the passages upon selection in the sidebar. What may sound trivial is actually a more complex problem since websites are often different when next visiting them. Such changes may occur from updates of the



(a) highlightings of first context enabled (green)



(b) highlightings of the other two contexts enabled as well (blue and pink)

Figure 8.12: ConTextMarker Prototype II.

content management system injecting different meta-tags into the page or by adding a different advertisement, etc. Re-finding the right passages is thus a non-trivial task. Work on Prototype II was supported by Rudolf Koch (re-highlighting) and Michael Kraus (experiments with pdf2htmlEX) as part of their student research assistant jobs at the time.

ConTextMarker Prototype III. So far, users could add context-sensitive notes to individual items (e.g. a website) in a context. A third version of the ConTextMarker, currently in pre-prototypical state (i.e. only code snippets available, not combined as a full application), also allows for adding notes to the context itself. As shown in Figure 8.13, the highlighter available using the context (right-click) menu is in this prototype. In the middle section of the right menu, one can see the options to *add new context note from selection* or *add new webpage note from selection*. The latter is identical to the functionality of previous prototypes. Another option above these two is *note append*. Text can be added to the last note the user has added or written. Regardless whether this was a note to the context or one to an individual item, the selected passage will be appended to it. This allows for easily assembling a kind of summary text by using note appending on all relevant passages. Other options shown in the screenshot are not related to highlighting: adding a webpage to the context (either as a link or a copy) and refreshing the context injections (mostly for debugging or if a websocket connection was lost).

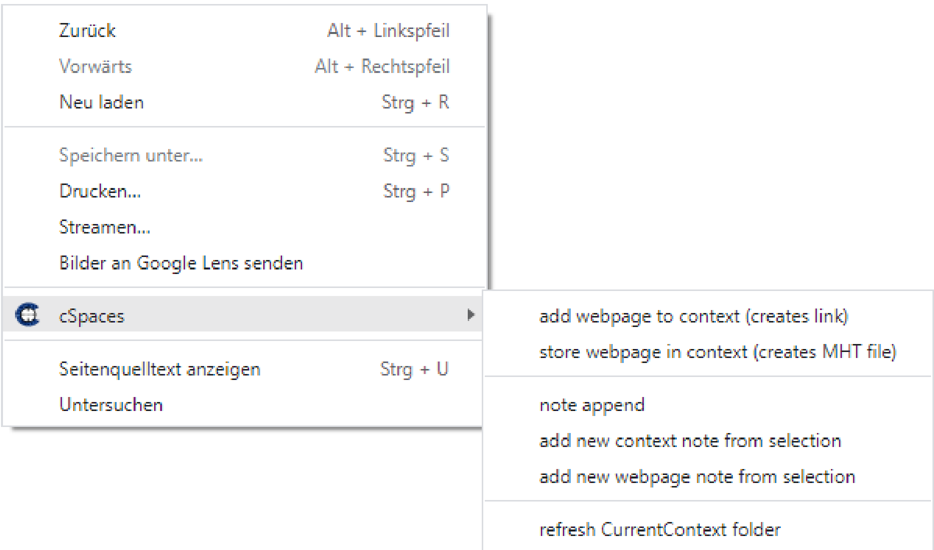


Figure 8.13: ConTextMarker Prototype III.

User Survey. In the cSpaces multi-month user study, participants that have used cSpaces (and have thus experienced working *with* and *in* context spaces for several months) were asked about contextual highlighting and notes. Although the ConTextMarker was not part of the study, all participants knew the app and were also shown screenshots as a reminder. The asked questions were mainly about Prototype III: They found contextual notes to be easy to understand (Q_{M1}, median: *strongly agree*), which would allow them to write things down exactly where they would like to have them (Q_{M3}, mean and median: *quite agree*). Most participants stated to not have seen such functionality in applications before (Q_{M2}, median: *strongly disagree*). The only participant who *quite agreed* restricted the given answer by adding "not in this extent". All answers are depicted in detail in Figure 8.14.

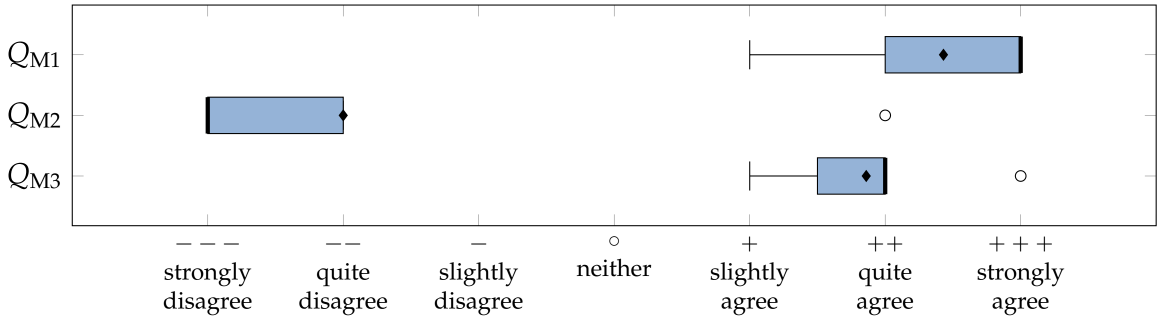


Figure 8.14: Multi-month User Study: Survey Result for Contextual Notes.

Q_{M1}: The concept of contextual notes (vs. general ones) is easy to understand.
Q_{M2}: I have already seen contextual notes in existing software applications.
Q_{M3}: Contextual notes would allow me to write things down exactly where I would like to have them.

(n=7; full questionnaire document in Sec. A.4; details on box plot semantics in Sec. B.2)

With the ConTextMarker given as a tool to participants (among others), the next sections report about two of the largest studies performed as part of this PhD project.

8.3.2 Cognitive Offloading Effects

Cognitive Offloading. Have you ever turned your head to read a tilted headline more easily? If this is the case, you have been performing a form of *Cognitive Offloading* [289]. Every strategic use of physical actions like tilting your head or using a calculator to reduce cognitive demands is defined as Cognitive Offloading. Only two decades ago, people still learned phone numbers by heart in order to be able to call somebody. Today, one simply selects a person's name in the smartphone and a connection will be established (assuming that the number has been entered to the list of contacts before). That way, several devices are used as an extended external memory store. Research in cognitive psychology showed that such memory offload can have comparable benefits for subsequent cognitive performance as discussed for intentional forgetting in Storm and Stone [365] and Runge et al. [294]. As shown in Runge et al. [294], the possibility to store information externally can even be seen as an implicit cue to intentionally forget the offloaded information as long as you can rely on the device to continually store it [365]. In the cSpaces scenario, one may think of contexts that "tell" the users what they have done the last time when they were working in/with them: what documents have been read or written, what are open tasks to be performed, etc. Such implementation would allow the user to rely on the Semantic Desktop as their external memory for all reached work progress. The intentional forgetting of externally stored work progress could then benefit subsequent tasks and ease the switch from one context to another. Cognitive psychology experiments like Tempel and Frings [372] revealed that intentionally forgetting about one task can enhance subsequent cognitive performances like encoding and recall of word material. A prominent explanation for these benefits is memory inhibition of intentionally forgotten information [38]. The main research question to be investigated in this study was whether these findings could be transferred to more complex scenarios like IM and KW supported by a Semantic Desktop like cSpaces. In other words: does allowing users to intentionally forget about their recently irrelevant contexts – thus performing cognitive offloading by relying on cSpaces to keep the information for them – increase their performance on the current context?

Experiment Setting and Procedure. Together with colleagues of the Managed Forgetting project [234], an experiment was set up, in which participants were asked to use cSpaces in order to do two subsequent web researches: one about a historical person (Simón Bolívar) and the other about a country (Laos). Participants could browse through a limited number of standardized web pages, while being supported by the assistant. The cSpaces prototype, depicted in Figure 8.15, mainly consisted of three parts: the context selection (1), a widget to view and store items in the context (2) and a reverse-chronological list of activities performed in the selected context (3). Visited pages were automatically added to the activity history by the system. Other activities like taking notes were listed as well. Clicking the green plus next to a website of the activity list (G) would add that page to the context. Red crosses could be used to remove items from the context or the activity list. Participants could use Prototype I of the ConTextMarker (see Sec. 8.3.1) to highlight relevant sections of a webpage, and by clicking the highlighter button (B) the text was added as a note to website in the context (F). If the website had not been added to the context yet, this was done automatically in the process (D). In addition, participants could also write notes manually by clicking the notepad icon (E) next to each web page in the context. This opened a small text area right below the item to enter the note. After each research phase, the participants were asked to switch to another context, which was done using the context selection field (C). Additionally, they were asked to click the tab closing button (H), another browser add-on ensuring that all browser tabs and windows except for the experiment instructions were closed when switching to the next context. The system ensured that these steps were actually taken by not allowing further inputs or browsing in addition to showing respective instructions on the screen.

In both web searches, participants were given a profile showing which facts they should look for. These were eleven facts per topic like the manner of death or place of birth of Bolívar or the birth rate or form of government of Laos. The facts were spread across several standardized websites, for example excerpts of Wikipedia [398] with hyperlinks removed (to ensure that all participants only browsed the same given set of pages). Each search started at a search screen, depicted in Figure 8.16 (A) that was deliberately implemented to look like *Google Search* [Sw49], which is typically known by participants (logo blurred in the screenshot). Result lists were static but different for each topic, thus participants' search skills were not relevant. Everyone saw the same number of results and in the same order.

Before the two search tasks, there was a short interactive tutorial, shown as a green box at the top of the sidebar (highlighting B in the screenshot). Participants had to perform several actions like adding a website to a context or highlighting text to get to know the system. Thus, by the start of the experiment's main phase, participants had already used each feature at least once, some of them at least twice. Ticked checkboxes for each type of action informed about the tutorial's current state of completion. The experiment was conducted in sound attenuated cabins equipped with a single workstation at the University of Trier. cSpaces was deployed together with a portable version of the *Opera Web Browser* [Sw98] ensuring the right add-on configurations and a permanent privacy mode (i.e. the app does not store a browsing history

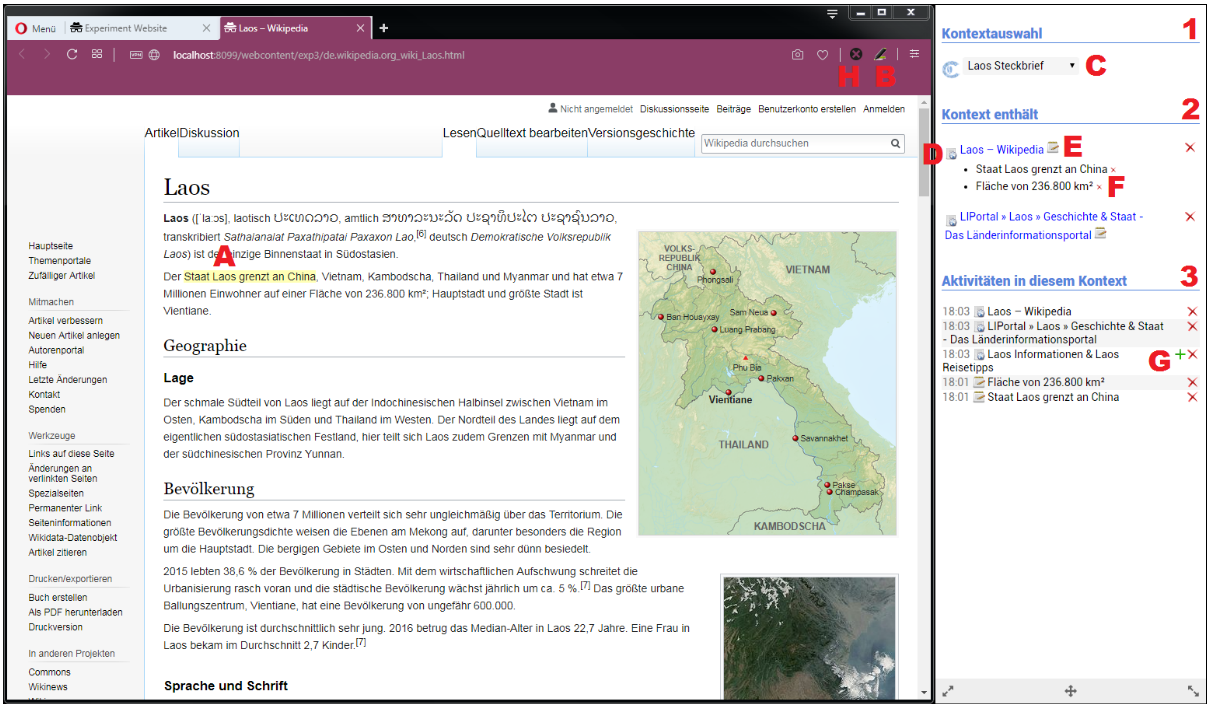


Figure 8.15: Cognitive Offloading Study: cSpaces Sidebar.
Highlighted sections of the sidebar: *1: context selection / 2: content of selected context (websites and notes) / 3: history of activities in selected context (reverse-chronological order).*
Highlighted features: *A: text selected by the participant is highlighted and converted into a note associated with the current website (website is added to the context if yet not part of it) / B: browser highlighter extension / C: drop-down menu for context selection / D: items (websites) in selected context / E: button to add a note to particular website (notepad icon) / F: button to delete note from particular website (red cross) / G: buttons to add item of history list to context (green plus) or remove item from history list (red cross) / H: button to close all browser tabs and windows except for the experiment instructions.*
(Note: This screenshot has already been published in Gauselmann et al. [126].)

or highlights already visited links, etc.).

The first research phase was interrupted after 4 min (the second after 3 min) by a math task, in which participants had to solve modular arithmetic problems as stated by Bogomolny [40] (also referring to Gauß [127]), which are known to “place heavy demands on working memory” [(24) referring to (15, 16)]. Participants did not know in advance that this interruption would happen and there was a time limit of seven seconds for solving each equation as well as a delay of 500 ms before the next one was shown. After switching the context from their particular research task to the mathematical one, the system would either keep or delete their research context. Participants got the instruction which of the two cases was chosen by the system. Keeping or deleting was randomized, so if a participant’s research context was deleted in their first trial, it was saved in their second one and vice versa. The order of the research tasks (Bolívar or Laos first) was also randomized. After the interruption by the math task, participants were asked to enter results into the fact sheet and afterwards, they switched back

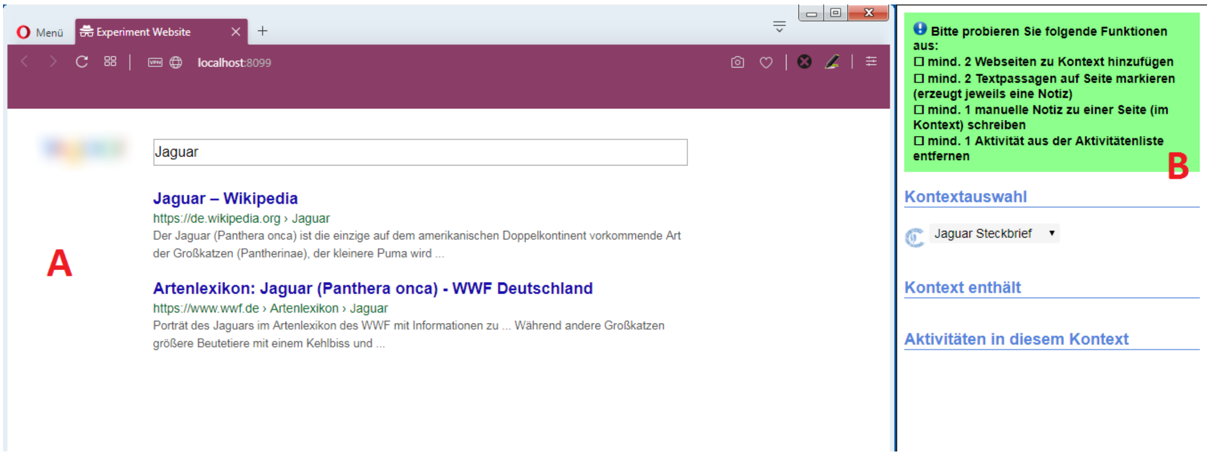


Figure 8.16: Cognitive Offloading Study: Static Search Results and Initial Tutorial. Highlightings: *A: screen of search results (deliberately implemented to look like* Google Search *[Sw49] [logo blurred]) / B: interactive tutorial (ticked checkboxes for each type of action inform about the tutorial's state of completion)*

to continue their research in a second phase. In cases of keeping the research context, the sidebar contained exactly the same content as before the switch. In deletion cases, participants started the second phase with an empty sidebar. The second phase also ended with a math task and displaying of the fact sheet to possibly enter further answers. In summary, the individual phases of the experiment's main phase are shown in Figure 8.17: two research topics (trials) split into two inquiries, both interrupted or ending, respectively, with a math task (arithmetic blocks highlighted in blue) and showing of a fact sheet to be filled out (test).

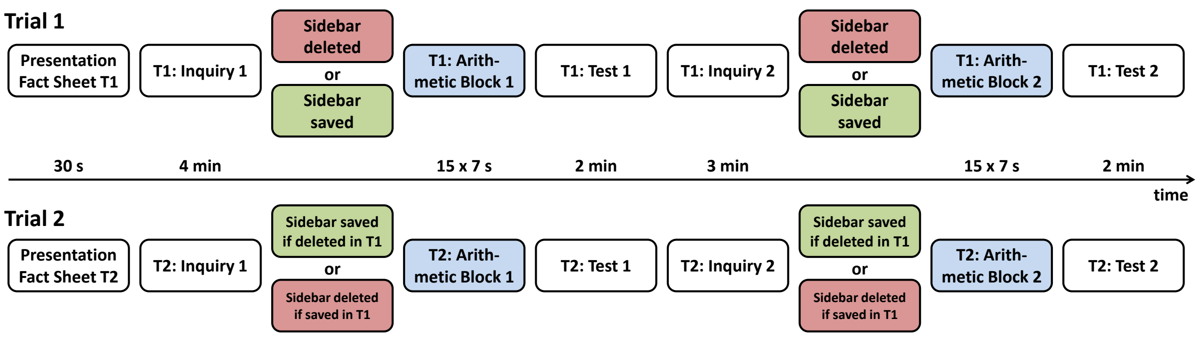


Figure 8.17: Cognitive Offloading Study: Experiment Setup.
(Note: A variant of this figure has been published in Gauselmann et al. [126].)

Results. The study involved 48 persons. As depicted in Figure 8.18, the mean number of correctly answered equations in the first inquiry was 10.31 if the research context was saved, while it was 9.42 in the case of its deletion. No such effect could be observed in the second inquiry (see discussion below). A value of 0.393 of *Cohen's d* [68] indicates that this is a weak effect. However, the difference in means is significant. A paired t-test [366] was used to compare the first inquiry of the two groups, G_S (context saved) and G_D (context deleted): $n = 48$, $\bar{x}_S = 10.31$, $\bar{x}_D = 9.42$, $s_S = 3.44$, $s_D = 3.21$, $r_{x_S x_D} = 0.77$, $p < 0.01$.

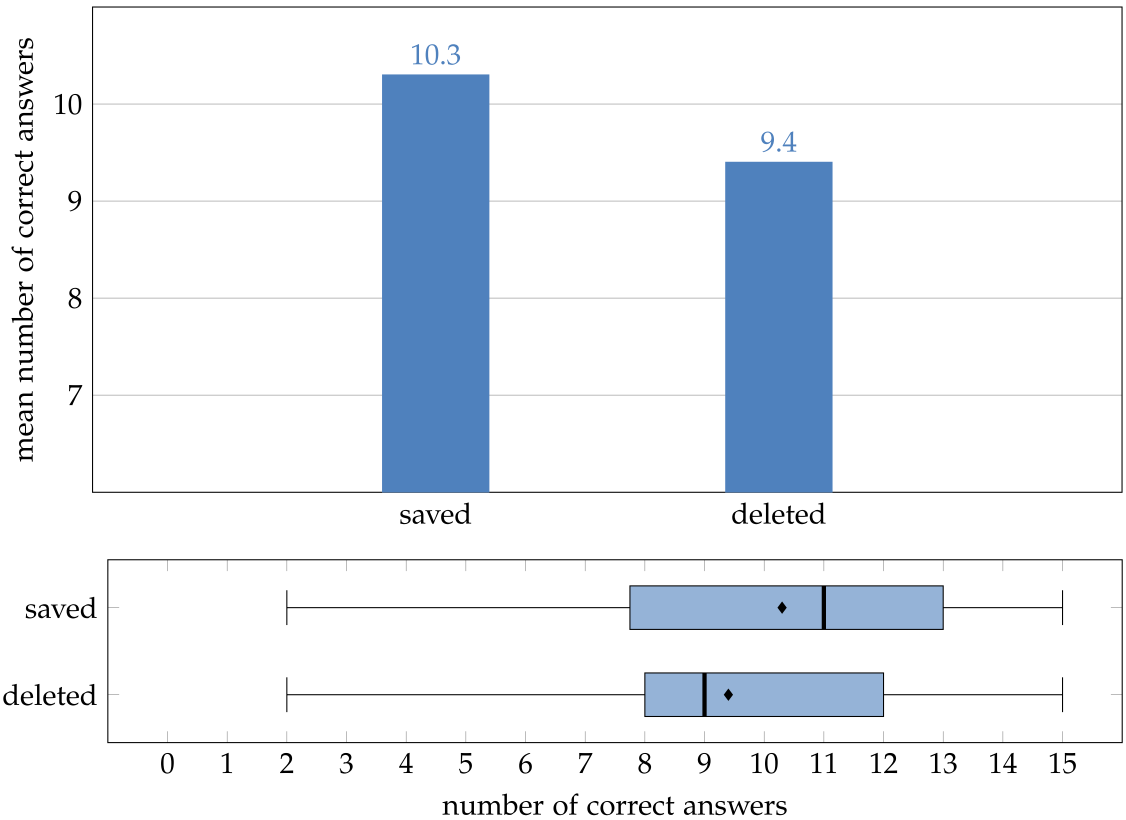


Figure 8.18: Cognitive Offloading Study: Number of Correctly Solved Arithmetic Problems (in Inquiry 1).
(n=48; details on box plot semantics in Sec. B.2;
Note: A variant of this figure has been published in Gauselmann et al. [126].)

Discussion. In situations of the research context being saved, participants could "cognitively offload" their so far collected facts in the research task to the cSpaces sidebar and thus free working memory capacity, which then was additionally available for the math task (i.e. a task completely unrelated to the previous one). The effect only occurred in the first inquiry, while no significant effect could be observed in the second one. This can be explained by participants typically only looking for a "last few" missing facts in the second inquiry, i.e. taking lots of notes in the sidebar was thus not necessary. This interpretation is supported by the measured use of the sidebar, which was less in the second inquiry than in the first one.

For further details, especially also from a cognitive psychology perspective, readers are kindly referred to the journal article about this experiment [126].

fNIRS-based Follow-up Experiment. In a follow-up experiment, the cognitive load of participants is intended to be more directly measured by means of **functional Near-Infrared**

Spectroscopy (fNIRS), "a neuroimaging technology for mapping the functioning human cortex" [115]. All other experiment conditions are identical. A technical challenge here was getting the communication of cSpaces and the fNIRS device (photos in Fig. 8.19) to work (succeeded after a few iterations). By the time of writing this thesis, the study was not yet complete. Again, experiments are conducted at the University of Trier, in this case especially supported by Christoph Geissler, who recently completed an fNIRS-focused PhD thesis [128].[3]

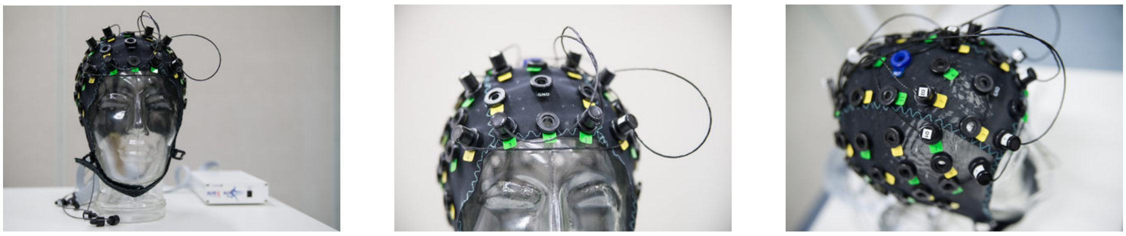


Figure 8.19: Cognitive Offloading Study: Functional Near-infrared Spectroscopy Device. (Image source: fNIRS Laboratory website, Department of Cognitive Psychology, University of Trier, URL: https://www.uni-trier.de/universitaet/fachbereiche-faecher/fachbereich-i/faecher-und-institute/psychologie/professuren/cognitive-psychology/laboratories/fnirs-lab-1.)

8.3.3 Effects on Task Resumption/Switching

A second study was conducted to investigate effects of using cSpaces on task resumption/switching. The experiments were actually part of a master thesis with a slightly different focus [337]. However, gathered data in the form of user activity logs and screen recordings (collected at the University of Trier as in the previous experiment) also allowed for a data-driven secondary study to quantify participants' so-called **Task Resumption Lag (TRL)** [10]. TRL is understood as the time a person needs to re-adjust their mind to the former context in order to continue where they left off.

Experiment Setting and Procedure. The experiment was conducted with 51 participants (65% female, mean age of 26.3, mostly students of humanities) in a knowledge-intensive multi-tasking scenario. Participants had to plan a barbecue evening which involved getting an overview what people would like to eat and drink, what was brought to the party by guests and what still had to be ordered. Additionally, they had to check various vendors for prices and offerings to finally fill an order list containing everything to be ordered for the party (see Fig. 8.20). For this, they had to read files, mails and websites. For the sake of a more controlled experiment, the former were also moved to the browser thus being available as if using a respective cloud service. As a consequence, logging each action was reliable and granting or denying access to each information item could be fully controlled.

The technical setting was basically the same as in the previous experiment (see Sec. 8.3.2) with the cSpaces prototype tailored to particularly support web research activities. There were, however, three different groups, each having their own type of sidebar with varying sets of features as shown in Figure 8.21.

[3]During this thesis' review process, the study could be completed. In short, the fNIRS study confirmed the results of the previous experiment. A journal paper with further details has been published [129].

Bestellung für das Grillen am 03.08.2019

Bitte jeweils eintragen, bei wem wir bestellen, wie viel (in Liter, Kilogramm, Flaschen oder pro Stück) und was es kostet!

was?	wo bestellen?	Menge	Gesamtpreis
Mineralwasser			
Saft			
Wein			
Bier			
Radler			
Grillkäse			
Grillfleisch			
Bratwürste			
Grillsauce			
Baguette (oder anderes Brot)			
Maiskolben			
Aubergine			
Zucchini			
Paprika			
Champions			
Antipasti			
Salat (vegan)			
Salat (vegetarisch)			
Salat (sonstige)			
vegetarisches Grillgut			

Figure 8.20: Task Resumption Study: Order List to Enter Results.

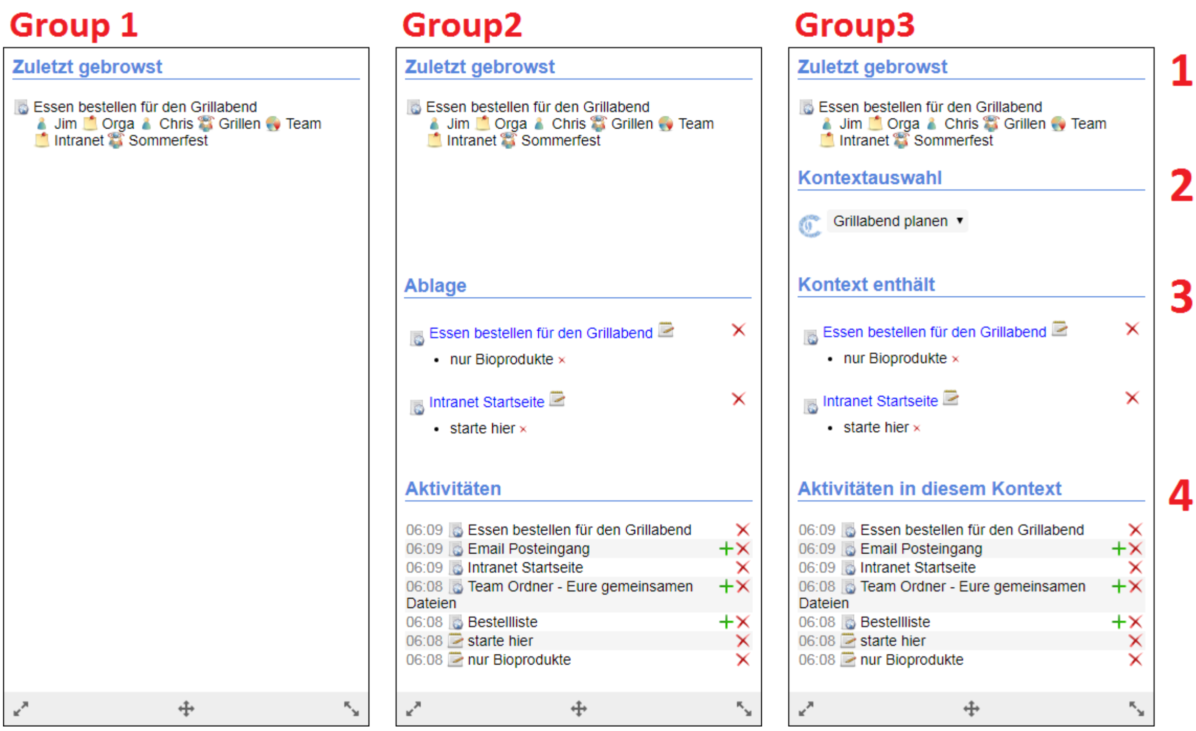


Figure 8.21: Task Resumption Study: Used cSpaces Sidebars.

Highlighted sections: 1: last browsed website with brief content analysis (all groups) / 2: context selection (group 3 only) / 3: selected context (group 3) or context-insensitive tray (group 2) to store websites and notes / 4: history of activities in selected context (group 3) or in general (group 2).

G_1, as a control group, only had the content analysis widget (1), i.e. a non-interactive sidebar. G_2 had the additional possibility to add items to a single context-free folder (3) as well as adding notes to them. Their activities were also captured in a history list (4). G_3 also had these possibilities but with the additional help of items and activities being separated by contexts. G_3 thus was the only group being supported by all four widgets depicted in the figure.

Figure 8.23 also shows the sidebar types of the three groups but with an explicit focus on context switches. One may see that the two depicted contexts (one highlighted in yellow, the other in pink) mix up in the sidebar of G_2, whereas they stay separated in the one of G_3. G_3 may switch between contexts as depicted in the right half of the figure. Note that the list of contained items in the context (3) is ordered chronologically, whereas the list of activities in the selected context (4) is in reverse-chronological order.

After a short tutorial on using the prototype, users began working on the main task. They were given the information that they would be interrupted to switch to a second task – from the perspective of the experimenter a distraction task – which was about working on a wiki entry about autonomous driving. Each participant spent in total 20 minutes on the main and 12 minutes on the distraction task. Three switches between the two were triggered leading to an overall number of 153 observed context switches as well as one hour of video material and 1000 logged activities on average per participant. The experiment's different phases are depicted in Figure 8.22.

Measuring TRL in this secondary study was a hard task since it completely relied on captured evidences in this case. Since participants did not state when they felt their switch back was complete, this was assumed in two cases: 1) when getting back to pages actually containing content needed to fill the remaining lines of the order list and 2) adding new items or notes to the current context. To detect 1), all 40 web pages used in the study were classified. They were either pages containing offerings and prices (17), amounts of food or beverages brought or requested by guests (5), navigation pages (10), spam without any relevant information (7) or the order list to enter results (1). Next, this classification was applied to the activity logs in order to calculate the TRL for each of the 153 context switches. Last, these results were compared to the captured video material and corrected where necessary. For instance, this was the case when users navigated through several pages very quickly, staying on each page for less than a second. Even if a page classified as content was among those, it was assumed that the detention time was too short to actually conceive anything written on that page. Thus, the next content page visit or sidebar addition in the activity log was used to calculate the TRL.

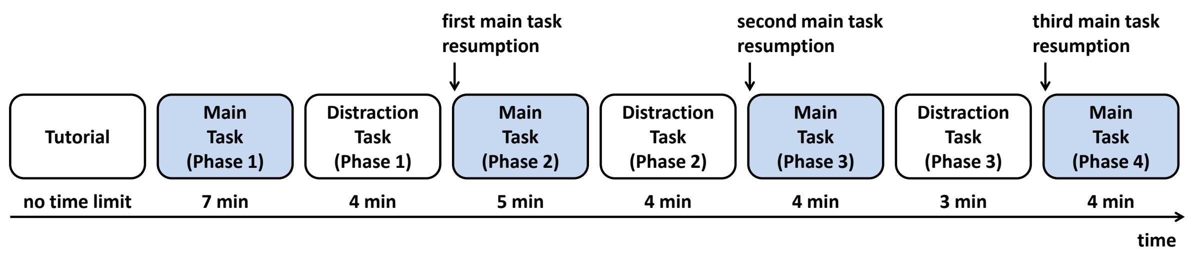


Figure 8.22: Task Resumption Study: Experiment Setup.

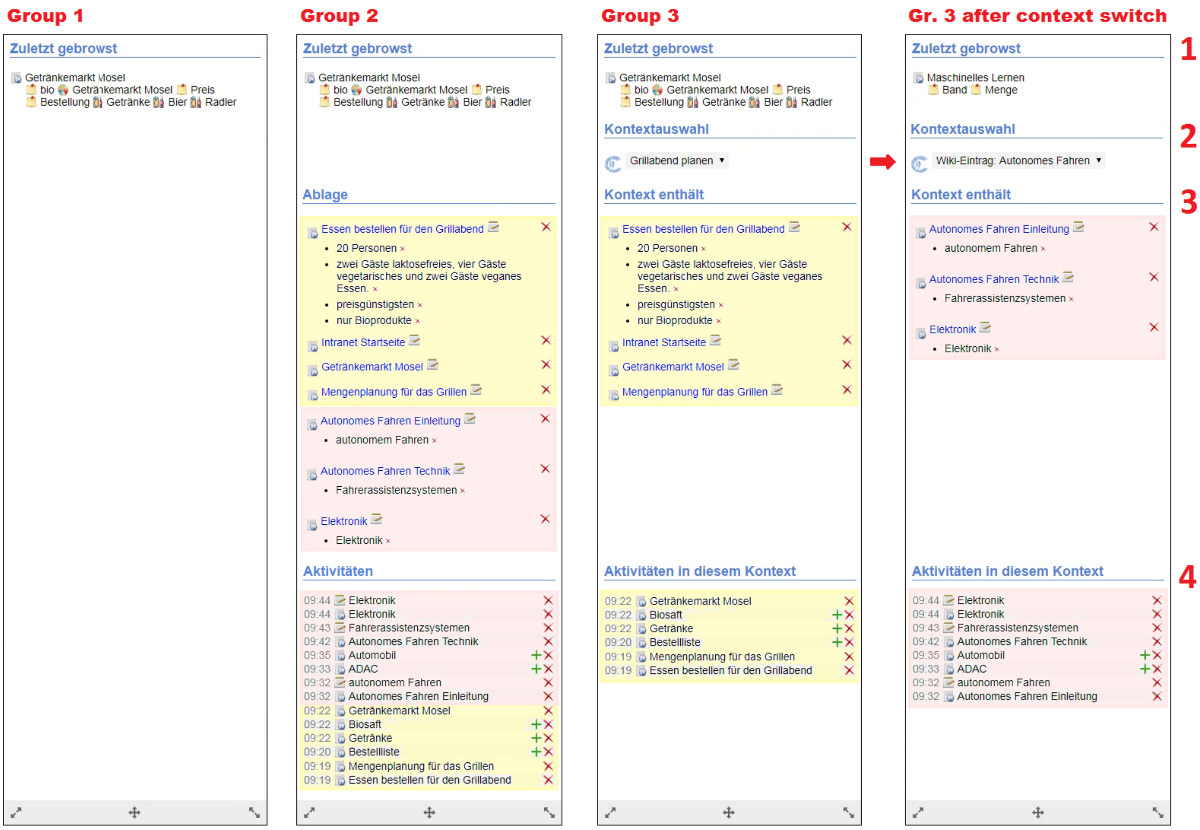


Figure 8.23: Task Resumption Study: How Context Switches Affect Sidebar Content. Since the sidebar of group 2 does not allow context switches, the content of the first (yellow highlighting) and second context (pink) get mixed up. Note that websites and notes are ordered chronologically (sidebar section 3), while the history of activities is in reverse-chronological order (sidebar section 4). Context switches in group 3 (right half of the figure) ensure that content and activities of different contexts do not mix up. Highlighted sections: *1: last browsed website with brief content analysis (all groups) / 2: context selection (group 3 only) / 3: selected context (group 3) or context-insensitive tray (group 2) to store websites and notes / 4: history of activities in selected context (group 3) or in general (group 2).* (Note: A variant of this figure has been published in Jilek et al. [183].)

Related Work. Shifting file and mail interaction to the web browser for the sake of having a more controlled experiment (as explained above) has the consequence that works about web-based multi-tasking become most relevant. *SearchBar* [259] and *Multitasking Bar* [390] are two exemplary tools, each investigated in a two-session study, involving 16 participants solving several web research tasks while being interrupted from time to time. Both tools were integrated into the browser to capture, store and present users' search and browsing history. Participants could create topics or tasks, respectively (contexts in the terminology of this thesis), so subsequent actions were automatically associated with the selected topic/task. Both studies gathered quantitative as well as qualitative feedback, e.g. about the tool usage, task resumption, re-visitation and re-navigation. However, they both did not measure and compare task resumption times when switching tasks after a few minutes. They rather had a particular focus on task resumption after about one week of pause in between sessions. Both studies showed that their assistants had various positive effects on users' work, especially improving their performance.

Results and Discussion. The average TRL for group G_3 having a context-sensitive sidebar was 10.5 s, followed by the TRL of G_2 of 11.0 s and finally a value of 18.3 s for the control group G_1 (results depicted in Fig. 8.24).

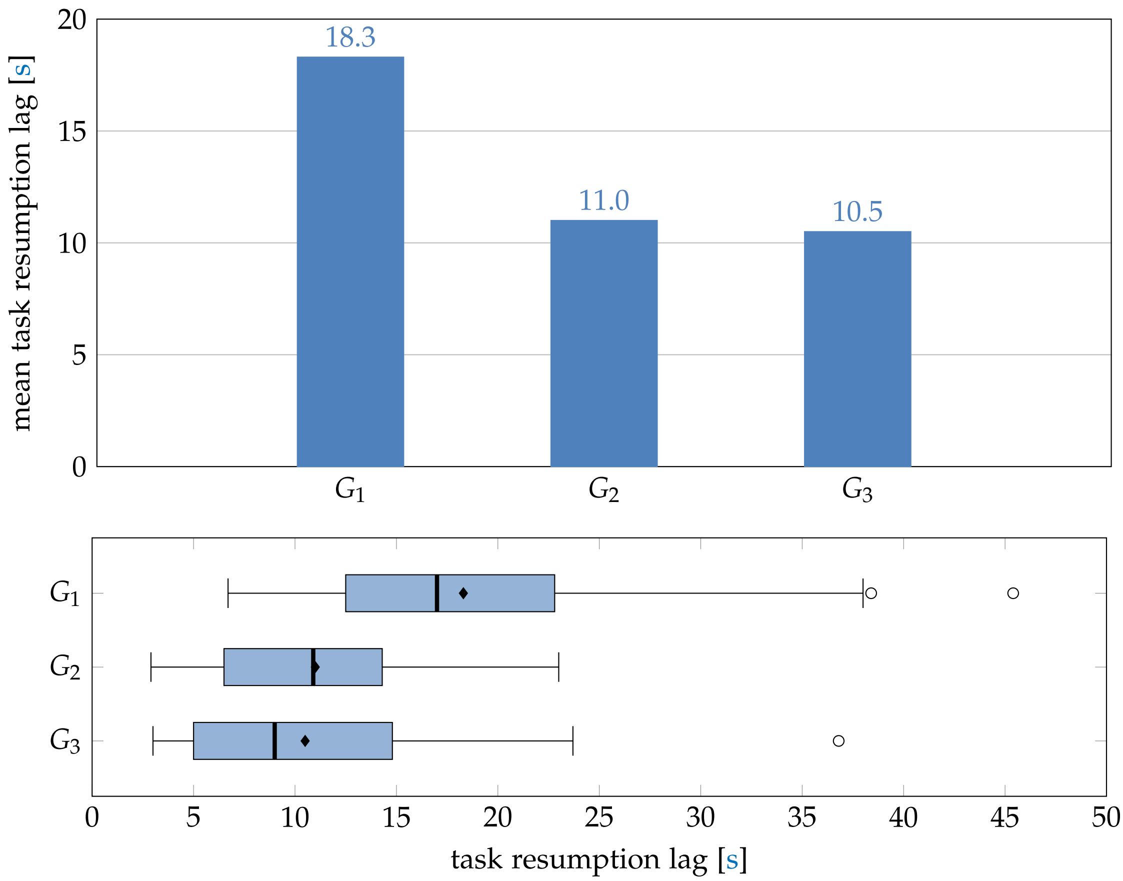


Figure 8.24: Task Resumption Study: Task Resumption Lags.
G_1: control group with non-interactive sidebar / G_2: group with context-insensitive sidebar / G_3: group with context-sensitive sidebar.
(n=51; details on box plot semantics in Sec. B.2)

cSpaces-supported groups were thus about 40% faster in resuming their former context than the control group. The small difference between G_2 and G_3 (not significant) can be explained by only having two different contexts, of which one even was a distraction task and discovered as such by some participants: rare cases of G_2-participants taking the time to tidy up their sidebar after returning from the distraction to the main task could be observed. They removed items and notes before continuing. It is worth noting that in post-experiment interviews, 78% of G_2 participants stated (or at least insinuated) the need for a context-sensitive sidebar not mixing things up. More contexts would have presumably been necessary to see a larger TRL difference between the groups G_2 and G_3. Comparing G_2 and G_3 to the control group G_1 yields *Hedges' g* [158] of 0.97 and 1.02, respectively, indicating a large effect in both cases. Results are significant. G_1 vs. G_2 and G_1 vs. G_3 were compared in unpaired t-tests [366]: $n_1 = 63$, $n_2 = 27$, $n_3 = 51$, $\bar{x}_1 = 18,273.14$ ms, $\bar{x}_2 = 10,969.59$ ms, $\bar{x}_3 = 10,548.14$ ms, $s_1 = 8,196.19$ ms, $s_2 = 5,770.1$ ms, $s_3 = 6,741.81$ ms and additional u-tests [235] having $p < 0.001$. Note that four participants were excluded from G_3 since they did not use the sidebar to do the tasks, i.e. they had zero or one sidebar actions during the whole experiment.

For further details, especially also from a cognitive psychology perspective, readers are kindly referred to this experiment's conference paper [183].

8.3.4 Insights of a Multi-month User Study

Scope and Goals. As explained in Section 8.1.1, several cSpaces support measures need some time to have passed. This either rules out short-term studies or forces bringing out effects artificially. Therefore, a medium/long-term study was planned as a complemented to the short-term experiments. However, with the multi-month user study being embedded into the PhD project, not every feature, especially those being developed or fine-tuned in the project's last phase, could make it into the study. As mentioned in Section 8.2, some features were also not yet stable enough to already be used 24/7 by arbitrary participants, possibly with a non-technical background. While the study was in progress, a few additional features were still added whenever they were ready. Nevertheless, the focus of this multi-month user study thus shifted more towards the aspect of enabling users to actually work *with* and *in* their different context spaces. (That is why the study is mentioned in this section although addressing other topics as well.)

With regard to gathering data and user feedback for evaluation, the study consisted of three lanes:

- **Gathering quantitative data**: For example, the *Stream Context Miner* (see Sec. 6.3.2) was evaluated with data obtained in the study. Further exploitation of the collected dataset is part of future work. By the time of writing, some colleagues, e.g. Mahta Bakhshizadeh, already use the dataset for their research.

- **Gathering qualitative feedback on features used by the participants**: Those cSpaces features available in the study could be used by the participants for up to five months. They were thus asked to provide qualitative feedback on these experiences.

- **Asking participants to estimate the potential of features still in an early state or envisioned**: Participants have experienced multiple features of cSpaces for several months. Although these features were only a subset of all ideas and prototypes developed in this PhD project, they got deeper impressions and insights of how things work and what is or could be possible. For example, some of the features not in the study are based on components that the participants actually used. For other features, participants saw demonstrations of early prototypes and thus got impressions that they could combine with their experience from the study. They were thus asked to extrapolate this (multi-month) experience with the system in order to assess the potential of features still in an early state or envisioned.

Setting. The study was conducted from August 5th, 2021 to January 7th, 2022 (156 days) and involved seven participants, all of them colleagues of DFKI SDS (see remark on potential courtesy bias below in this regard). In order to get realistic data, participants were asked to use the cSpaces application in their daily work. The extent of this inclusion was up to each participant to decide since cSpaces is still a comparably fragile research prototype after all, i.e. a loss of productive data (although rather unlikely) could not be ruled out. Usage statistics are provided in Table 8.1.

	P1	P2	P3	P4	P5	P6	P7	Total
days of use	7	9	45	26	38	91	33	249
days of availability	154	154	152	156	149	139	51	955
therefrom working days	109	109	107	111	106	98	37	677
num. of captured events	2,551	1,795	5,663	952	5,733	25,767	4,091	46,552
num. of created cSpaces	12	15	28	22	27	57	12	173
therefrom hierarchical	6	1	4	2	2	4	1	20
num. of browsed files	3	11	19	67	18	47	0	165
num. of browsed websites	294	374	1,218	40	1,417	4,209	841	8,393

Table 8.1: Multi-month User Study: Usage Statistics.
Evaluation period from 2021-08-05 to 2022-01-07. Abbreviations: *num.: number*.
Note: Some participants joined the study later or submitted their evaluation data a few days later than others leading to individual *days of availability* for each participant. The feature of *hierarchical contexts* was only available for a few days towards the end of the study, thus their number is comparably low. Individual participants' vacation days were not considered in the calculation of *working days*.

One can see that some participants used the app almost every work day (93% or 89% of their work days). Other participants that used the app less regularly still had days of intense usage also yielding hundreds of tracked events. The number of usage days over all participants accumulated to 249 days (individual vacation days not considered). In total, 173 contexts have been created (20 of them hierarchical[4]), 165 files and 8,393 websites have been browsed using the app leading to 46,552 events being observed. Privacy preservation measures were as described in Section 6.1.2 (in short: transparency of measures, possibility to turn off the observation, only manual data upload and data anonymization).

The cSpaces prototype used in the study is depicted in Figure 8.25. It is similar to the ones used in previous studies (see Sec. 8.3.2 and 8.3.3), with the main difference of allowing users to create their own context spaces, in particular also hierarchical ones. Looking at the GUI in detail one may see the currently selected context (A), a button to create a new context (B), a possibility to turn the observation on and off (C), the context/folder hierarchy (D), a button to create a new folder (E), a section of context tags (F) as well as the items contained in the context (G). A list of activities performed in the context is below the item section (bottom left) but not visible in the screenshot. The selected context is dynamically injected into the web

[4]The feature was only available for a few days towards the end of the study.

browser making its web-related items directly available there (H).

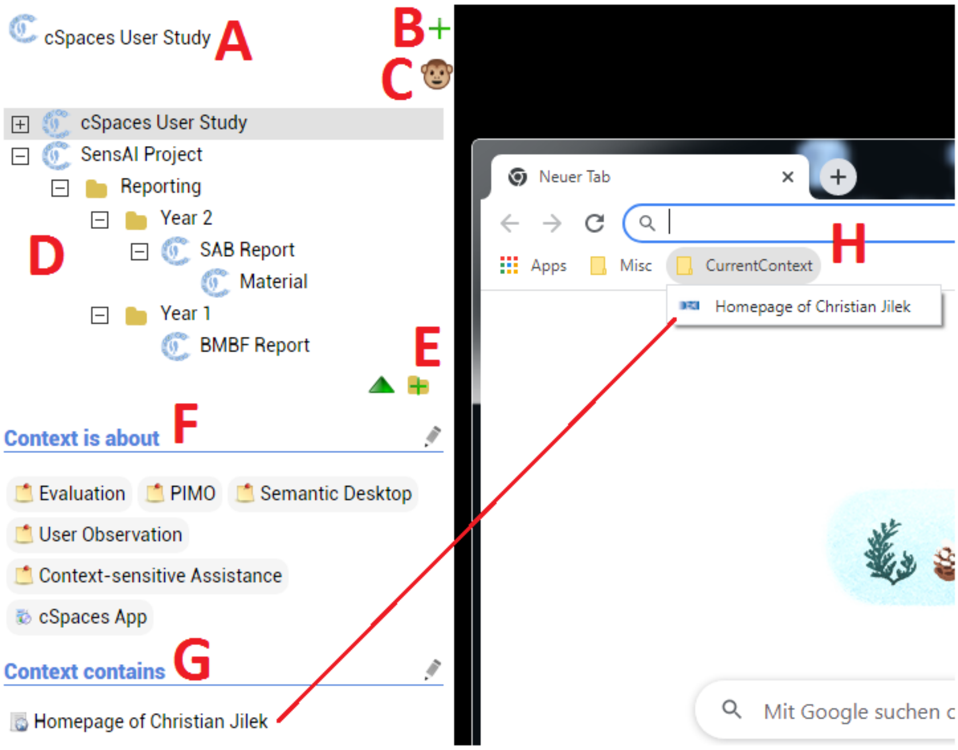


Figure 8.25: Multi-month User Study: Used cSpaces Sidebar.
Highlightings: *A: selected context (space) / B: button to create new context (space) / C: button to enable or disable observation / D: context (space) hierarchy / E: button to create new folder / F: "context is about" tags / G: items contained in selected context (space) / H: context (space) injection into the web browser*. Not visible in the screenshot (section below G): *activity history of selected context (space)*.

Remark on Results. First results of the study have already been discussed in previous sections about privacy preservation (see Fig. 6.3) and contextual notes (Fig. 8.14). As in these cases, results are sorted by topic and follow step by step in the remainder of this chapter.

Further details on the study as well as the full questionnaire document can be found in the appendix (Sec. A.4).

Results for the Idea of Context Spaces. With regard to the current section's topic, participants were asked whether they found the idea of working *with* and *in* context spaces easy to understand. All of them *quite* or *strongly agreed* (Q_{M18}, median: *strongly agree*). None of them had seen an idea like context spaces before (Q_{M19}) as shown in Figure 8.26.

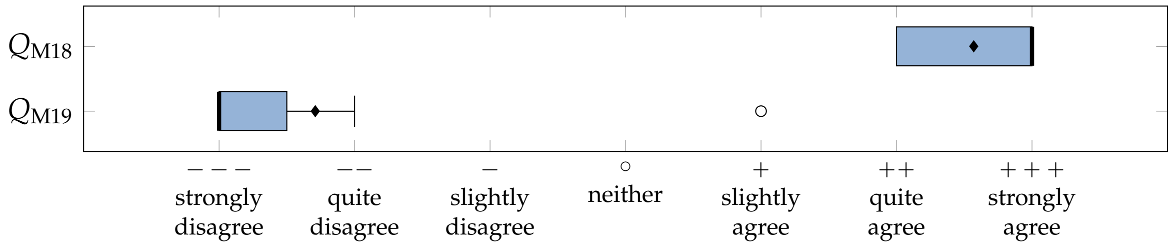


Figure 8.26: Multi-month User Study: Survey Result for the Idea of Context Spaces.

Q_{M18}: The idea of (working with and in) Context Spaces is easy to understand.
Q_{M19}: I have already seen something like Context Spaces in existing software applications.

(n=7; full questionnaire document in Sec. A.4; details on box plot semantics in Sec. B.2)

Results for Working with and in Context Spaces. With regard to benefits of working *with* and *in* context spaces, participants were asked the following:

- whether information items were right where they would like to have them with cSpaces (Q_{M20}/Q_{M21}),
- whether cSpaces were more aligned with their mental model than with traditional file, mail and web browsing apps (Q_{M22}/Q_{M23}),
- whether cSpaces allowed for faster context switches than with traditional file, mail and web browsing apps (Q_{M24}/Q_{M25}) and
- whether cSpaces helped in keeping their information sphere more tidied up (Q_{M26}/Q_{M27}).

By explicitly or implicitly addressing keeping an overview and things more tidied up as well as enabling focus, these items address research question RQ1. All four items were asked with regard to the prototype used in the study as well as the envisioned one. Note: when asking about the envisioned prototype, features introduced in later sections of this chapter, e.g. reorganization measures, also had an influence on the answers discussed here. Figure 8.27 shows the answers to all eight items. While the results for the current prototype (blue ■) are slightly on the positive side (mean and median between *neither* and *slightly agree*), the projections for the envisioned one are very much on the positive side (red ■): Three of four items have a median of *strongly agree* and the fourth one a median of *quite agree*. Mean values of all four items are between *quite* and *strongly agree*. The author therefore concludes that the current prototype is already helpful (several participants *quite* or *strongly agreed* to multiple items) and participants see great potential in the additional features that are still in development or envisioned.

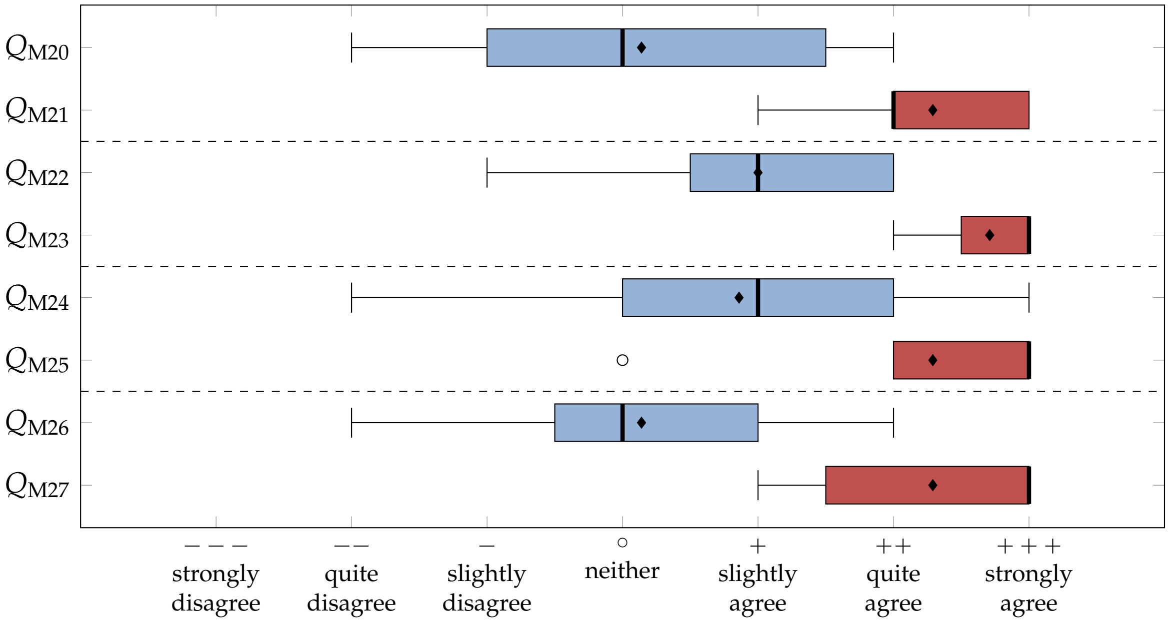


Figure 8.27: Multi-month User Study: Survey Result for Working with and in Context Spaces (current ■ vs. envisioned prototype ■).

Q_{M20}: ■ With the *current* prototype of cSpaces, information items (like files, bookmarks, etc.) are right where I would like to have them.

Q_{M21}: ■ With the *envisioned* prototype of cSpaces, information items (like files, bookmarks, etc.) would be right where I would like to have them.

Q_{M22}: ■ Context spaces as provided by the *current* prototype are more aligned with my mental model than with traditional file, mail and web browsing apps.

Q_{M23}: ■ Context spaces as provided by the *envisioned* prototype would be more aligned with my mental model than with traditional file, mail and web browsing apps.

Q_{M24}: ■ The *current* prototype of cSpaces enables faster context switches (i.e. it takes less time to resume another context) than with traditional file, mail and web browsing apps.

Q_{M25}: ■ The *envisioned* prototype of cSpaces would enable faster context switches (i.e. it takes less time to resume another context) than with traditional file, mail and web browsing apps.

Q_{M26}: ■ The *current* prototype of cSpaces helps in keeping my information sphere more tidied up.

Q_{M27}: ■ The *envisioned* prototype of cSpaces would help in keeping my information sphere more tidied up.

(n=7; full questionnaire document in Sec. A.4; details on box plot semantics in Sec. B.2)

Participants were further asked whether the mechanism of a "sashboard", i.e. the dynamic transition between sidebar and dashboard (see Sec. 3.3.3), would make their work more efficient (Q_{M5}). One participant *strongly agreed*, while all others *quite agreed* as depicted in Figure 8.28.

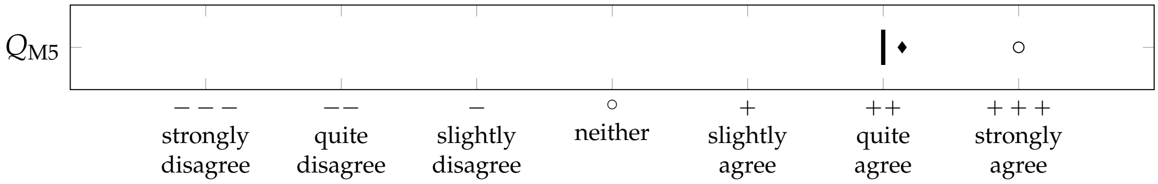


Figure 8.28: Multi-month User Study: Survey Result for cSpaces GUI Switches.

Q_{M5}: Being able to easily switch between a context's compact (sidebar) and more sprawling representation (dashboard) is a feature that would make my work more efficient.

(n=7; full questionnaire document in Sec. A.4; details on box plot semantics in Sec. B.2)

Results for Usability. There were also twelve items asking about the usability of cSpaces, which were taken from the *USE questionnaire* [232]. As shown in Figure 8.29, the assessment of

- *usefulness* (Q_{M33}–Q_{M35}, blue),
- *ease of use* (Q_{M36}–Q_{M38}, red) and
- *satisfaction* (Q_{M42}–Q_{M44}, purple)

were slightly positive: all mean values and medians are between *neither* and *quite agree*. The category of *ease of learning* (Q_{M39}–Q_{M41}, green) was perceived very well with almost only *strong agreements*.

Remark on Potential Courtesy Bias. When conducting experiments with colleagues, there is the potential problem of a *courtesy bias* [220], i.e. colleagues giving more positive answers than they would have with an unknown experimenter or the prototype of an unknown person. As the term "courtesy" suggests, this is done to be polite, please or not disappoint the interviewer, avoid negative consequences, etc. Further explanations as well as brief history of the term are given in León et al. [220]. Although the author cannot prove the absence of such a bias, the following measures were taken beforehand trying to avoid or at least reduce it:

- In an initial briefing, the author strongly encouraged participants to give objective assessments, especially underlining that constructive negative feedback often leads to improvements.

- Some items were asked with respect to the current prototype as well as the envisioned one. Participants are thus able to differentiate their criticism in a more fine-grained way.

- All participants are professional researchers knowing the bias as well as its negative consequences for validity.

- A standardized usability questionnaire (USE [232]) was incorporated into the final interview knowing that cSpaces was still a comparably fragile research prototype with several functions implemented in a less convenient way than users are accustomed to with typical desktop applications. Too high usability scores could thus be an indicator for the existence of a courtesy bias in the data.

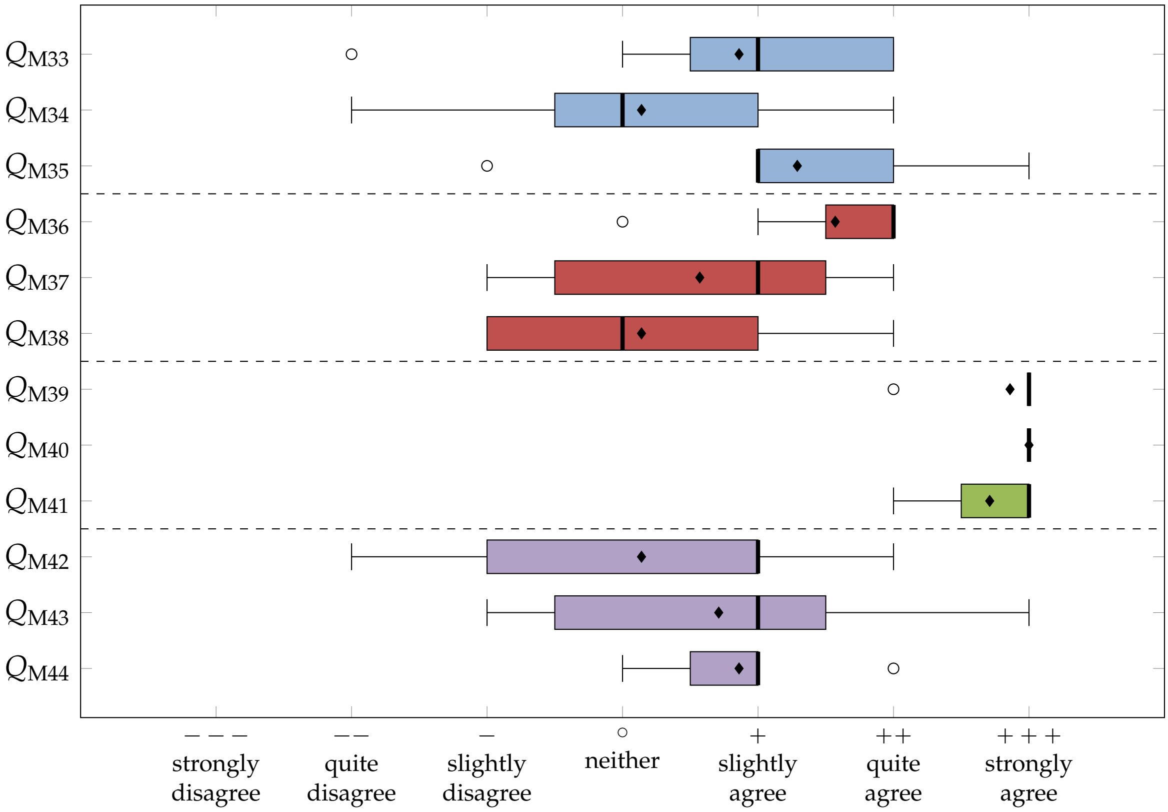


Figure 8.29: Multi-month User Study: Survey Result for Usability of cSpaces.
Q_{M33}–Q_{M35}: Usefulness,
Q_{M36}–Q_{M38}: Ease of Use,
Q_{M39}–Q_{M41}: Ease of Learning,
Q_{M42}–Q_{M44}: Satisfaction.

Q_{M33}: It helps me be more effective.
Q_{M34}: It helps me be more productive.
Q_{M35}: It is useful.
Q_{M36}: It is easy to use.
Q_{M37}: It is user friendly.
Q_{M38}: It requires the fewest steps possible to accomplish what I want to do with it.
Q_{M39}: I learned to use it quickly.
Q_{M40}: I easily remember how to use it.
Q_{M41}: It is easy to learn to use it.
Q_{M42}: I am satisfied with it.
Q_{M43}: I would recommend it to a friend.
Q_{M44}: It is fun to use.

(n=7; questions from USE questionnaire [232]; details on box plot semantics in Section B.2)

Although an experiment with a group of known participants is susceptible to such a bias, it is still a cost-effective approach, sometimes the only one affordable: Table 8.1 shows that there were 249 days of active use of the system, which would typically lead to five-digit amount of costs for "regular" (unknown) participants.

In conclusion, there are several items that also got negative scores, and several participants justified and explained these assessments in a constructive way in the free text items. With regard to the last item mentioned above, usability scores only were slightly in the positive range (except for ease of learning, which was perceived very well). The author would thus assume that if a courtesy bias exists in the data, it is marginal enough to not invalidate the author's conclusion that the cSpaces prototype used in the study was already perceived helpful and that participants saw a lot of potential in features that are still in development or envisioned.

8.3.5 Postponed cSpaces Unified Browser Study

The author has set up a last study regarding working *with* and *in* context spaces, which has been postponed last not least due to the pandemic situation existing in the last phase of this PhD project. It is a research prototype enabling users to browse files, mails and bookmarks together in a single browser with a single unified context/folder hierarchy. Alternatively (control group), all items can also be browsed with the traditional three systems: file system, mail client and web browser.

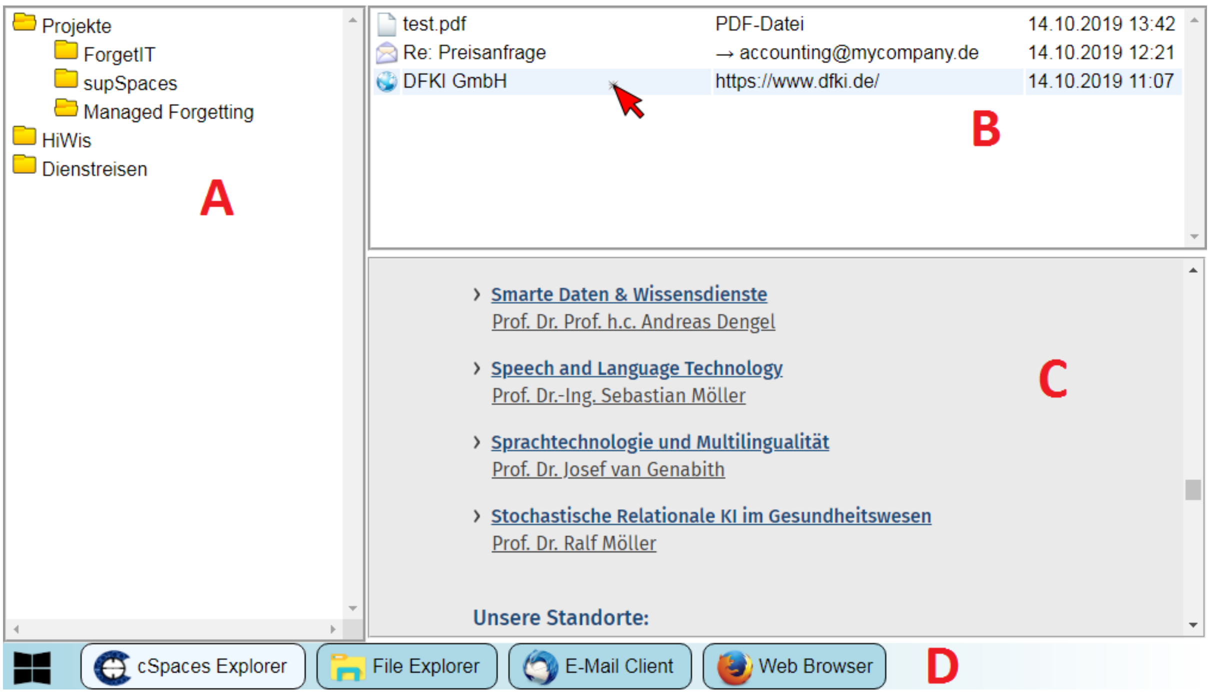


Figure 8.30: cSpaces Unified Browser: Experimentation Prototype
Highlightings: *A: context/folder hierarchy / B: information items contained in selected context or folder / C: view of selected information item's content / D: application tabs (group 1 only has the* cSpaces *explorer with a single context/folder hierarchy, group 2 only has the other three and may switch between them, which updates the folder hierarchy accordingly).*

The app's front-end is developed in *HTML5* [275] and depicted in Figure 8.30. Its GUI is meant to resemble the typical look and feel of an email client with a folder/context hierarchy on the left (A), the content of a clicked folder in the upper portion of the main part (B) and each clicked item's content in the lower portion of the main part (C). To ensure both approaches (context browser vs. traditional systems) are comparable, all interaction (including the opening the items) is done in the same application. For example, if a user is currently in the file system and clicks on the email client tab (D), the file folder hierarchy (A), currently still filled with the file folders, is filled with the email folders (widgets B and C are cleared due to the switch). Clicking on any of these folders shows its content in the item widget (B). While the three traditional systems only display their specific items (files, mails or bookmarks), all types are available together in the cSpaces browser as shown in the screenshot: *test.pdf* (a file), *Re: Preisanfrage* (an email) and *DFKI GmbH* (a website). Filtering to only view one type at a time is possible (the filtering icons are missing in the screenshot).

Note: For the sake of explaining the app, all four application tabs are depicted in the screenshot (D). Actually, there would either be just the context browser (called *cSpaces Explorer* in the screenshot) (test group) or the three traditional systems (control group).

Users having a unified browser are presumably faster in performing their tasks since they can directly navigate their different contexts. If they wish to only view the emails of a specific context, they may simply click the email filter instead of switching to their email client and navigating to the respective folder first.

Goals of the investigation would be:

1. Check if the hypothesis holds that users with a single context browser are faster than ones with traditional systems.

2. If that is the case: quantify how much faster these users are compared to the others.

3. Also gather qualitative feedback whether users would accept such a system after decades of non-existence of such a solution.

Normally, each item would open up in its respective application. A shared in-app content window (C) is used for the sake of control and detailed logging of the experiment. It also allows for easier conducting the experiment remotely (i.e. participants via the internet). Some back-end refactoring is additionally necessary to enable the remote execution, for which the experiment was not planned initially (see postponement statements above).

A major challenge in the experiment's design is setting up tasks that allow for a fair comparison of both approaches. The vacation replacement of a colleague could be a suitable scenario. Participants would need to orientate themselves in three folder hierarchies (vs. one in the context browser) to find all necessary items in order to answer questions coming from a customer.

8.4 User Support by Means of Condensation and Summarization

The next three sections present various support measures sorted by their type: condensation and summarization (the current Section 8.4), temporal reorganization, fading out and resurfacing (Sec. 8.5) and permanent reorganization (Sec. 8.6). The first ones in each category were typically realized with CoMem (cSpaces not yet existing at the time).

8.4.1 PIMO Diary – Condensation Using Retrospectively Mined Contexts

Initial Version. A first version of *PIMO Diary* was developed as part of the author's diploma thesis [178]. It as a CoMem-based app that generates diaries from a user's PIMO. Basically, diary generation is a variant of retrospective context mining. But in contrast to the *Contextifier* (see Sec. 5.3.3), PIMO Diary works one level of abstraction higher, taking the user's PIMO as an input instead of the information items themselves. On the one hand, this eases the task since a lot of relations are explicitly stated already. On the other hand, there are additional challenges: Typically a user states a time interval the diary should be generator for as well as the number of desired diary entries. PIMO Diary then starts by calculating the similarities of all things in the given time interval. They are iterated from highest to lowest similarity checking whether items should be merged. The most similar things are the first ones to become a cluster/context. Depending on their type, some things are more likely to become the core of such a cluster, e.g. events are typically good candidates. Except for situations, in which several events belong to the same project. Then it is rather the project, which becomes the core of the context. The similarity metric takes text and concepts into account. Regarding the former, labels/titles are treated with a higher weight than body/content text since they are typically a kind of human-made summary. PIMO Diary was inspired by the idea of *memory landmarks* [288, 172], i.e. "special landmarks or anchor events for guiding recall and for remembering relationships among events" [172]. Another idea is that singular events are either irrelevant or very special because of the rarity, e.g. meeting the president of a state or a music or film star, etc. In contrast, meeting the representative of a company, who does not contact you again after an initial meeting, makes the event a rather unimportant but nevertheless rare one.

As mentioned before, PIMO Diary iterates over item pairs sorted by similarity. Each of the two items of such a pair could meanwhile already be part of their own cluster. And as clusters (mined contexts) become larger, further heuristics are therefore applied. For example, centroid distances are additionally taken into account. If they are too large, no merge will be performed. Instead, the next pair is checked for a potential merge. Step by step, the system converges towards the target number of entries but possibly not able to exactly hit that number. If, in the end, too many contexts are available, the system votes, which ones to display and which ones to drop. This depends on a context's impact, i.e. type of core elements, number of contained items, etc. Finally, the desired number of diary entries is returned together with a ratio of how much of the user's PIMO in that particular time interval is covered by these entries. Low ratios are a hint for the user to increase the number of desired entries to cover more items of that time.

Evaluation. PIMO Diary was evaluated with four participants that had used CoMem for about four months and used the diary for another three weeks. Qualitative feedback was gathered, which had been very positive.

Improved Version. One of the first tasks of this PhD project was to enhance PIMO Diary with regard to some of the underlying algorithms (e.g. headline generation) and speed optimizations. The latter was particularly important since the app was shown at the German technology fair CeBIT[5] in 2015 [239] as part of the ForgetIT project [120]. A paper presenting the improved version of PIMO Diary was published in Jilek et al. [184]. Its GUI is depicted in Figure 8.31. Each diary entry has a header stating its covered time period, a headline as well as its most important items and an image (if available) (A). Further items can be accessed by unfolding the entry. The entries do not have generated sentences but rather a list of the most important keywords (B). The most relevant concepts are shown in each entry's upper right corner (C). On the right, there are basic and detail settings to adjust the diary generation (D). Last, there is an overall (concept) context for the whole time period (E).

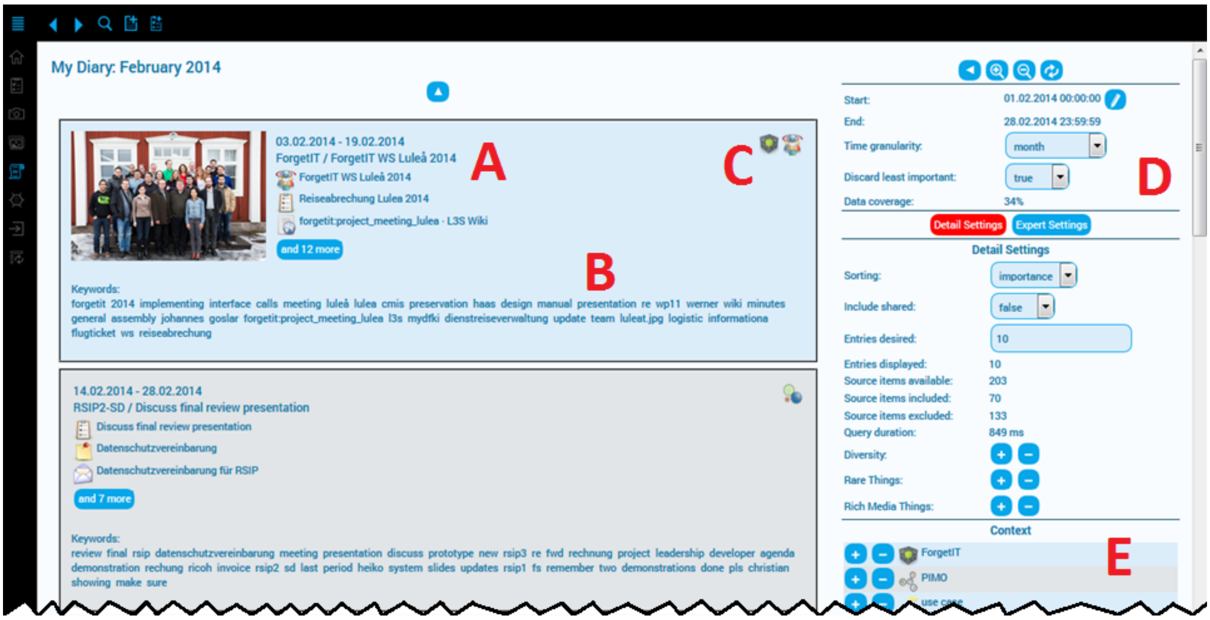


Figure 8.31: PIMO Diary: GUI.

Highlighted sections: *A: a diary entry's time period, generated headline, important items [other items available by unfolding] and image [if available] / B: most important keywords of the entry / C: most important tags of the entry / D: basic and detail settings / E: overall context of the selected time period.* (Note: A variant of this screenshot has already been published in Maus et al. [242].)

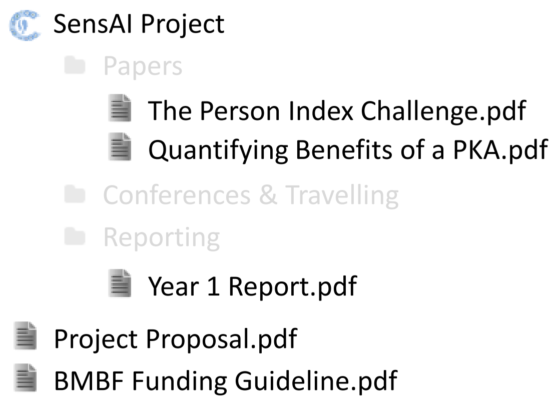
Further Use for Other Support Measures. PIMO Diary's clustering capabilities can be used by other support measures, e.g. the reorganization measure of context splitting (see Sec. 8.6.2). Another idea for condensation – which is only triggered if items have lost relevance – was to remove all items covered by a diary entry from the KG and instead keep the entry as a condensed "memory landmark". Thus, a short summary of a project, event, etc. is still available – with keywords, item names, possibly a photo – but the original items are (currently or permanently) not available. If the contained items regain relevance (and are still available, e.g. in an archive or forgotten area of the system), this condensation step could be reverted.

Parts of PIMO Diary have also been used to provide basic functionality with regard to indexing, vector representations and helper functions that ease similarity calculations, etc. in the Contextifier (Sec. 5.3.3) and the Stream Context Miner (Sec. 6.3.2).

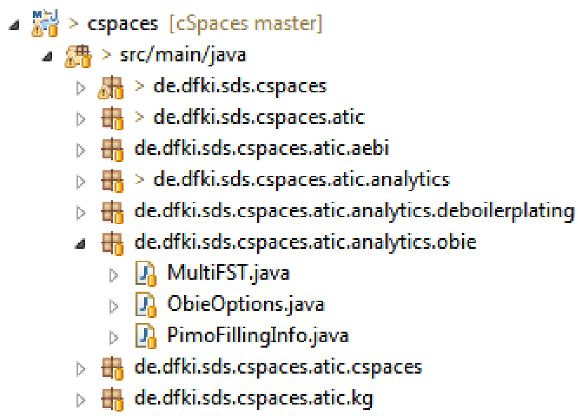
[5] https://www.cebit.de/

8.4.2 Condensation by Flat Context Views

A second condensation measure is still in a pre-prototypical state. It is called *Flat Context Views* and was inspired by *flat package views* available in software development environments like the *Eclipse IDE* [Sw38]. As depicted in Figure 8.32b, the content of the *OBIE* package, containing the *Multi-FST* (introduced in Sec. 6.2.2) and other classes, is shown at the same level as the base package (`de/dfki/sds/cspaces`) despite actually being three levels deeper in the hierarchy (`/atic/analytics/obie`).



(a) Flat Context View.



(b) Flat Package View.

Figure 8.32: Flat Context View Inspired by Flat Package View (as for example available in the *Eclipse IDE* [Sw38]).

This idea could be applied to cSpaces as well: Imagine browsing the *SensAI Project* context as depicted in the Figure 8.32a. Having a flat view would not only show the direct content of that context (here: proposal and funding guideline) but also a most relevant selection (high MB) of each sub-context (here: two papers and the year 1 report). Only if items in child contexts are above a certain MB threshold they would make it into the flat view. In this example, no item of the *Conferences & Travelling* context had a high enough buoyancy. When returning to the *SensAI Project* context after some time of absence, a user is likely able to see the currently most relevant items right away. Borders between parent and child contexts are temporarily dissolved as illustrated in the example by sub-contexts being drawn and labeled in gray. This Flat Context View is meant as a complementary condensation feature: it is possibly enabled by default when returning to a context but can be disabled if a user browses deeper or explicitly turns it off. The amount of contained items of a context could be used as a threshold to turn off the feature by default: rather empty contexts will less likely cognitively overwhelm the user when returning to them.

Survey Result. In the multi-month user study, participants were asked whether Flat Context Views would help them being less overwhelmed when resuming a (non-sparsely filled) context. Most participants *slightly* or *quite agreed* (Q_{M8}, median: *quite agree*) as shown in Figure 8.33.

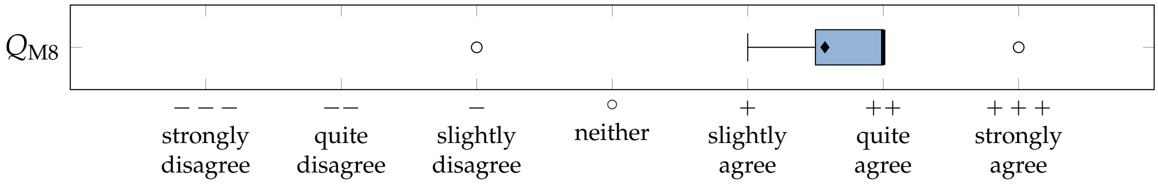


Figure 8.33: Multi-month User Study: Survey Result for Flat Context Views.

Q_{M8}: Flat Context Views are a feature that would help me being less overwhelmed when resuming a (non-sparsely filled) context.

(n=7; full questionnaire document in Sec. A.4; details on box plot semantics in Sec. B.2)

8.5 User Support by Means of Temporal Reorganization, Fading Out and Resurfacing

A second category of support measures are temporal reorganization as well as mechanisms of fading out and resurfacing (when losing or regaining relevance, respectively) triggered by Memory Buoyancy (MB).

8.5.1 Fading Out and Resurfacing in CoMem

First fading out and resurfacing mechanisms were realized as part of CoMem. They are guided by and fully rely on changing MB values, whose conceptual and technical details were described in Section 7.2.1. As mentioned in that section, the development was led and mainly conducted by Sven Schwarz with only minor contributions by the author.

Figure 8.34 shows the example of the calendar entry of a workshop conducted in the ForgetIT project [120]. Viewing the entry during the meeting is depicted on the left-hand side. All involved persons (top), tasks (middle) and emails (bottom) associated with the event are shown. After eight months (middle part of the figure), tasks and emails have already dropped in relevance and are thus not shown anymore. What remains are primarily persons involved in the project as well as the presentation slides and the project (thing) itself. After two years (right-hand side), the presentation is not visible anymore (possibly since it was just one presentation among many in the project), whereas photos taken at the event (probably often referred to and shown to colleagues) as well as the project itself remain. The same is true for several of the involved persons.

A second example especially also showing MB rising is illustrated in Figure 8.35. It shows the MB graph of a task (labeled *WP9*, in blue) and two subtasks, *T9.4* (in green) and *T9.5* (in pink). The old task T9.4 is revisited, thus rising in relevance (see highlighting A), when writing a deliverable in task T9.5 (section between A and B) and during the final project review in April (D). Phases of fading out can also be seen, for example the sections after A and C for the green graph and after B and between C and D for the other two. They are reminiscent of the *forgetting curve* by Ebbinghaus [100], which – like the MB calculation – is typically also described as a reciprocal function ($\frac{1}{x}$ or $\frac{1}{log(x)}$).

These fading out and resurfacing mechanisms have also been used by *PimoCloud* [314] in CoMem to automatically upload and download items and thus synchronize users' local devices accordingly, i.e. upload and deletion of local instances on relevance drop and (re-)download when (re-)gaining relevance.

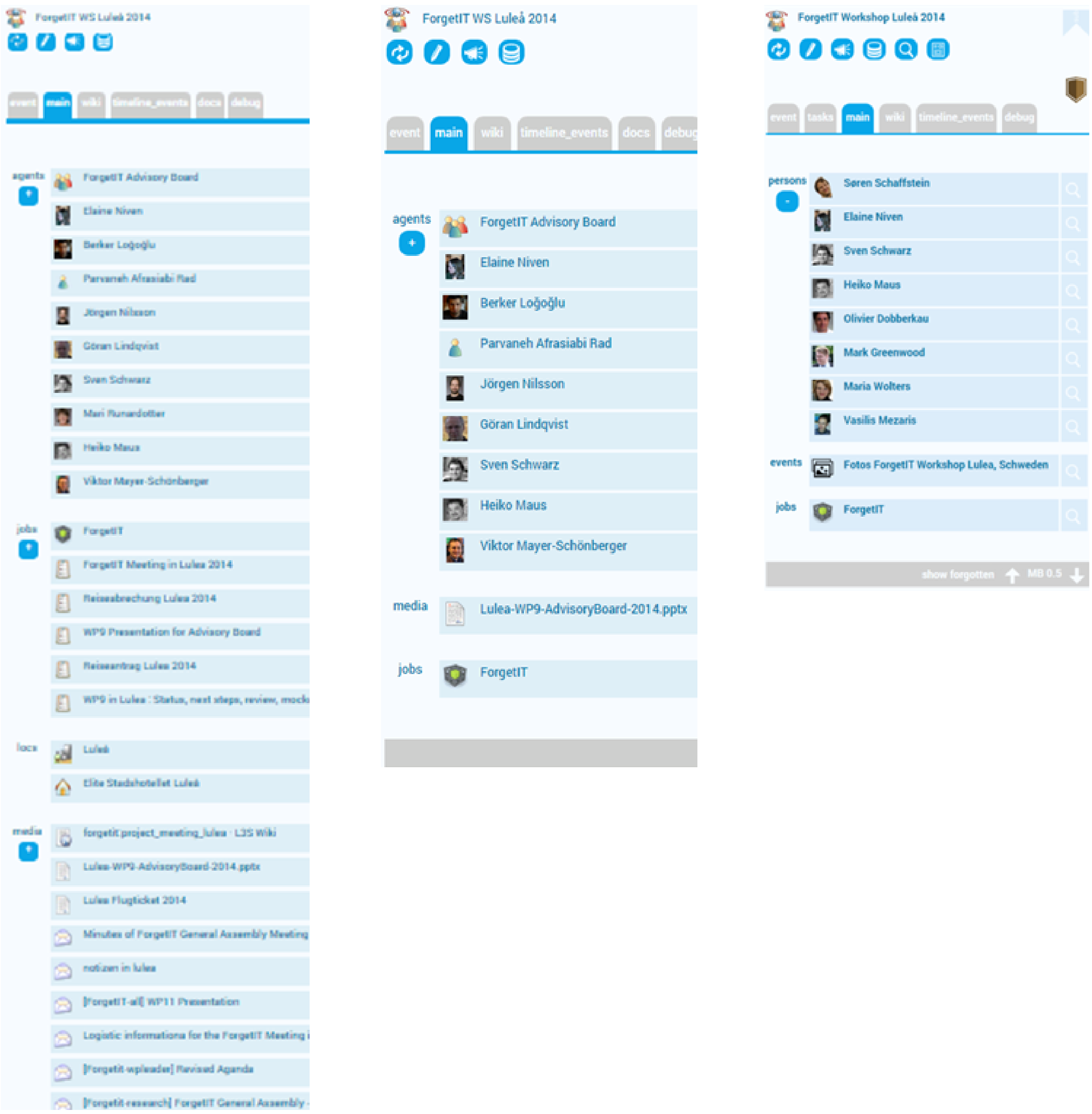


Figure 8.34: Fading Out and Resurfacing in CoMem: Calendar Example.
The example shows the calendar entry of a workshop during the event (left), after eight months (middle) and after two years (right). First tasks and emails fade out (first to second image), later also the workshop's presentation and only some of the persons (contacts) and photos remain (second to third image).
(Note: These are screenshots by Heiko Maus already published multiple times [182, 185, 240, 244].)

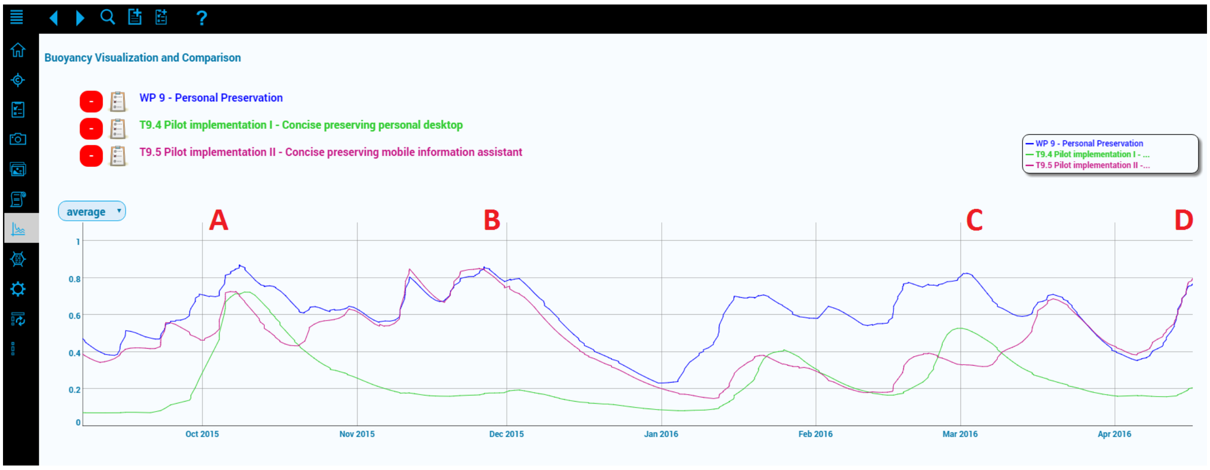


Figure 8.35: Fading Out and Resurfacing in CoMem: Memory Buoyancy Visualization. The green task for example resurfaces at positions A and C and fades out again shortly after each of these positions. The other two tasks are mainly active between A and B, around C and from position D on. Their fading out sections are after B and between C and D. (Note: This is a screenshot by Heiko Maus already published multiple times [182, 240, 244].)

8.5.2 cSpaces Context Overview

Also based on the idea of fading out and resurfacing is a *Context Overview* for cSpaces. When users resume their work after a break or at the beginning of a new work day, for example, this feature provides them with an overview of those contexts with the highest MB. The overview can therefore also be seen as a welcome screen. It consists of a set of tiles, each of them showing a context summary consisting of tags, high buoyancy items, last activities, etc. A context's *Last Focus* (introduced in the next section), could also be shown in such tiles. Users can set the default amount of such tiles according to their liking, whereas further contexts can be loaded using a *more* button. Clicking a context will bring up the sidebar with the selected context already loaded. A first prototypical implementation, supported by Desiree Heim as part of her student assistant job at the time, is depicted in Figure 8.36. The screenshot was cropped to only show three tiles for the sake of readability – a typical number of tiles would be nine, twelve or more (depending on screen size and user preferences).

Survey Result. In the multi-month user study, participants were asked whether such an overview would be a useful feature to get or keep an overview of currently relevant contexts. One of them *quite agreed*, while all others *strongly agreed* (Q_{M4}).

Figure 8.36: Context Overview.
This screenshot was cropped to only show three tiles for the sake of readability.
(Note: This is a screenshot provided by Desiree Heim.)

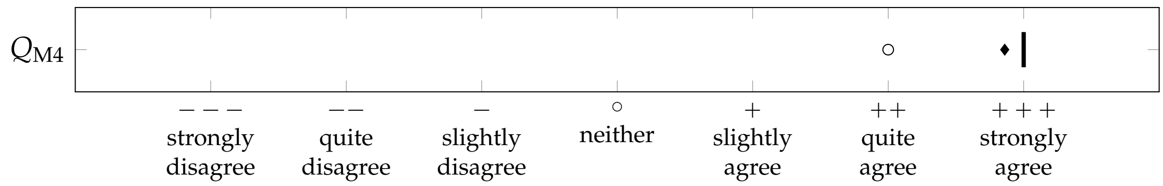


Figure 8.37: Multi-month User Study: Survey Result for Context Overview.

Q_{M4}: Context overview would be a useful feature to get/keep an overview of currently relevant contexts.

(n=7; full questionnaire document in Sec. A.4; details on box plot semantics in Sec. B.2)

8.5.3 cSpaces Saving Last Focus

Last Focus was already mentioned in the last section and in the introduction of cSpaces' context model (see *focal aspect* in Sec. 5.1). The basic idea is that a context presents its "golden thread" to the user upon revisitation.

In cSpaces, MB is the primary driver for ranking, hiding, reorganizing, etc. But when returning to a context (resumption), it may nevertheless be helpful to show the user just those resources and activities that were last used or performed, respectively. Such a Last Focus view would be more driven by *least recently used (LRU)* timestamps than by MB values and would serve as a complement. In future versions, the scope of Last Focus could be enhanced to cover more aspects like open/active applications, the last persons a user was in contact with in the respective context, etc.

In a more business-related scenario like the one of *supSpaces* [370] (see Sec. 9.2.1), clerks were working on solving technical problems (IT support scenario). If they could not solve a certain problem until their shift ended, they had to write a respective note in the shift log, so clerks of the next shift could take over where they left off. Last Focus would also be useful in such scenarios possibly replacing (or at least semi-automating) such a shift log.

Survey Result. In the multi-month user study, participants were first asked whether they understood the difference between LRU and MB, which they mostly did – all *strongly agreed* except for one (Q_{M6}). Then they were asked whether looking at the Last Focus would help them remember more easily where they left off when returning to a context. Most participants *quite* or *strongly agreed* (Q_{M7}, median: *quite agree*, mean close to *quite agree*). The results are depicted in Figure 8.38. Last Focus is still in pre-prototypical state.

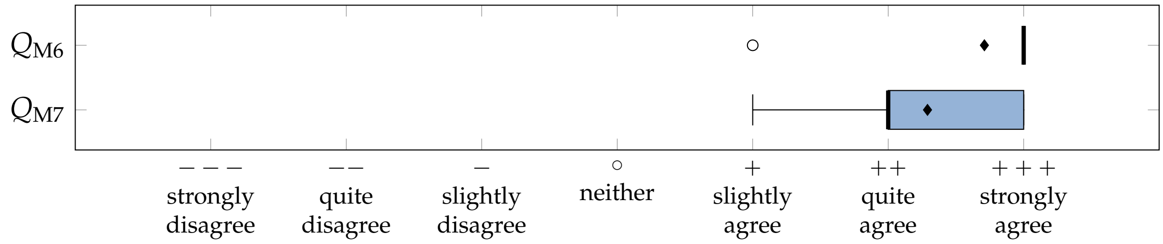


Figure 8.38: Multi-month User Study: Survey Result for Last Focus.

Q_{M6}: I see a difference between Memory Buoyancy and Least Recent Used values.
Q_{M7}: Looking at the last focus would help me remember more easily where I left off when returning to a context.

(*n*=7; full questionnaire document in Sec. A.4; details on box plot semantics in Sec. B.2)

8.5.4 cSpaces Temporal Hiding and Reorganization

The last support measure in this category is temporal hiding and reorganization. Note that the border between temporal and permanent hiding/reorganization is rather fluent. "Temporal" here primarily means driven by Memory Buoyancy, i.e. things fading out and resurfacing. However, if things are not stimulated again and thus MB does not rise again, faded and permanently removed seem identical, what differs then is actually just the intention. Things would resurface if MB rose again. In contrast, measures like context splitting or merging, as discussed in Sections 3.3.1 and 8.6.2, are right away meant to be permanently (except for cases, in which a user calls an undo function).

A very simple form of temporal reorganization is depicted in Figure 8.39. Context injections, as explained in detail in Section 8.2.2, make the current context available as a *CurrentContext* folder in the file system (center) and web browser (right). Switching contexts (via the sidebar, left-hand side) dynamically updates the injections and thus also the content of these folders. The figure shows a context switch from the *Managed Forgetting* to the *supSpaces* context. Six files and two bookmarks of the former project are replaced by two files and a bookmark of the latter.

As mentioned, the above example is just a simple form of reorganization. A more typical case is depicted in Figure 8.41. The upper half shows the real example of a project folder on the author's computer. When revisiting this folder after two years of disuse, the system could present only a condensed version as depicted in the lower half of the figure. Instead of being overwhelmed with all the available items, only those files and folders considered most relevant by the system (highest MB) are shown. However, all other items are still available but hidden by default using a generated *forgotten* folder. For the sake of the example's simplicity, only a single folder was used inducing a hard cut between forgotten and still active. However, the “forgotten area” could be clustered in a more fine-grained way, e.g. with folders like *created X months ago*, *associated with hot topics* (i.e. those items’ local MB is rather low but their global one is high), etc. Examples are shown in the Figure 8.40.

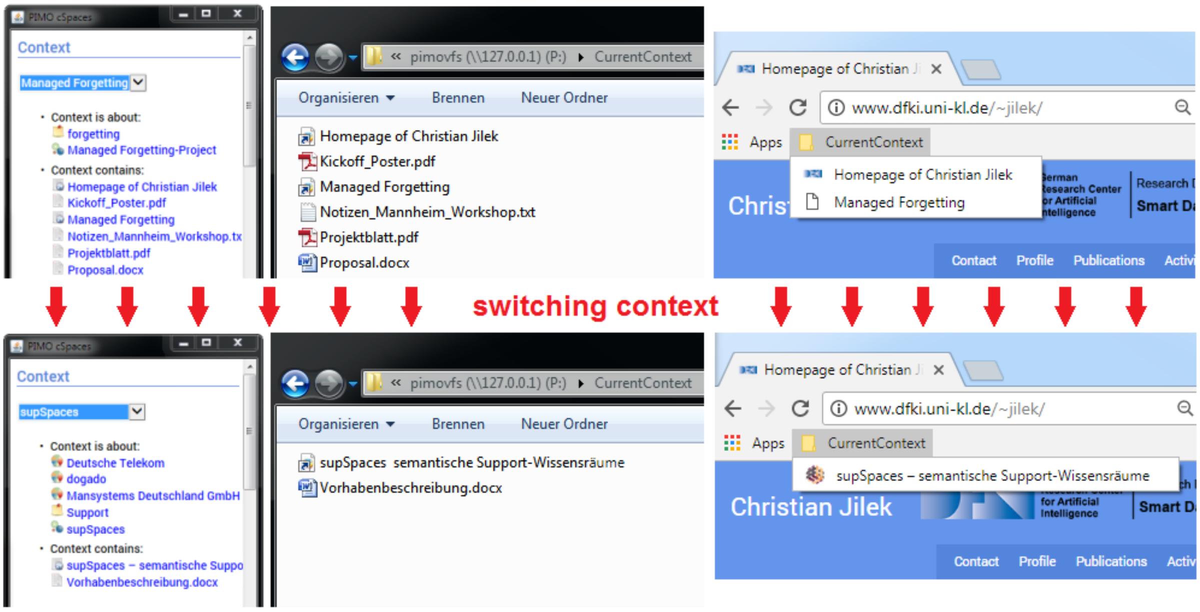


Figure 8.39: Reorganization: Basic Example of Temporal Reorganization.
Switching contexts updates the sidebar’s content (left) according to the newly selected one as well as the injections into the file system (center) and web browser (right). Upper screenshot: apps before context switch, lower screenshot: apps after context switch.
(Note: This figure has already been published in Jilek et al. [187].)

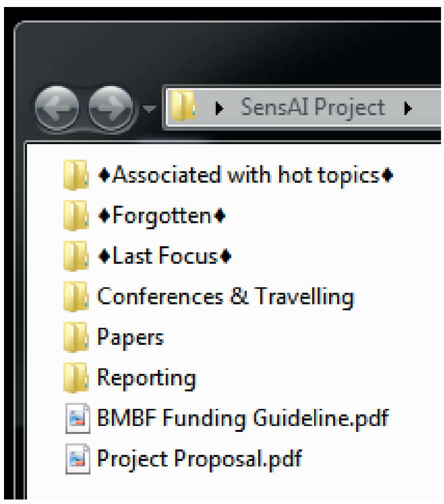


Figure 8.40: Reorganization: Intelligent Folder Injections.
In this example, they are set apart from regular folders using a diamond symbol as their first and last character (configurable).

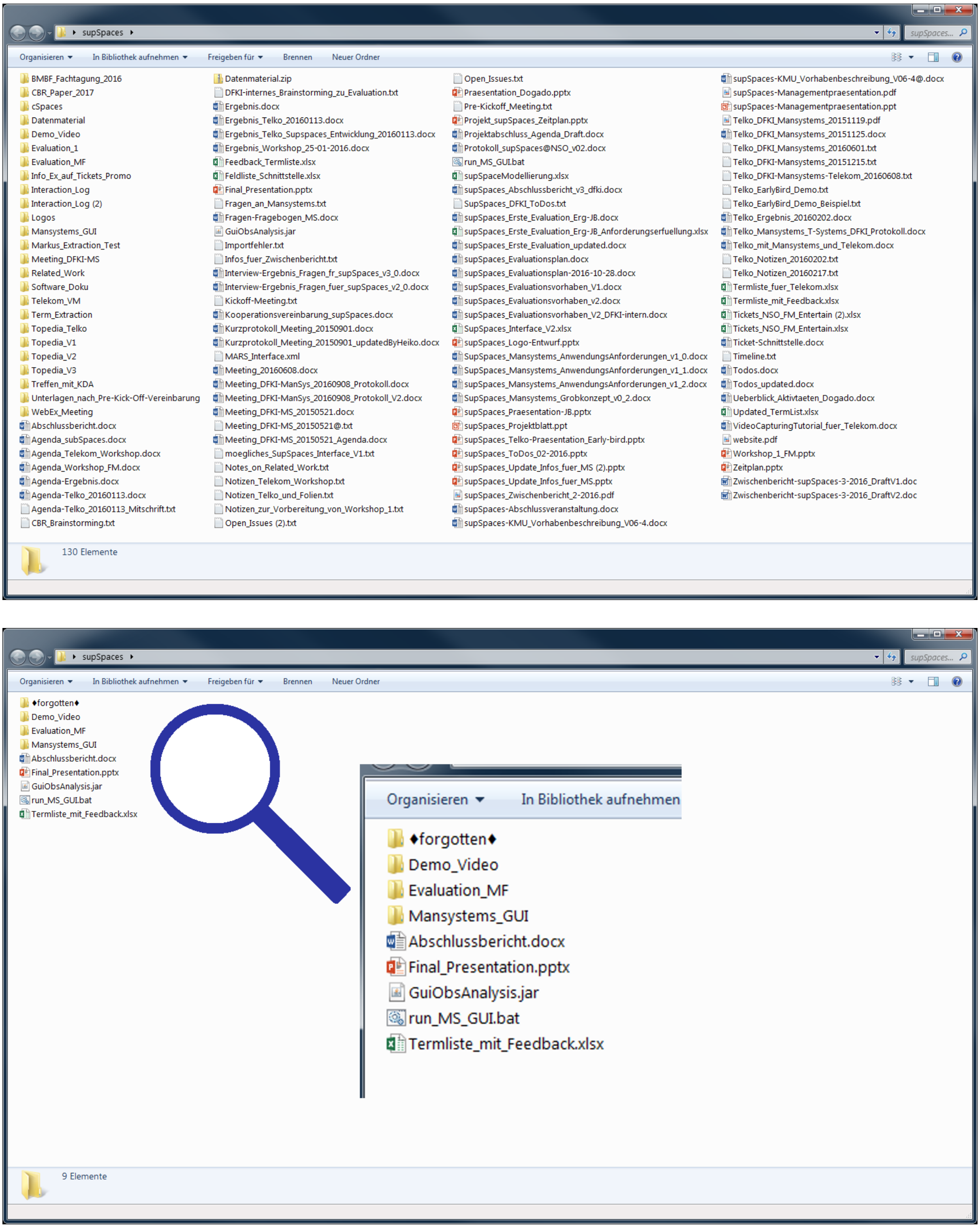


Figure 8.41: Reorganization: Example of Reorganization.
Revisiting a project folder after two years of disuse. Upper screenshot: regular system, lower screenshot: cSpaces hiding the least relevant items by default but also providing possibilities to re-access them using a “forgotten” folder (IFI).

Since such folders are realized using virtual file system technology, injecting them can be performed in milliseconds without actually moving or copying any data on the harddisk. Using such **Intelligent Folder Injections (IFIs)** – as the author called them –, a context could present its "golden thread" to the user, especially its *Last Focus* as mentioned in the previous section. The labels of IFIs deliberately start and end with a presumably rarely used character (like the black diamond symbol) to set them apart from regular folders. As mentioned, IFIs are virtual subdivisions of contexts according to certain criteria. They are primarily meant as a simple means to access the results of temporal or permanent hiding, condensation or reorganization measures performed by the system.

Upon revisitation, cSpaces needs to assemble the context's current appearance – or "in-carnation", in other words. The respective component is therefore called the **Incarnator**. Dependent on user settings (e.g. only show high buoyancy items by default, hide the rest), the folder's actual content as well as all relevant IFIs are calculated (or at least injected): for performance reasons, a combination of pre- and lazy calculation is intended in future versions.

cSpaces reorganization features are in an early state. Readers are kindly referred to a demonstration video available online[6] to get an impression of the system's dynamics. In the video, simpler IFIs just using creation and modifications dates for clustering are used.

Survey Results. Participants of the multi-month user study were asked about IFIs. The first item (Q_{M9}) asked whether they found IFIs easy to understand. Whether they are an easy way to work with a highly autonomous assistant like cSpaces as well as the results of its support decisions (like hiding, condensation or reorganization) was the topic of the second item (Q_{M10}). Most participants *quite* or *strongly agreed* to both items as depicted in Figure 8.42.

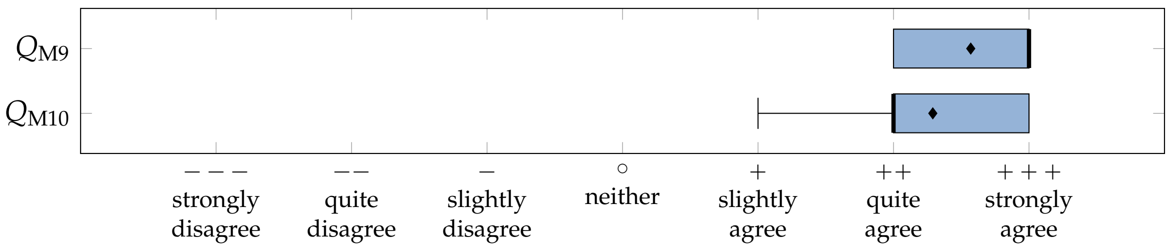


Figure 8.42: Multi-month User Study: Survey Result for Intelligent Folder Injections.

Q_{M9}: Intelligent Folder Injections are easy to understand.
Q_{M10}: Intelligent Folder Injections are an easy way to work with a highly autonomous assistant like cSpaces and the results of its support decisions like hiding, condensation or reorganization.

(n=7; full questionnaire document in Sec. A.4; details on box plot semantics in Sec. B.2)

Three further items were about the cSpaces approach in general: The first one (Q_{M30}) asked whether using cSpaces to only view the current context while hiding all others reduces cognitive load while working. The second one demanded participants to assess whether

[6]possible locations: https://www.dfki.uni-kl.de/~jilek/demo/cspaces/#video or https://arxiv.org/src/1805.02181v1/anc/demo_video.mp4

cSpaces are a suitable user interface to work with a highly autonomous assistant that is capable of self-(re)organization and forgetting (Q_{M31}). The item Q_{M32} additionally asked whether they could think of a better solution to subtly provide contextual meta-information to the assistant. Answers to the first item were all positive with a median and mean of or close to *quite agree*. Participants either *quite* or *strongly agreed* to the second item, while the third one was answered rather negatively. Results are shown in Figure 8.43.

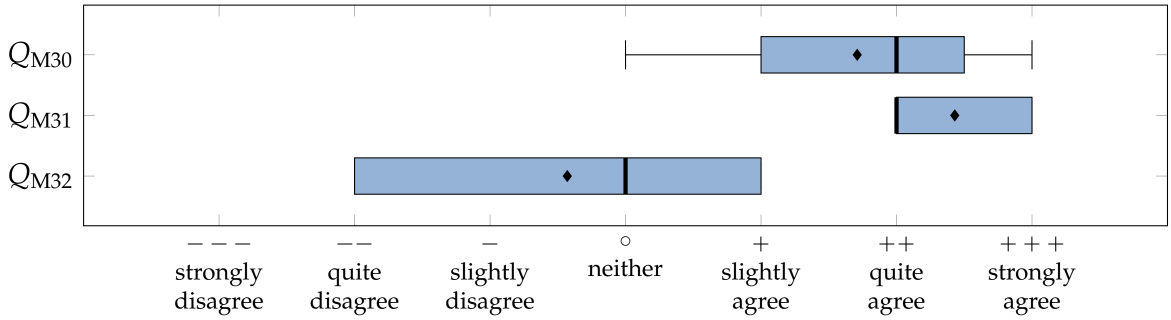


Figure 8.43: Multi-month User Study: Survey Result for Self-organizing Context Spaces.

Q_{M30}: Using Context Spaces *in general* (not necessarily the current prototypical implementation) to only view the current context (and its content) while hiding all others (and their content) reduces my cognitive load while working.
Q_{M31}: Using Context Spaces *in general* (not necessarily the current prototypical implementation) are a suitable user interface to work with a highly autonomous assistant that is capable of self-(re)organization and forgetting.
Q_{M32}: I could think of a better solution to subtly provide contextual meta-information to the assistant.

(n=7; full questionnaire document in Sec. A.4; details on box plot semantics in Sec. B.2)

8.6 User Support by Means of Permanent Reorganization

The last category of support measures discussed are by means of permanent reorganization (also see remarks on temporal vs. permanent at the beginning of Section 8.5.4).

8.6.1 Preservation Prototype II: General Preservation in CoMem

Note: Section 8.6.1 is in large part a revised version of a report by the author published as part of Zhu et al. [407, Sec. 6.2].

This section is a continuation of Section 7.3.1, which presented the *Preservation Prototype I* that aimed at photo preservation. The second prototype discussed in the following takes more types of information items into account and thus also uses a slightly different calculation (more parameters, different weights, etc.) The next paragraph first presents more details on the Preservation Value calculation basically valid for both prototypes.

Details on Preservation Value Calculation. The six Preservation Value dimensions, investment, gravity, social graph, popularity, coverage and quality, are used in both prototypes. As mentioned, the Preservation Value is a weighted sum of aspects of these dimensions. The weighting basically works as follows:

- Suppose a resource only has low evidence values for each dimension, then the sum of these evidences should also be quite low.
- An exceptionally high value in one dimension, e.g. a user spent a lot of investment on a resource, should definitely lead to a high Preservation Value, no matter how high or low the other evidences are – one extraordinary high value can "pull up" the overall score.
- If a resource only got scores that are slightly above average but this is true for most of the dimensions, then its Preservation Value should be relatively high – the individual scores should sum up ("escalate") to a rather high value.
- Resources having a combination of mostly low and some average values should only have a low to average Preservation Value.

After gathering a value for every factor, which also includes a normalization step (details below), they are added according to their belonging to one of the six dimensions. To combine evidence factors in a way that fits the requirements described before, an approach by Schwarz [343] was utilized, which is based on the *Dempster-Shafer Theory of Evidence* [(343) referring to (138, 403)] and has previously been used in similar scenarios. According to this approach, two evidence scores v and w ($v, w \in [0,1]$) are added as follows:

$$v \oplus w := 1 - (1 - v) \cdot (1 - w)$$

As stated in Schwarz [343], the approach applies basic probability theory: A higher-level evidence relies on lower-level ones. Its probability P(A) is therefore calculated as the probability of the union of the lower-level evidences $P(E_i)$, which are assumed to be independent (*):

$$\begin{aligned}
P(A) &= P(E_1 \cup E_2) \\
&= P(E_1) + P(E_2) - P(E_1 \cap E_2) \\
&\overset{*}{=} P(E_1) + P(E_2) - P(E_1) \cdot P(E_2) \\
&= 1 - 1 + P(E_1) + P(E_2) - P(E_1) \cdot P(E_2) \\
&= 1 - 1 + P(E_2) + P(E_1) - P(E_1) \cdot P(E_2) \\
&= 1 - (1 - P(E_2) - P(E_1) + P(E_1) \cdot P(E_2)) \\
&= 1 - (1 - P(E_1)) \cdot (1 - P(E_2))
\end{aligned}$$

Each evidence factor evaluation should lead to a value in $[0, 1]$. Some items already have values in this range, e.g. photo ratings. To others, e.g. the length of a wiki text, normalization functions like the following are applied:

$$n_1(x) := \begin{cases} 1 - \frac{1}{ln(x+e)} & \text{if } x \geqslant 0 \\ 0 & \text{otherwise} \end{cases}$$

To additionally cope with varying user behavior (generally short vs. long annotations, few vs. most pictures highly rated, etc.), the resulting values are further normalized in a subsequent step. All values in a certain category, e.g. the wiki text length, are divided by their maximum value in this category (in this case, the value calculated for the longest text). According to the Preservation Value dimensions explained earlier, the different factors are then added in order to get combined evidences for investment (e_I), gravity (e_G), social graph (e_S), popularity (e_P), and quality (e_Q). The calculation of e_I is considered as an example in the following: Given the normalized values for the number of annotations (n_A), the wiki text length (n_T), the number of writing actions (n_W) and the number of rating actions (n_R) as well as their respective weights w_A, w_T, w_W and w_R, the combined evidences for investment e_I are defined as follows:

$$e_I(n_A, n_T, n_W, n_R) := w_A \cdot n_A \oplus w_T \cdot n_T \oplus w_W \cdot n_W \oplus w_R \cdot n_R$$

The other combined evidences e_G, e_S, e_P and e_Q are calculated analogously. Next, these combined values are summed up.

$$c(e_I, e_G, e_S, e_P, e_Q) := 0.8 \cdot e_I \oplus 0.8 \cdot e_G \oplus 0.65 \cdot e_S \oplus 0.5 \cdot e_P \oplus 0.5 \cdot e_Q$$
$$f(e_I, e_G, e_S, e_P, e_Q) := 0.5 \cdot e_I \oplus 0.5 \cdot e_G \oplus 0.65 \cdot e_S \oplus 0.8 \cdot e_P \oplus 0.8 \cdot e_Q$$

$$e_I, e_G, e_S, e_P, e_Q \in [0, 1]$$

Basically, for a given resource x, the Preservation Value $PV(x)$ is equal to $c(...)$ for *curators* and to $f(...)$ for *filers* (see personas in Sec. 7.3.1). The *safe* variants of these profiles require another process step to ensure *coverage*. For photo collections, this means that all photos having the highest Preservation Value in their respective collection will receive a Preservation Value of 1, which is the highest value possible. The Preservation Value calculated in a first pass is then overwritten accordingly. In a final step, the algorithm assigns each resource to a matching preservation class ranging from *gold* (preserved) to *wood* (non-preserved) by applying respective thresholds. These classes could be used by other components developed in the ForgetIT project, details are omitted here (e.g. see Mezaris et al. [251]).

The preservation calculation details presented so far basically apply to both prototypes with the exception that Prototype I used slightly different weights to calculate the overall sum[7]. Especially, gravity was not considered since participants' PIMOs were comparably empty (thus connectivity, important projects and closeness to important things did not find application)

[7] weights in $c(...)$: $w_I = 1.0$, $w_G = 0$, $w_S = 0.5$, $w_P = 0.75$, $w_Q = 0.5$, weights in $f(...)$: $w_I = 0.5$, $w_G = 0$, $w_S = 0.5$, $w_P = 0.75$, $w_Q = 0.75$

and there was only one type of information item to consider (photos).

The next paragraph discusses how the second prototype further differs from the first one.

Preservation Prototype II. As mentioned, the second prototype takes more types of information items into account basically supporting everything stored/mentiond in CoMem. Additionally, the scope of possible system interactions (or used system features, respectively) is much higher. On the one hand, this means that there is much more data available to base preservation suggestions on. But on the other hand, fine-tuning of the algorithm is more difficult due to the increased number of parameters. First, the *cold start problem* (present in the photo study) is already solved for most users at DFKI SDS since they already used the CoMem for a longer time. Second, DFKI users may choose their preservation strategy on their own. They may use the four previously introduced strategies, such as *Safe Curator*, *File & Forget*, etc. as presets but may additionally check or uncheck individual items (heuristics or rules) in the different dimensions. It is therefore possible to set the algorithm to measure *investment* only by the number of annotations, disregarding the length of a resource's wiki text and other factors, for example.

A major difference to Prototype I is that the *gravity* dimension is evaluated with especially the *type-based heuristic* supporting various information items: Bonuses between 0.45 and 0.5 are assigned to life situations, projects and conferences, 0.2 to 0.25 to events, jobs, persons, agents, videos and notes and 0.1 to documents. Since contracts are typically a rare document in CoMem (at least at the time), they are always preserved by getting a type-based bonus of 1.

The summarization of the different factors to combined evidences is done analogously to Prototype I, except for the normalization. Prototype II's algorithm uses the following functions, whereas the constants a, b, c and d are set dependent on the currently evaluated factor[8]:

$$n_2(x) := max(\ a \cdot [1 - (x+c)^{-d}],\ 1\ ) \qquad (0 \leqslant d \leqslant 1)$$

$$n_3(x) := max(\ a \cdot [1 - \frac{1}{log_b(x+c)}],\ 1\ )$$

The preservation status of all things in the PIMO can be viewed using a similar widget to the *Time Capsule GUI* used in the photo study (see Fig. 7.7). It basically also is a vertical split screen and depicted in Figure 8.44. Hovering over the preservation icon (symbol of a shield) presents a detailed explanation of various aspects/dimensions as shown at the bottom of the figure.

Although no formal evaluation of the Preservation Prototype II has been conducted, first results obtained by using it at DFKI SDS were perceived as promising by involved colleagues. In combination with the PoF (Preserve or Forget) framework developed in the ForgetIT project (e.g. see Mezaris et al. [251]), the prototype can be used to automatically select items for preservation and send them to a long-term archive. Depending on user settings, this could be a permanent move, i.e. items are deleted from the user's devices.

[8]Stating the different cases for the domain of x is omitted for the sake of readability.

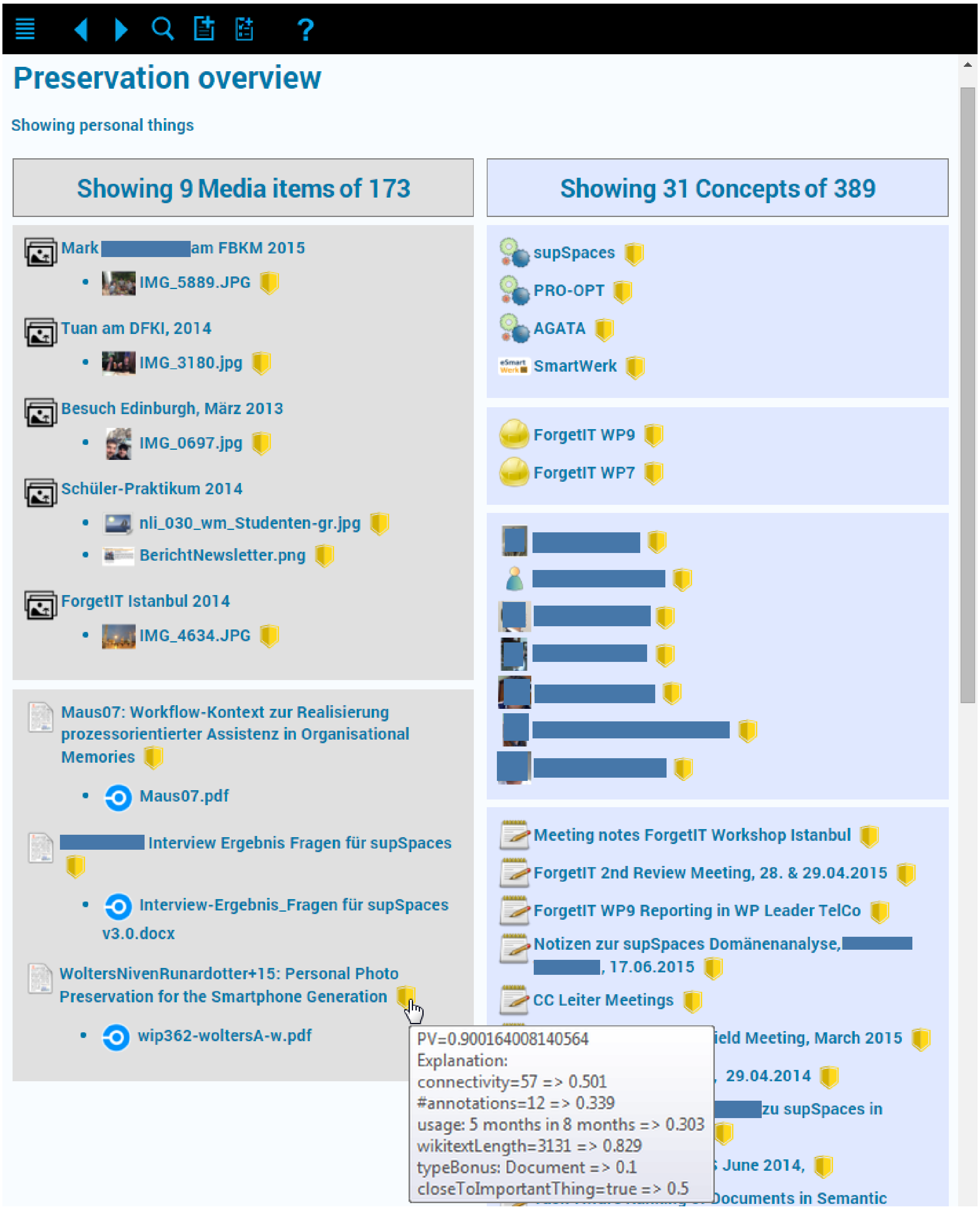


Figure 8.44: Preservation Prototype II: GUI and Explanations.
Abbreviations: *PV: Preservation Value*.
(Note: This is a screenshot by Heiko Maus. Variants of it have been published multiple times, e.g. [240, 245].)

8.6.2 cSpaces Automated Context Management

Section 3.3.1 introduced several self-(re)organization measures. This section particularly addresses context spawning and retraction as well as splitting and merging. For some features an early prototype is available, others are in a pre-prototypical state. If no prototype is presented, at least design and implementation ideas are discussed.

Context Spawning. A first example of context spawning is depicted in Figure 8.45. The example also gives a better impression of the cross-application aspect of working in context spaces enabled by cSpaces' deep integration. A user has created a calendar entry *PIMO Session* (highlighting A) having the description *PIMO Session with Heiko Maus and Sven Schwarz* and *Berlin* as its location. Additionally, the link to a website *www.comem.ai* is attached. Upon saving the calendar entry, the system gets the information via CalDAV [75], performs information extraction on its content and creates a new context space for it. Selecting the context in the sidebar (B) shows that this *context is about* the city of *Berlin*, the software (concept) *PIMO* and the persons *Heiko Maus* and *Sven Schwarz*. In addition, the attached website link has been added to the context's content (*context contains*). The selection of the context has already triggered a reorganization of the web browser (C) and the file system (D): In the *MyContexts* folder (D), there is now an additional folder for the *PIMO Session* context (injected via the SMB protocol [161, 254]). The attached link of the calendar entry has been injected into the *CurrentContext* bookmark folder via WebExtensions [261] (C) making web-related content directly available in the browser. Note: The scenario of the screenshot is also available as an online video – see the figure's description for respective links.

A second example of context spawning has been hinted at in Section 6.3.2 about the Stream Context Miner. The tool voting for an "unknown" context could suggest that a new context should possibly be created. Doing further investigations in this regard was considered a topic for future work.

Context Retraction. The Stream Context Miner is also relevant for the feature of context retraction. Even if observed evidence snippets were assigned to a certain context, the tool's observation goes on and possibly comes to the conclusion that another context would have been the better choice now that more evidence is available. Thus, if the user has not confirmed the context mapping yet, the earlier decision is revised and the item is re-assigned to the better matching context.

Context Splitting. For context splitting and merging, the clustering capabilities of PIMO Diary (see Sec. 8.4.1) could be helpful. cSpaces could simulate the diary app's similarity calculations in the background. This would mean that only for every new item entering a context a similarity score to its co-elements in the context needs to be calculated (i.e. linear runtime complexity). Instead of taking an excerpt (for a certain time interval) of a user's PIMO, the diary app would take the context as an input. If the diary app finds meaningful sub-contexts (with respect to parameter settings that still need to be identified in future work), a context split could be proposed to the user.

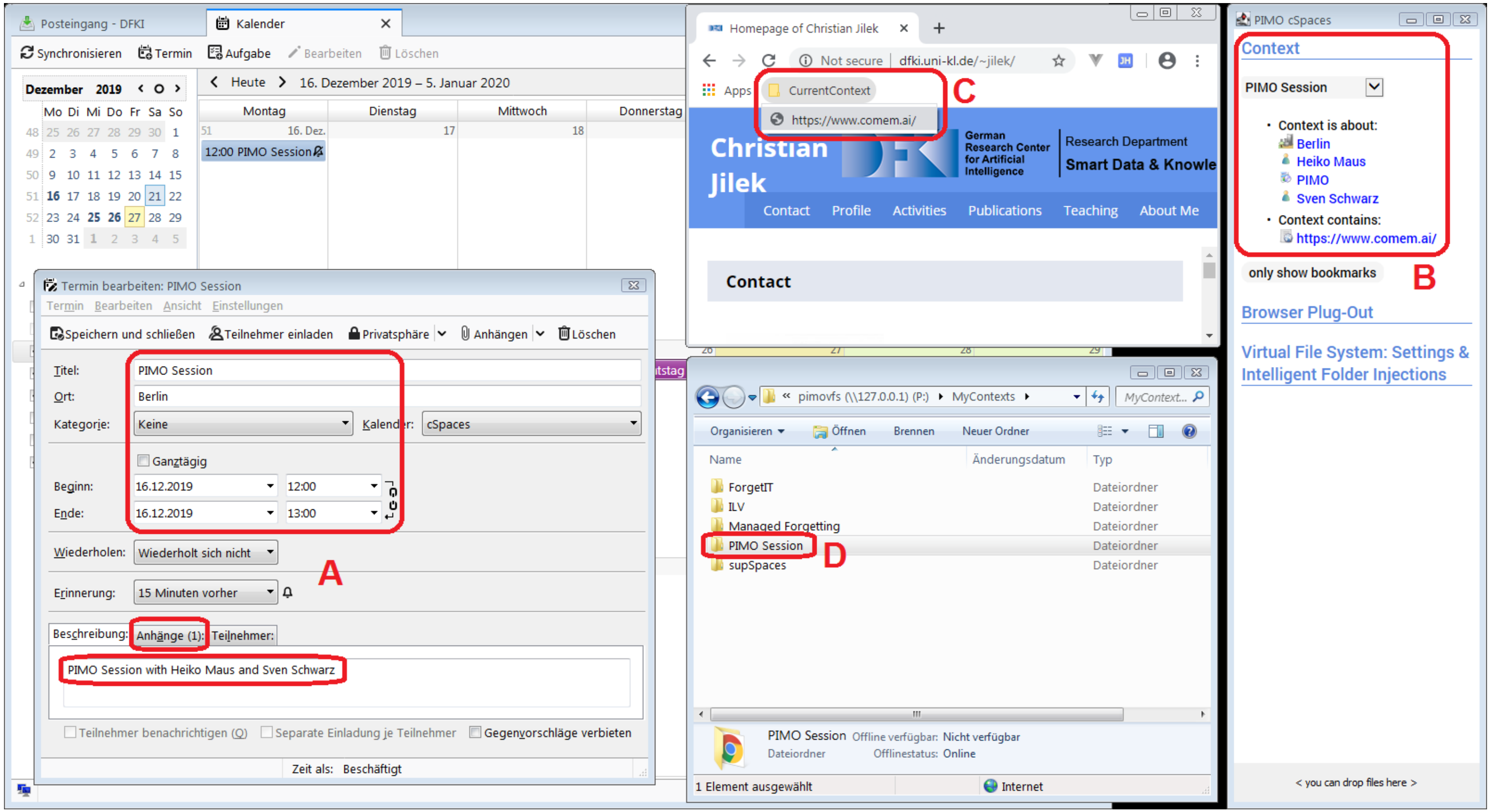


Figure 8.45: Automated Context Management: Example of Context Spawning.
Creating a calendar entry spawns a new context space and reorganization measures are performed upon its selection.
Highlightings: *A: creating a new calendar entry, in particular with a textual description and a link to a website / B: the spawned context space with the website as its content and automatically derived tags by* NER *(context is about) / C: upon selection, the context's content is injected into the web browser / D: the newly spawned context is also accessible as a folder in the file system.*
Also see corresponding video at `https://www.dfki.uni-kl.de/~jilek/demo/cspaces/#video` or `https://arxiv.org/src/1805.02181v1/anc/demo_video.mp4`.

Context Merging. In comparison, context merging is a simpler task. If two similar contexts (e.g. see the diary app's similarity metrics) lose relevance (MB drops), the system could propose or automatically perform a merge.

Full-automatically anticipating all contexts, sub-contexts and potentially helpful context reorganization measures is out of this thesis' scope – also remember the sensor and interpretation gap (Sec. 3.2) in this regard. cSpaces only seeks to increase the automation, i.e. the system should only act automatically in more certain cases, otherwise it should just propose an operation to the user, which they can accept or deny.

Survey Results. As shown in Figure 8.46, participants of the multi-month user study were asked about the usefulness of these support measures (items in blue ■) as well as their preference regarding the system automatically performing such measures (complemented by an undo function) vs. asking the user for permission first (item in red ■).

Most participants estimated the usefulness quite high (Q_{M11} to Q_{M13}): each item's median is *strongly agree*, whereas the mean value is in some cases a bit below, in others a bit above *quite agree*. All participants preferred automatically performing measures in combination with an undo function (Q_{M14}): the results were between *slightly* and *strongly agree* with a median and mean value of *quite agree*.

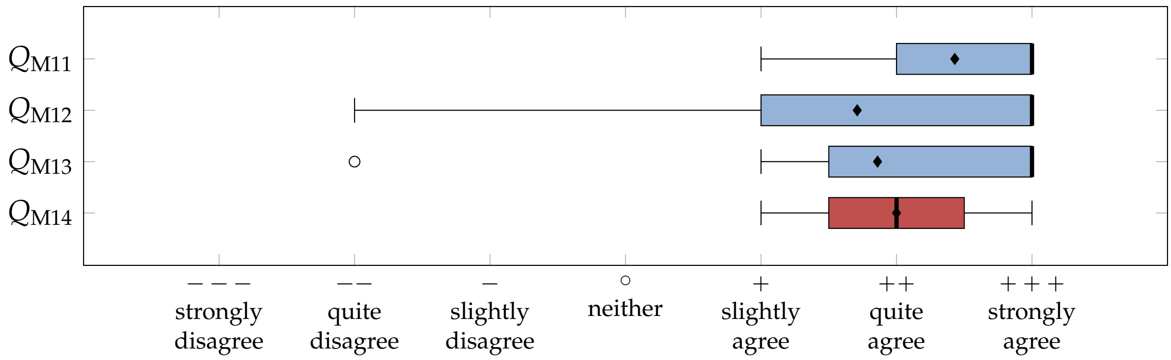


Figure 8.46: Multi-month User Study: Survey Result for Automated Context Management.

Q_{M11}: ■ Spawning & Retraction are a feature I would find useful.
Q_{M12}: ■ Splitting is a feature I would find useful.
Q_{M13}: ■ Merging is a feature I would find useful.
Q_{M14}: ■ Regarding self-(re)organization *in general*, I would prefer just doing over asking first, i.e. the system performs the support measures it finds appropriate without asking and has an undo function to correct it, instead of asking for user feedback before doing something. (Please remember: all reorganization measures are virtual, nothing is physically moved on the harddisk.)

(n=7; full questionnaire document in Sec. A.4; details on box plot semantics in Sec. B.2)

8.7 Searching Forgetting-enabled Information Systems (FIS) like cSpaces

Earlier sections (e.g. Sec. 3.1 have already addressed the topic of searching a forgetting-enabled information system (FIS). Remember the example of a user entering keywords into the search field of an FIS and no (or seemingly incomplete) results are shown. Users would then wonder whether they entered the "right" keywords or whether they really saved the information they are now looking for. In general, a balance has to be found between Managed Forgetting mechanisms that prevent users from being overwhelmed by the potentially high number of search results. But on the other hand, users still have to find the things they are looking for, especially if they come back to something not accessed for a long(er) time. Whereas *accessed* here especially means that a whole topic area of the KG has not been stimulated for a long(er) time – accessing related topics (i.e. in the KG neighborhood) would have raised the MB otherwise.

Several prototypes of FIS Search – sometimes also called *Forgetful Search* – have been developed as part of this PhD project. They are presented in the following. This section especially addressed research question RQ7.

8.7.1 FIS Search Prototype I in CoMem: The Memory Buoyancy Slider

A first FIS Search Prototype has been developed as part of CoMem. Like the fading out and resurfacing mechanisms (see Sec. 8.5), this search UI fully relies on Memory Buoyancy. Its development has also been led and mainly conducted by Sven Schwarz with only minor contributions by the author. The search interface is depicted in Figure 8.47. It is equipped with a slider to manually lower the MB threshold, which was intended to be used if a user explicitly wants to remember something that has already been forgotten by the system. Such things have a low buoyancy and do therefore not appear in the list of results. By lowering the threshold, more and more forgotten items are brought up and appear in the list, see highlightings A to B to C in the figure. But unexpectedly (at first), users started to habitually drag the slider in the low MB direction instead of rephrasing their query if sought things were not found in the first place. Obviously an indicator showing whether there are more relevant items was missing. This problem was addressed in a second prototype as discussed in the next section.

8.7.2 FIS Search Prototype II: Coverage Indication and Contextual Clustering

The challenge of establishing trust in such a highly active system is closely linked to the question of how to visualize what is actually forgotten (*forgotten* in the sense of Managed Forgetting, i.e. only a condensed version is still remaining or something is hidden by default or moved to an archive due to low MB). Users wondering why certain results do not show up in their list of search results could be informed that there is currently no search result in the "active" part of their data but something in the forgotten area. Possible measures could be visualizing how much of their data is covered by the search results and how much of these results belong to the active or forgotten area, respectively. Both areas could also be thematically/contextually clustered providing brief overviews that give hints whether further exploration of a cluster could lead to the sought items. Additionally, if the user enters an exact

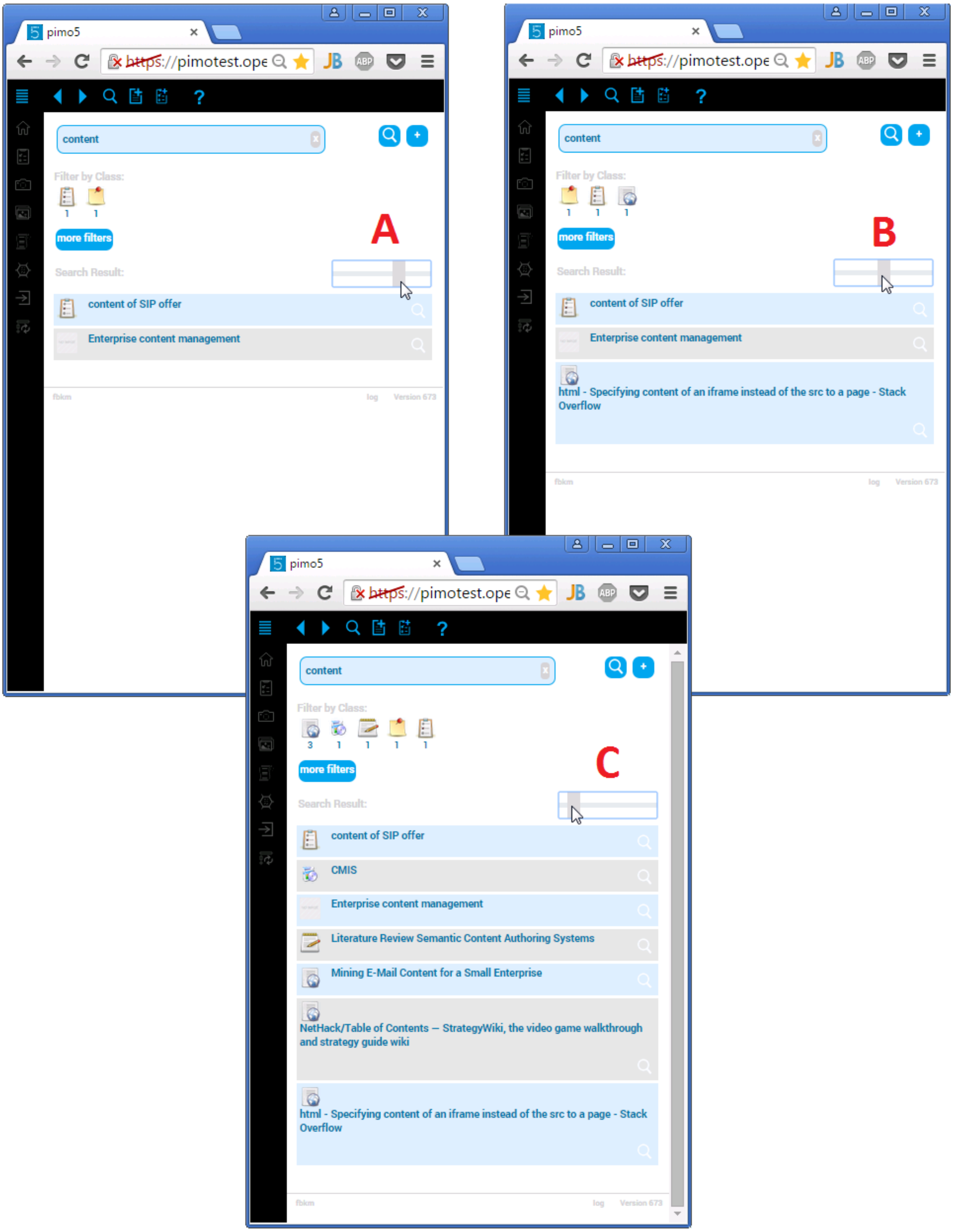


Figure 8.47: FIS Search Prototype I: The Memory Buoyancy Slider.
Dragging the MB slider to the left (A to B to C) lowers the MB threshold of items to display and thus brings up more forgotten results.
(Note: These screenshots have already been published in Jilek et al. [182].)

match of a thing's label, e.g. the full name of a person or project, then it would be justified to directly show actually forgotten items since the user seems to have remembered something that they have not browsed/used for a long(er) time.

A prototypical implementation of these improvements, *FIS Search Prototype II*, was developed as part of the bachelor thesis by Rudolf Koch (née Novik) [274]. The app is standalone (not integrated to CoMem yet) based on a pre-version of the ATIC KG Store (see Sec. 5.2.1). Its interface is depicted in Figure 8.48. It consists of an input field to enter search terms (A), a coverage indicator (B), contextual clusters of non-forgotten (C) and forgotten items (D) and the list of search results (E). Two more features were developed that can be accessed via buttons at the bottom right (F): a timeline view (screenshot in the appendix, Fig. A.3) and an *Assisted Remembering Mode*. The latter basically collects the tags of potential search results and asks whether the search results have something to do with one of these topics. Depending on the user's answer, the assistant either excludes items tagged with topic (negative answer) or missing the topic (positive answer). Then, the next question is asked to further narrow down the search – an example is depicted in Figure 8.49.

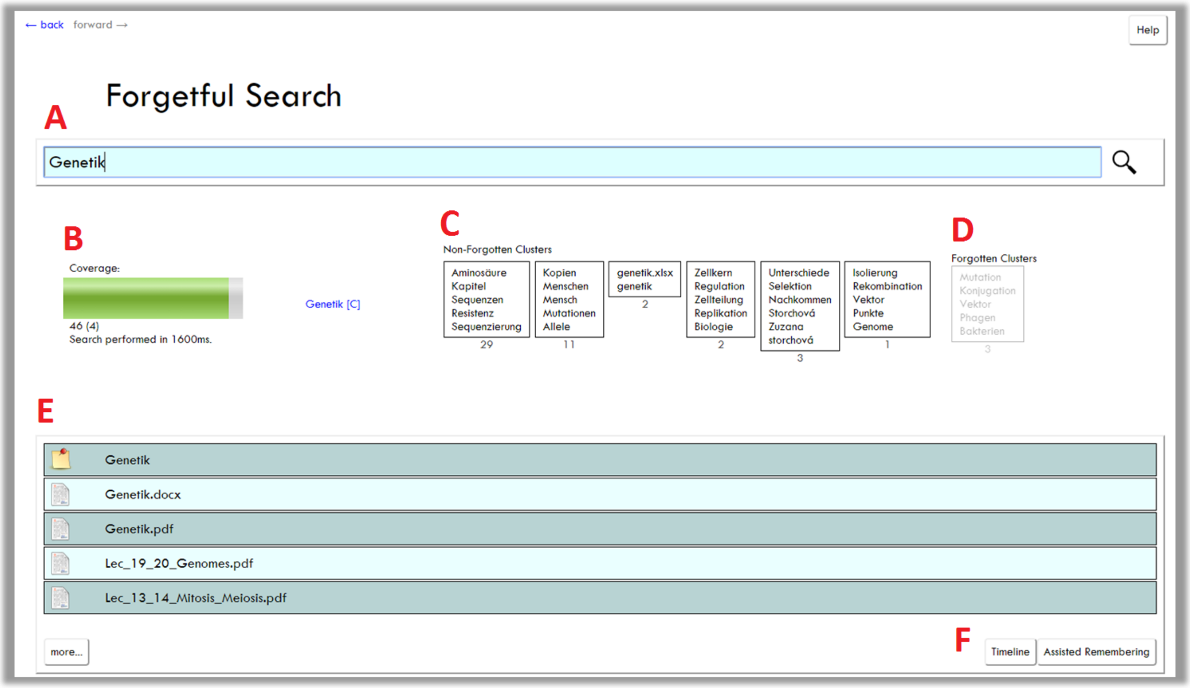


Figure 8.48: FIS Search Prototype II: GUI Realized in Novik [274].
Highlighted sections: *A: input field to enter search terms / B: coverage indicator / C: contextual clusters of non-forgotten items / D: contextual clusters of forgotten items / E: search results / F: buttons to access timeline view and assisted remembering mode.*
(Image source: supplementary material of Novik [274].)

Evaluation. The FIS Search Prototype II was evaluated with seven participants in a short-term experiment. Participants first bootstrapped a scenario using the *PIM Crawler* (see Sec. 5.3.1) and the *Semantifier* (Sec. 5.3.2). Since the *Contextifier* did not exist at the time, users were asked

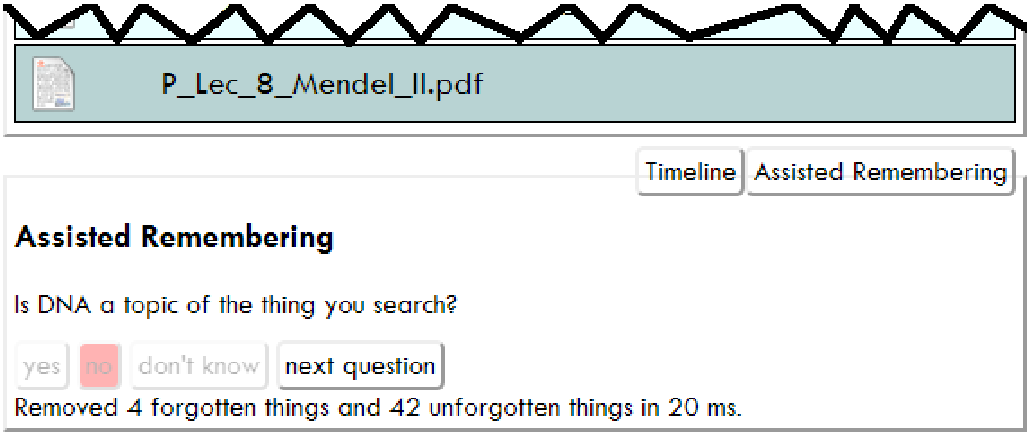


Figure 8.49: FIS Search Prototype II: Assisted Remembering.
(Image source: Novik [274], image cropped to focus on main aspect.)

to pick several folders that may serve as contexts. The were also asked to name those items that they have not accessed for a longer time, which were used as elements of the forgotten area. All other ones went into the active area. Participants were given a guideline ensuring that they step by step explore all features of the system, e.g. searching for forgotten as well as non-forgotten items, using the timeline view as well as the assisted remembering mode, etc.

Last, they were asked twelve standardized questions about the app's usability (with two items slightly rephrased by Novik [274]) and eight custom questions about its core functionality. As depicted in Figure 8.50, usability scores were quite positive: eleven of twelve mean values are in a range slightly below *slightly agree* to a bit above *quite agree*. Items getting slightly more negative scores are *fewest steps possible to accomplish a task* (Q_{S6}) and *learning quickly how to use it* (Q_{S7}). Scores for the core functionality were in general a bit higher than the ones for usability as shown in Figure 8.51. Three of eight items have mean values between *quite* and *strongly agree* (Q_{S15}, Q_{S19} and Q_{S20}), four between *slightly* and *quite agree* and a last, negatively phrased item (Q_{S16}) was rated between *slightly* and *quite disagree*.

In summary, participants had the impression to also find forgotten items if searching for them and forgotten items were only shown if that made sense. Participants stated in particular that they would entrust a forgetting-enabled information system with their data if they were given such search capabilities to retrieve them. Contextual clusters and the coverage indicator were perceived as very helpful (Q_{S20} and Q_{S19}, median: *strongly agree*, mean value between *quite* and *strongly agree*). Assisted Remembering and the timeline view were positively assessed as well but with slightly lower scores (Q_{S17} and Q_{S18}).

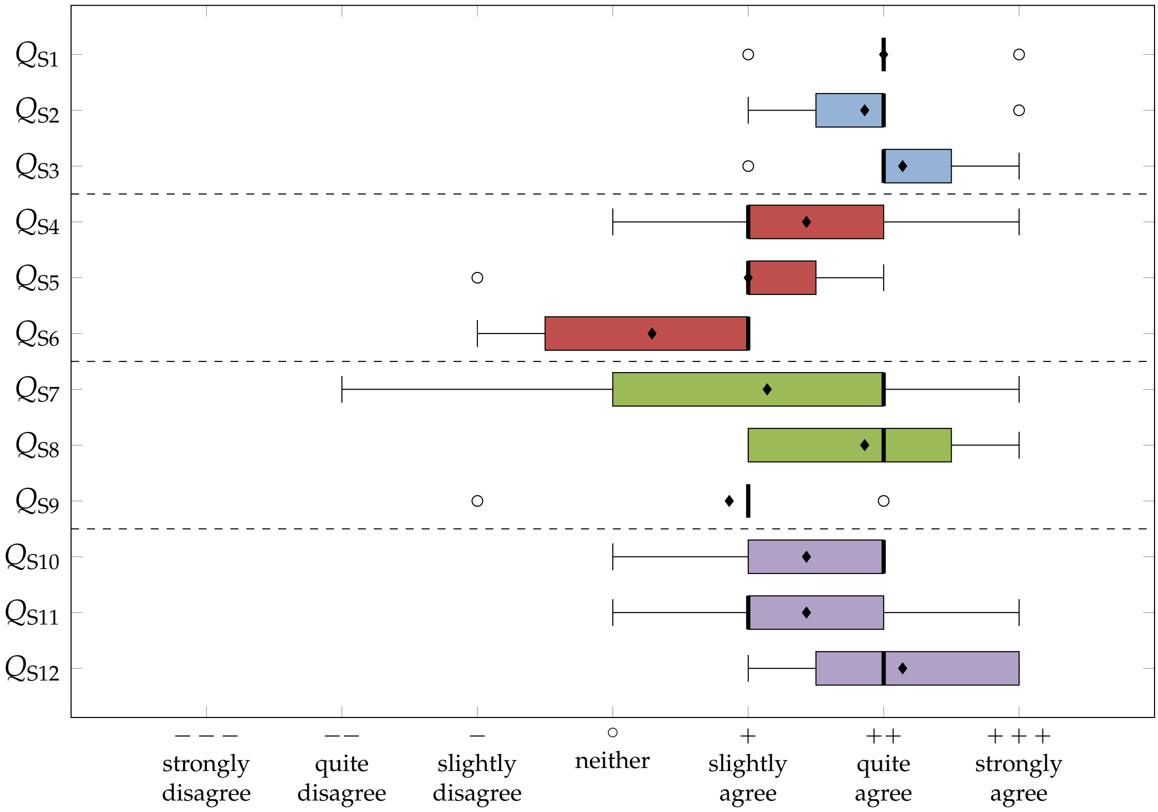


Figure 8.50: FIS Search Prototype II: Survey Result for Usability as Reported in Novik [274].

Q_{S1}–Q_{S3}: Usefulness,
Q_{S4}–Q_{S6}: Ease of Use,
Q_{S7}–Q_{S9}: Ease of Learning,
Q_{S10}–Q_{S12}: Satisfaction.

Q_{S1}: In long term use, I think it will help me to be more effective.
Q_{S2}: In long term use, I think it will help me to be more productive.
Q_{S3}: It is useful.
Q_{S4}: It is easy to use.
Q_{S5}: It is user friendly.
Q_{S6}: It requires the fewest steps possible to accomplish what I want to do with it.
Q_{S7}: I learned to use it quickly.
Q_{S8}: I easily remember how to use it.
Q_{S9}: It is easy to learn to use it.
Q_{S10}: I am satisfied with it.
Q_{S11}: I would recommend it to a friend.
Q_{S12}: It is fun to use.

(n=7; questions from USE questionnaire [232], whereas items Q_{S1} and Q_{S2} were slightly rephrased by Novik; details on box plot semantics in Section B.2)

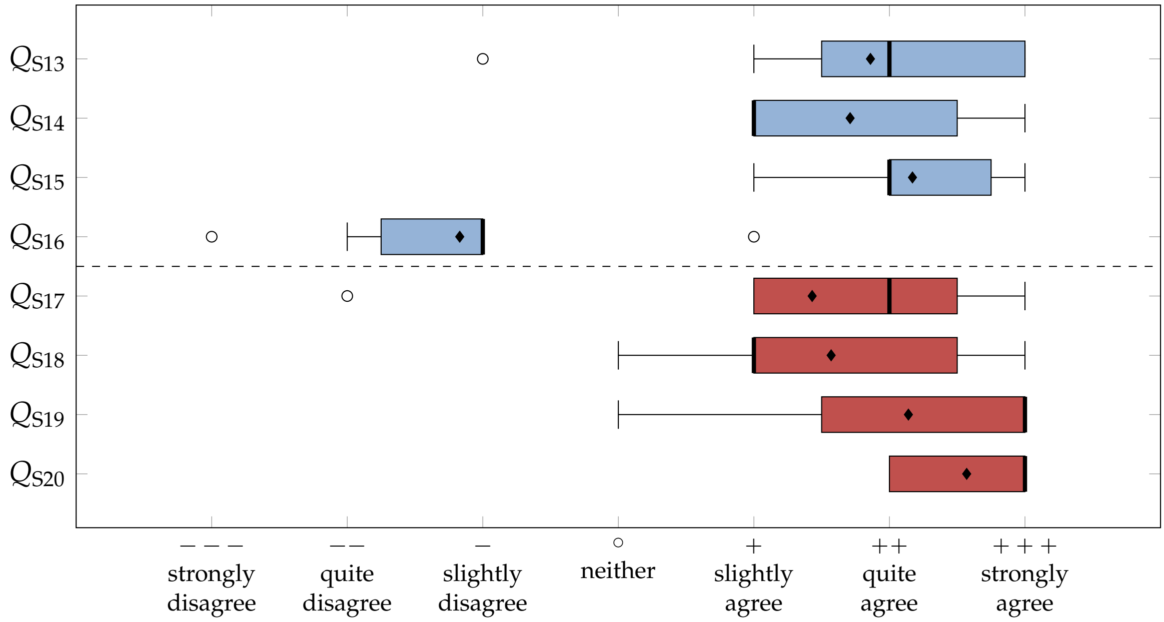


Figure 8.51: FIS Search Prototype II: Survey Result for Core Features as Reported in Novik [274].

Q_{S13}: I have the impression that I can find everything contained in the system, even if it is forgotten.
Q_{S14}: Forgotten things are showed only if it makes sense.
Q_{S15}: I would entrust a Forgetful Information System with my data if I could use Forgetful Search to retrieve it.
Q_{S16}: I would entrust a Forgetful Information System with my data if I could NOT use Forgetful Search to retrieve it.
Q_{S17}: I think that *Assisted Remembering* can help me to find forgotten things I search by narrowing down search results.
Q_{S18}: In my opinion, the *Timeline* is a good visualization of the temporal aspects of search results.
Q_{S19}: The *Coverage Indicator* helps me to orientate regarding the question how many of the search results are hidden by the Forgetful Information System.
Q_{S20}: I think that *Clusters* are a good approach to get an overview of the search results.

(n=7; details on box plot semantics in Section B.2)

8.7.3 FIS Search Prototype III: Integrated cSpaces

FIS Search Prototype II had a makeshift solution for contexts since cSpaces was not ready for integration at the time. A third prototype still in development aims at such a full integration, making context spaces directly available and especially browsable in the search. Search results are typically not shown as a single entry but together with information about the context(s) they are contained in. If several results belong to the same context, this may alter the ranking, bringing that context further up in the result list. A balance has thus to be found between individual items' relevance scores and their respective contexts' scores. Search will not be restricted to checking a list of results but users may also directly navigate through the different contexts.

The prototype is in a pre-prototypical state and its GUI is depicted in Figure 8.52. Its development was supported by Desiree Heim as part of her student research assistant job at the time.

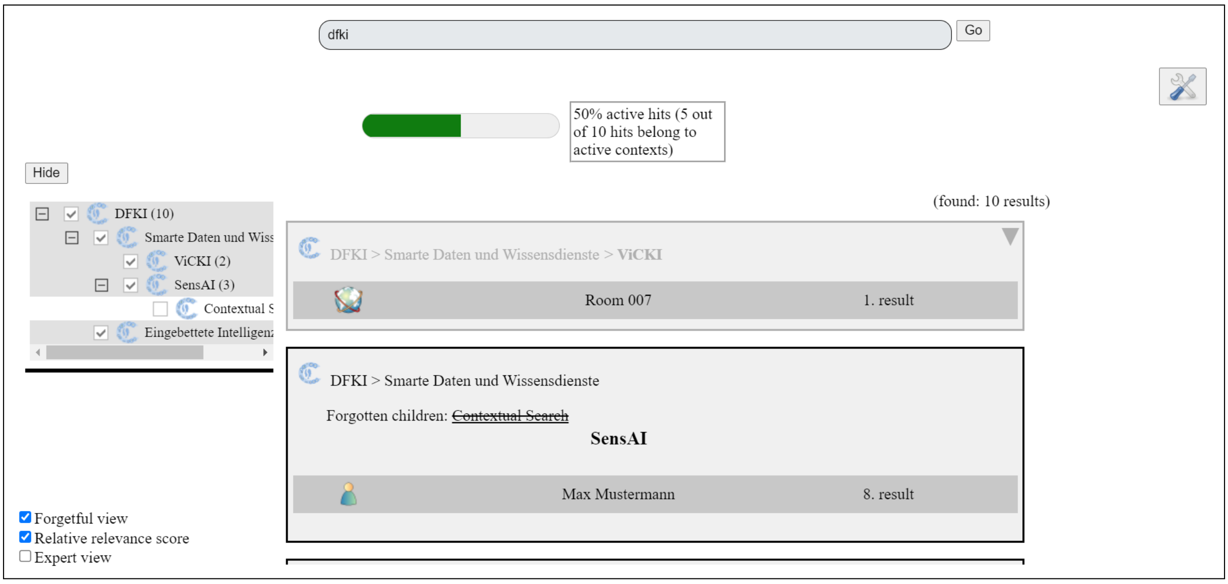


Figure 8.52: FIS Search Prototype III: GUI.

As already mentioned, being able to find forgotten items if they are searched for facilitates trust in the system. Trust is also the main topic of this chapter's last section.

8.8 Trust in Highly Autonomous Assistants like cSpaces

In the scenario of a highly autonomous AI-based assistant like cSpaces that reorganizes a user's information sphere, trust is particularly important – and therefore also the topic of research question RQ8. This comprises trust in search results and the system's actions and not losing the feeling of control, for example.

Online Study. A study on trust in an assistant like cSpaces has been conducted as part of the master thesis by Julia Knabe [207] with Anna-Sophie Ulfert-Blank[9] as the main supervisor and the author as a co-supervisor. The study was set up as an online inquiry, in which participants assumed the role of a knowledge worker being supported by cSpaces. As in the previous experiments and depicted in Figure 8.53, cSpaces was available in the form of a sidebar on one side of the screen. However, the study was completely conducted based on screenshots. In three different situations, participants were shown three different variants of status messages of the assistant. These messages were phrased to correspond to levels of the *Situation Awareness-based Agent Transparency (SAT) model* by Chen et al. [58]. A first variant

[9]Industrial Engineering and Innovation Sciences, Human Performance Management, TU Eindhoven, https://research.tue.nl/en/persons/anna-sophie-ulfert-blank (Post-doctoral researcher at Goethe University Frankfurt am Main at the time)

only stated the assistant's *goals and actions*, e.g. that a new context has just been created by the assistant. A second variant additionally explained a bit of *reasoning* behind the action, e.g. that the context has been created since the last four actions strongly addressed a certain project. And a third variant also gave *projections* on the assistant's future actions. For example, if additional evidences confirmed the switch to this new context, the last three actions performed by the user would also be assigned to the newly created context. The situations and messages (in German) are depicted in Figure 8.54.

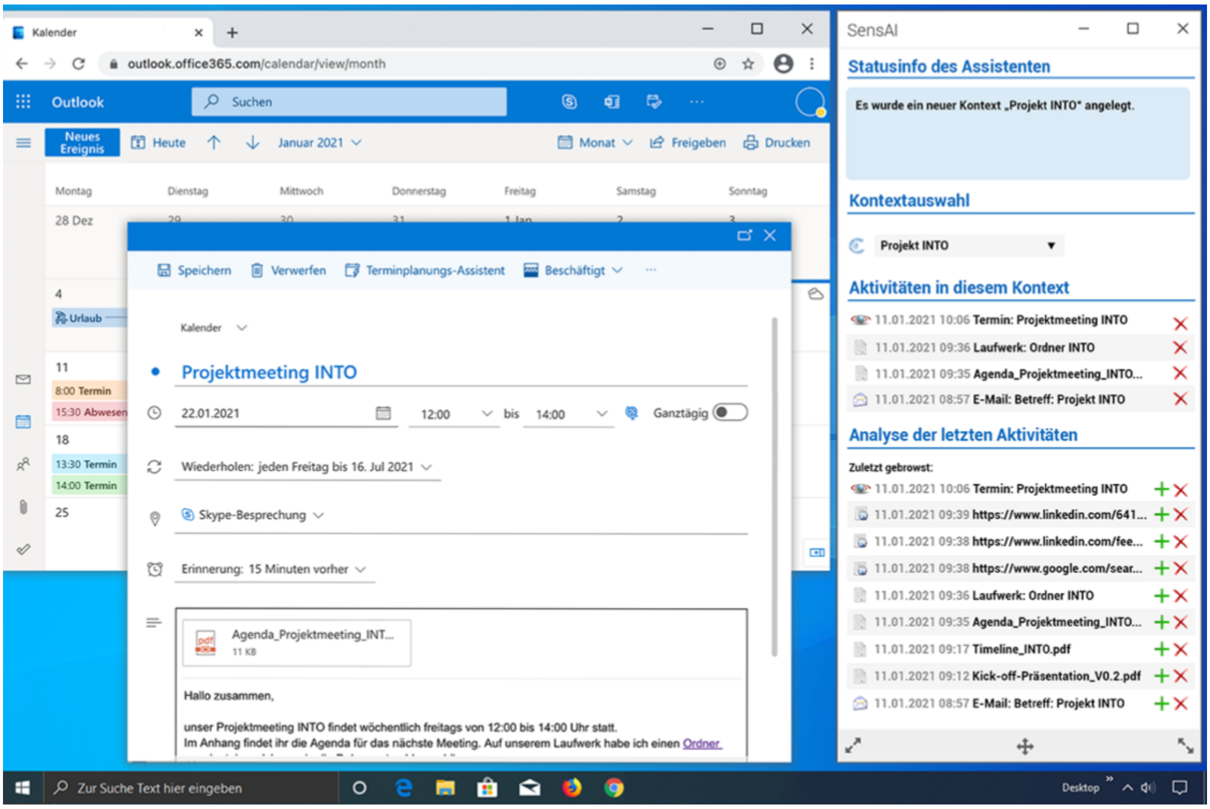


Figure 8.53: Trust Study: Scenario.
(Image source: Knabe [207].
Note: A variant of this image may be part of a publication currently in preparation [383].)

Participants then stated their perception of the system in various dimensions and also gave preferences. The most important results of the study are as follows:

- Trust and intention to use increased as a function of higher transparency.
- There was no moderation effect of propensity to trust.
- Transparency perceptions differed across situations.
- Participants formed a preference.

The study was conducted in two months involving 140 participants. For more details as well as concrete figures, readers are kindly referred to a paper currently in preparation [383][10].

[10]Especially concrete figures were omitted here to ensure that Ulfert-Blank et al. [383] (in preparation) is actually the first publication of the results (some journals might have strict rules also including preceding dissertations in this regard).

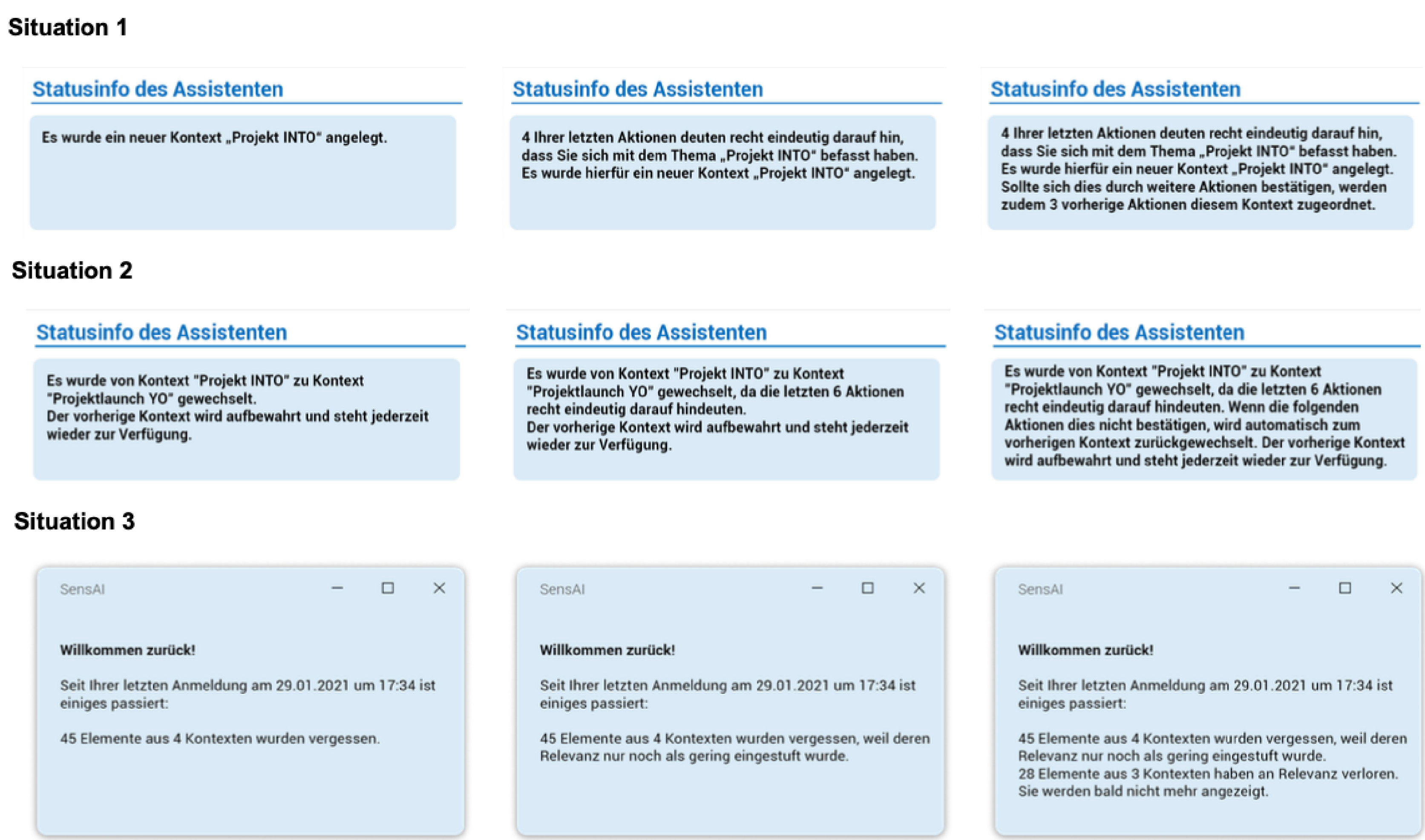


Figure 8.54: Trust Study: Variants of Status Messages of the Assistant as Reported in Knabe [207]. In three different situations (rows), the assistant informs the user about its actions with messages having different information content: 1. the assistant's *goals & actions* (left), 2. additional *reasoning* (center), 3. additional *projections* (right) [58]. (Image source: Knabe [207]. Note: This image may be part of a publication currently in preparation [383].)

In conclusion, more explanations were not necessarily perceived as more transparency, and for different situations different levels were preferred. Explanations by an assistant like cSpaces should thus be adaptable. Users should be able to regulate their amount. For example, in the beginning, when working with such a system is still a new experience, more explanations are probably helpful. After some time, when users are more accustomed to the system, they may prefer to regulate down their amount.

Related Work. Related publications are, for example, Shin [352] and Weitz et al. [396]. The main difference between them and cSpaces is the focus and setting of cSpaces being a self-(re)organizing and especially forgetting-enabled information system.

Survey Result of Multi-month User Study. In contrast to the online inquiry, participants of the multi-month user study have actually used cSpaces. They were thus asked whether they would trust the reorganization decisions of the envisioned system (i.e. a system having the complete bandwidth of envisioned measures). One participant *slightly disagreed*, another one *strongly agreed* and all other ones *quite agreed* (Q_{M28}). A last item, which is also related to trust, asked whether cSpaces needed more on-screen explanations for its support measures. Answers were mostly between *slightly disagree* and *slightly agree* with a median and mean value of *neither* (Q_{M29}).

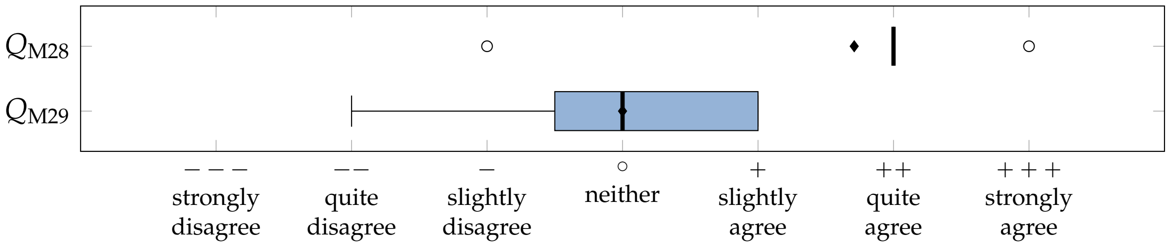


Figure 8.55: Multi-month User Study: Survey Result for Trust and Explanations.

Q_{M28}: I would trust the *envisioned* prototype of cSpaces in its hiding, condensation and reorganization decisions.

Q_{M29}: The support measures as provided by the *envisioned* prototype of cSpaces need more (on-screen) explanations.

(n=7; full questionnaire document in Sec. A.4; details on box plot semantics in Sec. B.2)

Part II Conclusion

Part II presented the thesis' main approach aligned to the cSpaces interaction cycle (see Sec. 1.4). For each of its steps/aspects, solutions have been designed, implemented and discussed. Some of them are still in an early or pre-prototypical state, while others are stable enough to be already used in daily work, including research and industry. The last part of the thesis, Part III, first gives an overview of the transfer of some ideas and components into practice and also discusses further potential of the approach (Ch. 9). Last, the thesis is summarized and concluded, and an outlook on possible future work is given (Ch. 10).

Part III

Conclusion and Outlook

9 Further Potential and Transfer into Practice

Part II presented this thesis' main approach. While some ideas and assets are still in an early or pre-prototypical state, others are mature and stable enough to already be used in daily work. This chapter discusses this transfer into practice (research and industry) as well as further potential and application scenarios only briefly investigated so far. The chapter is ordered according to the five descending levels of maturity mentioned in this thesis' introduction (Sec. 1.5):

1. the most elaborated ideas and stable prototypes of the main approach which have already found their way into practice, e.g. as software components, parts of feasibility studies or as a mere knowledge transfer in the form of consulting jobs (Sec. 9.1),
2. elaborated ideas realized in early prototypes (Sec. 9.2),
3. elaborated ideas with experimental code in pre-prototypical state,
4. elaborated ideas sketched in position papers (Sec. 9.3) and
5. early ideas (Sec. 9.4).

The current chapter (Ch. 9) is the first half of Part III of the thesis. Its second half (Ch. 10) summarizes and concludes the thesis and also gives an outlook on possible future work.

Note: Parts of this chapter have already been published [21, 188, 290, 373].

9.1 Transfer into Practice

cSpaces' range of features steadily grew over the course of this PhD project. Development and investigations were spread across various lanes. New experimental ones were added step by step, while existing ones steadily matured (see the sequence of maturity levels above).

Assets that found their way into practice can be assigned to nearly all steps/aspects of the cSpaces interaction cycle (see Sec. 1.4). An integration to CoMem typically implies usage in research and industry since CoMem is often a foundation of such projects as explained in more detail in Section 2.3.3. In the following, the most prominent examples are presented. However, some industry transfer projects cannot be mentioned since they are affected by non-disclosure agreements.

ATIC Knowledge Graph Storage, Indexing and Search. The ATIC KG Store (see Sec. 5.2) was developed to be mostly compatible to CoMem. Thus, several of its components could already be integrated while the overall store was still in development. Most prominently, these were the combination of structured and indexed search (see Sec. 5.2.3) with sub-stores based on the idea of *Multi-model Databases* (MMDB) [231], including vector representation and access management solutions. One of the first versions of the CoMem Search is depicted in Figure 5.5.

CoMem Search was first used in (as well as developed and tested for) smaller CoMem instances like the one at DFKI SDS, which, at the time of writing, had about 96 thousand resources and 1.5 million relations. For comparison reasons: a CoMem instance productively used in industry at enviaM [109], for example, contained 9 million resources and about 168 million relations at the time of writing. Although a bit of optimization is needed in the near future to better cope with the scaling, ATIC KG components used in the implementation performed quite well considering a scaling factor of more than 110.

Additionally, based on discussions with Markus Schröder and the author, Sven Schwarz developed a data docking (integration) solution inspired by ATIC's Native Docks and Registry (see Sec. 5.2.2).

Search prototypes based on ATIC KG components have meanwhile been used 24/7 in CoMem (including industry partners like enviaM [109]) for about six years. Other ATIC components were used as part of student projects and theses, e.g. [63, 65, 274].

Knowledge Graph Bootstrapping and Data Import. Industry projects at DFKI SDS often have a data import or KG bootstrapping task in their early phase. The same is true for research projects, whereas they may also have such a task in late phases, for example to set up scenarios for evaluation.

Tools like the *PIM Crawler*, *Semantifier* and *Contextifier*, introduced in Section 5.3, as well as Markus Schröder's *Hephaistos Toolkit* (see Sec. 5.3.4) are often used for such tasks. Before the availability of several of these tools, in particular Hephaistos, the author has developed data import tools for CoMem, for example based on Spreadsheets (e.g. see Fig. 8.1). Some of these tools have thus been used for up to seven years (in this case, not 24/7 but every once in a while).

User Activity Tracking and Plug-out Architecture. Semantic Desktop re-engineering, especially the introduction of *plug-outs* and AEBI has been discussed in great detail in Section 6.1.1. These tools are meanwhile in 24/7 use in CoMem for more than six years. Industry partners like enviaM [109] recently started using these plug-outs and infrastructure for first steps and experiments together with DFKI SDS.

Information Extraction. The IT-RTC-OBNER approach presented in Section 6.2.2 has been used 24/7 in CoMem (including installations at the customer) for more than six years now. At first, a pre-version of Prototype I was used, which was replaced by Prototype II (see Sec. 6.2.4) in 2019.

In a collaboration of Wacom [389] and DFKI that started in 2018 [388], the author served as a consultant guiding Wacom to develop a variant of IT-RTC-OBNER tailored for their *Digital/Semantic Ink* technology and intended to be integrated into *Wacom Notes*[1], a worldwide available product.

[1] https://www.wacom.com/en-en/products/wacom-notes

Working with and in Context Spaces (Context-sensitive Assistance). Regarding working with and in context spaces, several cSpaces user interfaces are also available in CoMem, although not to their full extent, yet. In Section 8.2.2, the *PIMO (CoMem) Desktop Application* (depicted in Fig. 6.2) was mentioned, which is meanwhile available for about 4.5 years. Additionally, the guiding mock-up of a context dashboard (Fig. 8.5) was presented after which a dashboard for the property management department of *enviaM* [109] was created. Figure 9.1 shows a first prototypical implementation by Sven Schwarz having the property (i.e. a piece of land) in its focus area and additional information like related files or mails, information of the land registry office, etc. in widgets around it. A second example of such a context dashboard was developed as part of the *supSpaces* project [370] presented in Section 9.2.1.

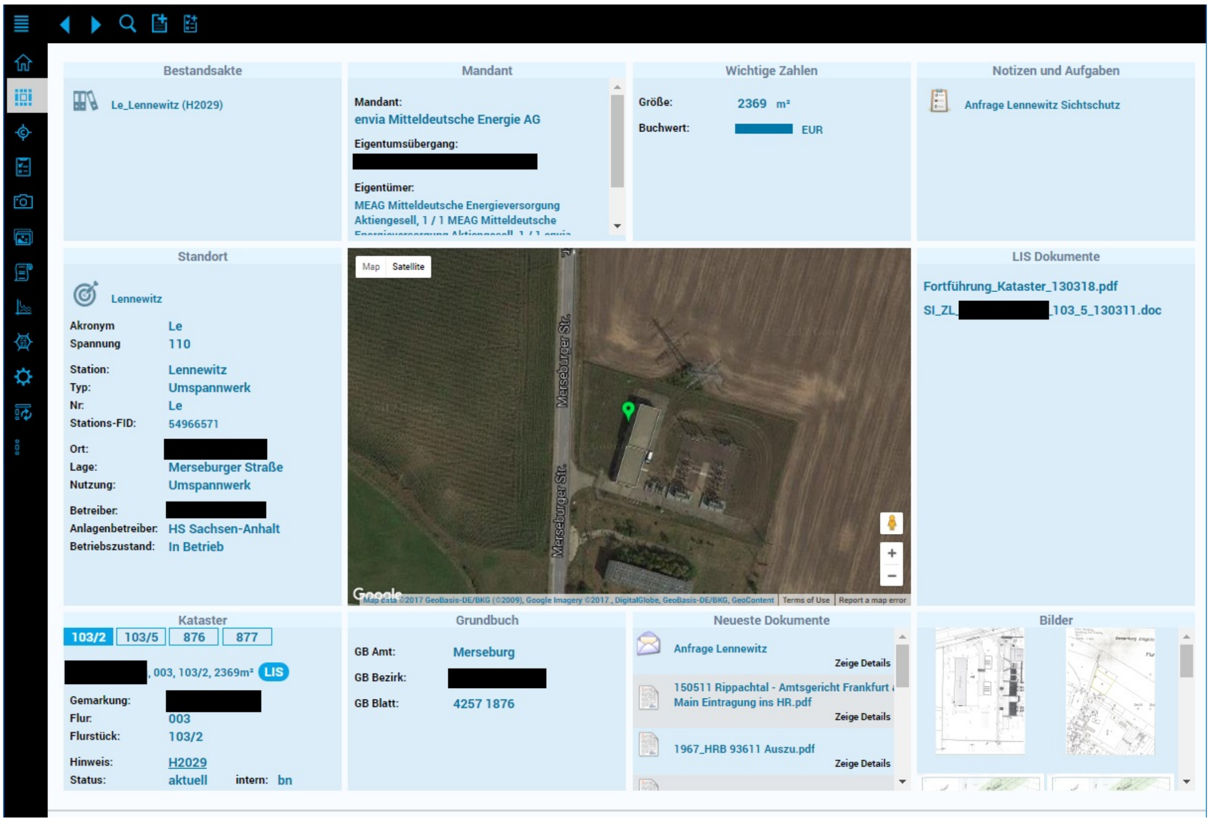


Figure 9.1: Prototypical Context Dashboard Realized in an enviaM [109] Project.
The property (i.e. a piece of land) in the focus area of the dashboard and additional information like related files or mails, information of the land registry office, etc. in widgets (tiles) around it. (Note: This is a screenshot by Heiko Maus. A variant of it has been published in Dengel and Maus [86].)

Managed Forgetting. The second phase (2020-2023) of the Managed Forgetting project [234] tries to bring some of the support measures to a corporate setting. This work is funded by the DFG and supported by enviaM [109] as a cooperating company in the experiments, which are still ongoing.

Further Dissemination and Teaching. Apart from core research activities and industry projects, the author also gave **invited talks and presentations** at various occasions including

- *Die Rheinpfalz*, newspaper interview, 2023 [233],
- *Die Wirtschaftszeitung*, newspaper interview, 2021 [357],
- *Sensorik X. HR-Expertenforum 2020* (invited talk) [181],
- *13th DBpedia Community Meeting 2019* (invited talk) [180],
- *KMU-innovativ: IKT Symposium 2016* (poster presentation) [179],
- *CeBIT 2015* (demo presentation) [239]

as well as (invited) talks at the *Universities of Trier* (2017, 2019), *Kaiserslautern* (2019) and *Frankfurt am Main* (2020).

The author also served as a **peer reviewer** (e.g. CHI, MobileHCI, WAHM @ UbiComp) and supported the teaching activities of the *Artificial Intelligence Group* (Computer Science Department, TU Kaiserslautern) headed by Prof. Dengel, for example as a **guest lecturer** of *Applications of Machine Learning and Data Science* (2022, 2023, 2024) and **(co-)supervisor of 37 student projects, seminar papers and theses**.

Further details can be found in the Author's CV available in the appendix.

9.2 Elaborated Ideas Realized in Early Prototypes

This section presents elaborated ideas that have been realized in early prototypes. For some of them (e.g. the supSpaces prototype), an early evaluation has already been conducted.

9.2.1 supSpaces – cSpaces in IT Support Scenarios

The term *supSpaces* is short for *support (knowledge) spaces for knowledge management in IT support*. It was a two-year, BMBF-funded project [370] of four partners:

- Mansystems (CLEVR[2] from 2020 on), a developer of service management applications,
- DFKI and two application partners,
- Deutsche Telekom[3], a German telecommunications company, and
- Dogado[4], a German cloud service provider.

Its focus was on supporting knowledge management in IT support scenarios. Such a scenario typically was a follows: A ticket is spawned, either by a monitoring software or by a customer. The core idea of supSpaces is to capture the context space emerging from such a ticket – in terms of cSpaces, the ticket is the context's nulceus. By claiming a ticket, a clerk creates and enters such a supSpace. Its GUI is a dashboard depicted in Figure 9.2. Highlighting A shows the claimed ticket. There is also a text area for taking notes in that particular context space (B). After the initialization of the context space, the system first performs information extraction on the ticket as described in Section 6.2.2. It shows identified named entities in the GUI's

[2] https://www.clevr.com/
[3] https://www.telekom.de/
[4] https://www.dogado.de/

middle section (C). Based on the ticket analysis, support tickets of the past (as well as their solutions – if available) and relevant resources are recommended (D). Clicking a resource loads its content inside the dashboard (E). Following CoMem's idea of being a meta-system "on top" of legacy systems, clerks may also follow a link to jump to the original location of the resource. Later versions also allowed for keyword based search using an early version of CoMem/ATIC search (see Sec. 5.2.3). By confirming relevant topics and helpful material, writing comments, associating items with the ticket's context, etc., clerks may easily document their progress, solution attempts (particularly helpful if the ticket cannot be solved before a shift change) and finally their solution. During the processing of the ticket, the supSpaces system live-updates its list of proposed supportive material after each captured user activity – one of the reasons why information extraction tasks should be completed within milliseconds (see IT-RTC-OBNER motivation in Sec. 6.2.2). Then, the system finally saves the supSpace, i.e. the ticket together with its rich context consisting of all information described before. If such a problem or a similar one occurs (again), the system will bring up the previously captured supSpace. Users can then decide whether the solution to the old problem can actually be transferred to the new one, which is now considerably easier given the old ticket's additional contextual information. The approach was mainly realized by means of *Case-Based Reasoning (CBR)* [2].

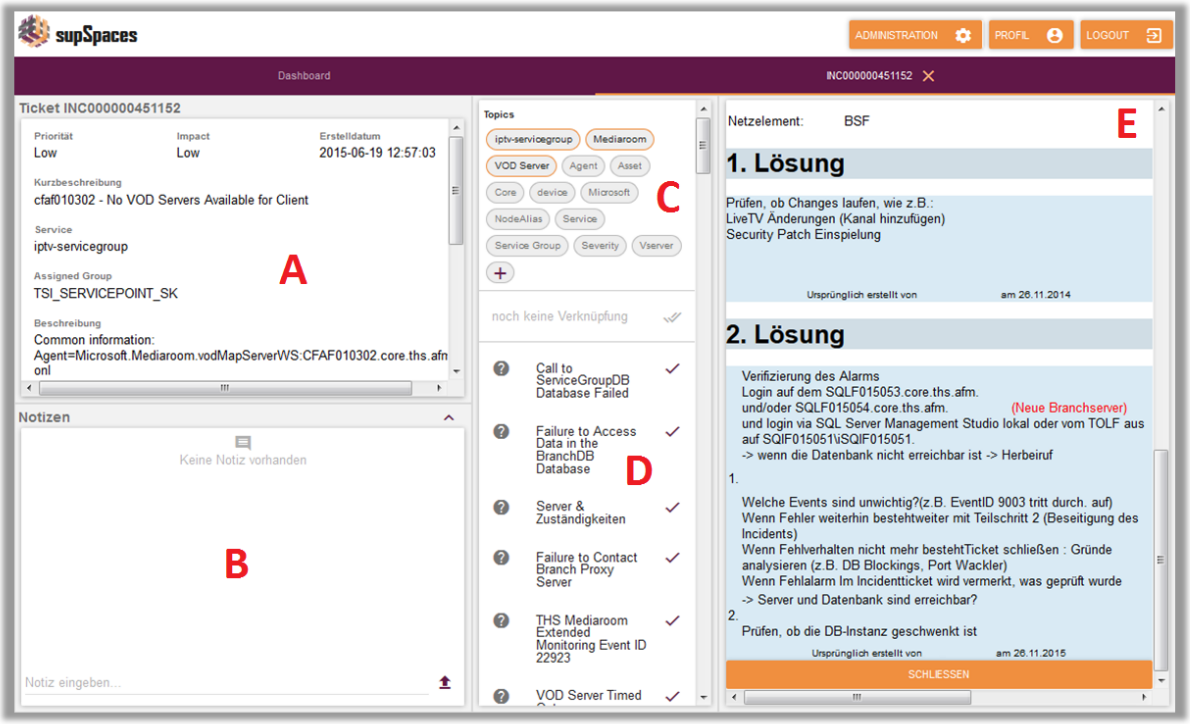


Figure 9.2: Prototypical Context Dashboard Realized in the *supSpaces* Project [370].
Highlightings: *A: the support ticket / B: notepad / C: suggested and confirmed topics (tags) / D: similar tickets (including solutions) and relevant resources (recommended based on the ticket itself and tags) / E: clicking an entity (in D) shows its content in this widget*. Later versions also allowed for keyword based search using an early version of CoMem/ATIC search (see Sec. 5.2.3).

Evaluation. The supSpaces prototype as described above was evaluated in an expert walk-through (structured interview). A clerk of Deutsche Telekom used the system to solve a set of (past) tickets and rated the system's different features while doing so. (Since there were several clerks solving tickets in the department, each clerk only knew a subset of all past tickets.) The overall feedback was generally very positive. As a side outcome, the interviewed expert stated that a context widget permanently visible at the edge of their screen would have been even more helpful since they were already switching a lot between windows opened in parallel of which the dashboard was yet another one. As mentioned in Section 8.2.2, this remark encouraged the author in his interest/decision to focus more on sidebar (and other) widgets as a complement to dashboards.

9.2.2 PIMO Bot – cSpaces-based and -dedicated Chat Bots

Among other tools, DFKI SDS also uses the group chat *Mattermost* [Sw73] for intra-team communication. Together with Sven Schwarz, Rudolf Koch and some students over the years, the author realized *PIMO Bot*, a chat bot integrated in Mattermost and connected to CoMem. *PIMO* Bot can be invited to arbitrary channels (open groups) or rooms (closed groups) or can be talked to directly in a private conversation. PIMO Bot can be configured to associate channels and rooms with certain topics and whenever new things are added in CoMem, it informs users of the respective channel or room as depicted in Figure 9.3. It also listens to users' commands to do a search, tag items, etc.

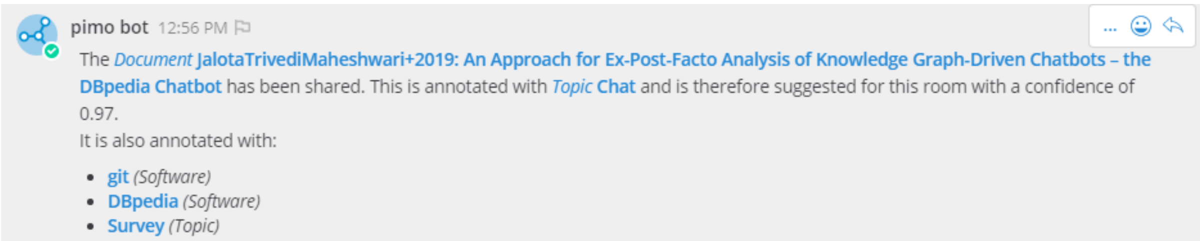


Figure 9.3: PIMO Bot.
In the depicted chat message, *PIMO Bot* recommends a newly shared resource of *CoMem* in a thematically relevant chatroom of DFKI's group chat *Mattermost* [Sw73].

PIMO Bot is a first attempt of realizing a context-sensitive chat bot for CoMem or cSpaces, respectively. In future work, such a bot could, on the one hand, keep an overview of all of a user's contexts and maybe give hints at deadlines etc. And on the other hand, it also switches from context to context with the user, telling them the context's "golden thread", i.e. what was the context's last state or last actions, what was planned to do next, etc.

PIMO Bot was used 24/7 in the DFKI SDS department for more than five years – at the time of writing, it is down for maintenance though (necessary adaptations to API changes).

9.3 Elaborated Ideas Sketched in Position Papers

This section presents elaborated ideas not yet available as a prototype but sketched in position papers.

9.3.1 cSpaces as Part of a Digital Twin of an Organization for Enterprise Modeling

Riss et al. [290] proposes cSpaces to be used as part of a *Digital Twin of an Organization (DTO)* for Enterprise Modeling (EM). In short, the position paper argues that there is still a gap between EM and real-time business process execution. While data and process mining approaches may be suitable means, a lot of context information is still lost, especially if the time between process execution and model adaptation is long. Additionally, there is a certain lack of a comprehensive picture of business processes as they actually take place: process derivations may be derived from event logs but often it is less obvious what caused them. Here, cSpaces could help with its rich information about a user's work context: items, actions, etc.

For further details, readers are kindly referred to Riss et al. [290].

9.3.2 Contextual States to Improve Context-sensitive Recommendation

The basic idea of *Contextual States* is depicted in Figure 9.4. The content of a context space (i.e. its tags and items) could be identical at two different points in time, t_1 and t_2. However, at t_1, the user came back to the context after a longer break, whereas in t_2, they just finished a web search activity. In the first scenario, providing a recapitulation (e.g. often or last

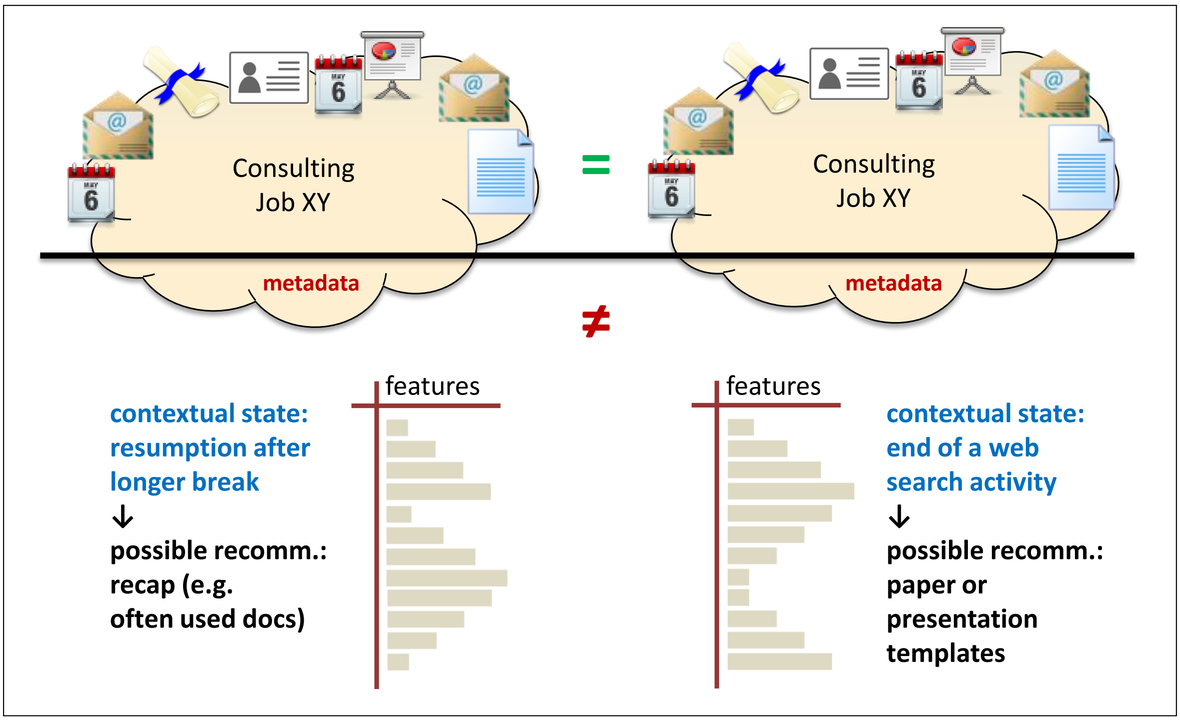


Figure 9.4: Basic Idea of Contextual States.
Abbreviations: *docs: documents / recap: recapitulation / recomm.: recommendation*.

used documents) could be a good recommendation by the system, whereas in the second scenario, showing paper or presentation templates could be helpful. One may observe that although the context space is identical on a content level in each scenario, it is different on a meta-level: Features like the time of a user's last action (no/short/longer break), the type of actions (reading/writing/...), what has been added/removed/tagged last, were there long phases of collecting items (adding) or long phases of reading (nothing added), note taking, etc. determine a *contextual state*. For different states typically different recommendations are favorable. One hypothesis is that such contextual states could improve context-sensitive recommendation. This and related topics are part of the PhD project of Mahta Bakhshizadeh, a colleague of the author. Further details can be found in Bakhshizadeh et al. [21], a position paper on the topic.

9.3.3 Inhibitory Spreading to Improve Proactive Information Delivery

The idea of *Memory Inhibition* has been introduced in Section 3.1. In short, it is the temporal suppression of currently irrelevant or misleading information in order to facilitate processing of relevant information [222]. The section also mentioned that in computer science, mechanisms closely related to activation or highlighting the most relevant items are more frequently observed than suppress "unwanted" information.

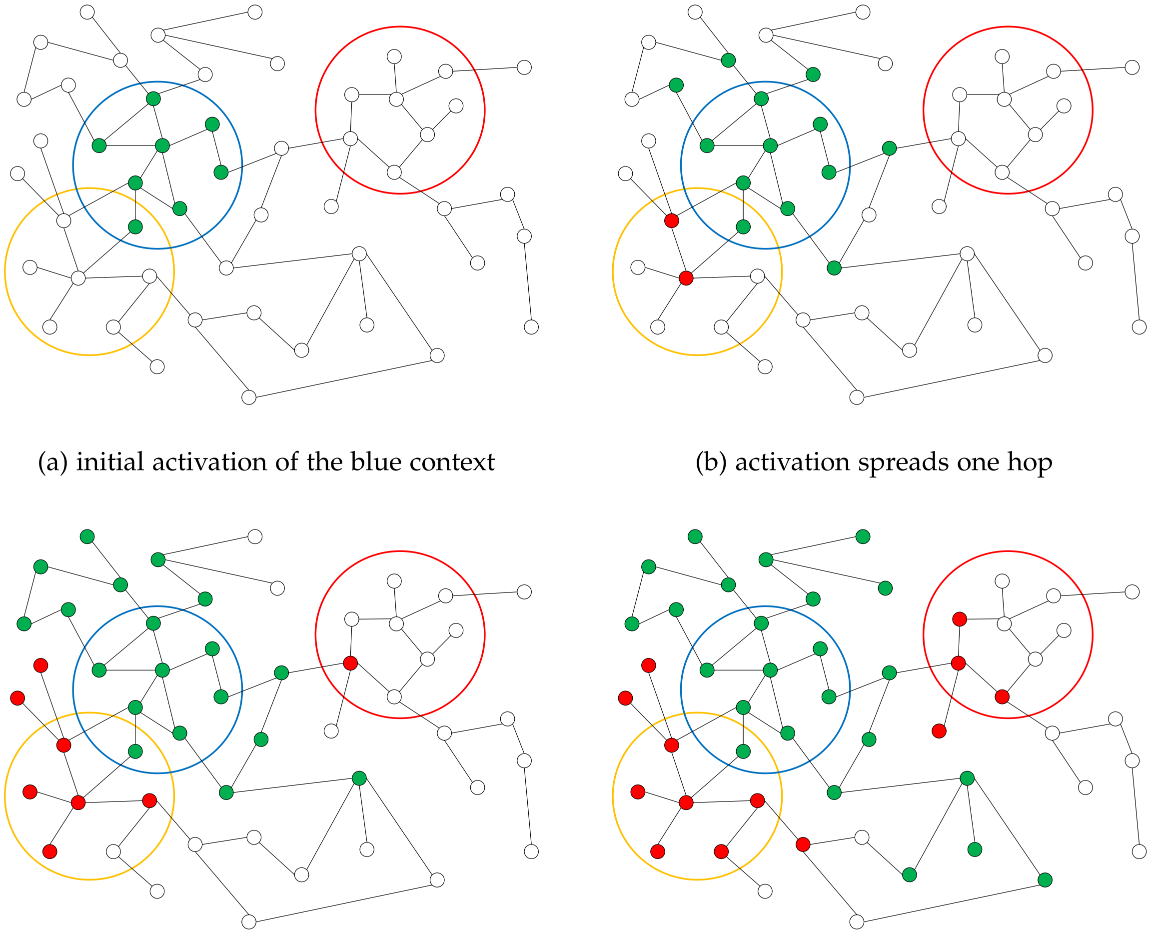
Apart from an overview and in-depth discussion of inhibition in cognitive and computer science, the position paper by Tempel et al. [373] also proposed *Inhibitory Spreading* as an example of how to enhance traditional algorithms, in this case *Spreading Activation* [70], with aspects of inhibition. Inhibitory Spreading is shown in Figure 9.5. The sub-figures 9.5a to 9.5d show how activation spreads across a network. There are three different contexts depicted as blue, green and yellow circles, which are in close neighborhood in the network. Using classic spreading activation, the red nodes would have been activated, which may be undesired since closely related but currently irrelevant contexts cause interference. Such interference is suppressed by inhibition. One could also say that inhibition applies an additional (down-)ranking of otherwise activated nodes and is thus an enhancement of the way activation is typically used in computer science. Nodes of irrelevant contexts are only activated if there is an overlap with an active one: see yellow and blue contexts.

One hypothesis to investigate in future work is that inhibitory spreading could improve proactive information delivery (PID).

9.4 Early Ideas

This section presents early ideas that still need further elaboration and for which no prototype is available, yet.

Real-time Anytime Disambiguation Exploiting cSpaces and PIM Data. In the presented IT-RTC-OBNER approach, disambiguation was out of scope. This had considerable impact on precision as shown in Figure 6.13. However, exploiting the high amount of contextual information available in cSpaces for disambiguation is an obvious idea. Disambiguation typically needs training samples, i.e. sample sentences with labeled entities – e.g. see *CoreNLP*



(a) initial activation of the blue context

(b) activation spreads one hop

(c) activation spreads a second hop

(d) activation spreads a third hop

Figure 9.5: Inhibitory Spreading.
Contexts are depicted as circles: the blue one is currently relevant, while the yellow and red one are not. Green nodes are activated, and red nodes would have been activated using classic forms of *Spreading Activation* [70], but are not since they belong to the currently irrelevant and therefore inhibited (i.e. suppressed) contexts. Nodes of these contexts are only activated if there is an overlap with an active one: see yellow and blue contexts.
(Note: This figure has already been published in Tempel et al. [373].)

[Sw113], *OpenNLP* [Sw16] or *spaCy* [Sw55]. Since cSpaces is an application to support IM and KW, one idea for future work is to exploit context (spaces) as well as available PIM data. Regarding the latter, remember statements in Section 5.3.2 about exploiting semi-structured information items like emails (sender, recipients) or calendar entries (attendees, location). Such PIM structures can serve as initial samples for a disambiguatior making the approach unsupervised. Since real-time capability (as well as inflection tolerance) is still a requirement and the disambiguator should already start its prediction after the first processed information item, the envisioned approach is a *real-time anytime disambiguator exploiting cSpaces and PIM data*. The idea is currently pursued by Desiree Heim as part of her research.

cSpaces in Purchase-to-Pay Scenarios. The author has supported research on using knowledge graphs to enhance KW support in Purchase-to-Pay[5] (P2P) [377] scenarios conducted by his colleague, Michael Schulze [338–340]. “P2P refers to the operational process that covers activities of requesting (requisitioning), purchasing, receiving, paying for and accounting for goods and services” [1].

Similar to the supSpaces scenario discussed in Section 9.2.1, cSpaces could be used to capture (working) contexts that emerge with each P2P entity, e.g. an invoice. Basically, whenever a clerk works with such an entity, the emerging context space is stored with the entity. Other clerks later working with said entity may then also access the respective context space providing them with further insights not available otherwise.

Active Notes – cSpaces Generation and Reorganization Driven by Notes. A last idea to be presented in this section is having generation and reorganization of cSpaces be driven by notes. Imagine writing a note and while writing, a context (space) is created. Information extraction is performed on the content leading to topics being associated with the context (as tags). Stating to-dos leads to sub-contexts being created. Mentioning documents or website makes cSpaces adding them to the context. Ticking to-dos or striking out a task’s line leads a “task-ish” sub-context to be marked finished, etc. Since they trigger reorganization measures, such notes are called *Active Notes* by the author.

Chapter 9 Conclusion

Part II of the thesis and especially Section 8 have shown benefits of using cSpaces (1) and that the approach has further potential (2). Chapter 9 has emphasized both findings by reporting on cSpaces ideas and assets that have already found their way into practice (ad 1) and presenting further potential and application scenarios (ad 2).

[5] or Procure-to-Pay

10 Summary, Conclusion and Outlook

As the second half of Part III, this chapter summarizes (Sec. 10.1) and concludes (Sec. 10.2) this thesis and gives an outlook on possible future work (Sec. 10.3).

10.1 Summary

The previous nine chapters have presented *Self-organizing Context Spaces (cSpaces)*, a self-(re)organizing and especially forgetting-enabled information system to support information management and knowledge work. It combines the ideas of a Semantic Desktop, Context Spaces and Managed Forgetting. Its core interaction cycle consists of eight steps or aspects, respectively.

Part I introduced into the topic and presented the general approach and vision of cSpaces as well as (general) related work. Next, in Part II, all parts and aspects of cSpaces have been introduced in more detail aligned to the steps and aspects of its interaction cycle. The first half of Part III addressed further potential and applications scenarios and the transfer into practice of several of cSpaces' ideas and assets.

This summarizing chapter, consists of two parts. First, this thesis' research questions are re-addressed compiling the findings of all previous chapters for each question individually. Second, the thesis' contributions are summarized in a compact way itemized by various dimensions.

10.1.1 Research Questions Addressed

This section compiles the findings of all previous chapters for each research question individually.

Research Question RQ1:

> **How to realize a self-(re)organizing and especially forgetting-enabled work environment that a) eases focus, b) eases keeping an overview and c) helps keeping the user's personal information sphere (more) tidied up?**

The answer to the main research question is basically the whole cSpaces approach with the eight steps/aspects of its basic interaction cycle – see Section 3 for the general approach and vision and Part II for each aspect of the approach in detail. Several studies, like the ones on cognitive offloading (Sec. 8.3.2) and faster task switching/resumption (Sec. 8.3.3), implicitly showed the achieving of the objectives a) to c). In addition, participants of the cSpaces multi-month user study were explicitly asked about them (Q_{M20}–Q_{M27}). As shown in Figure 8.27, results for the current prototype were slightly on the positive side, whereas participants' projections for the envisioned one were very much on the positive side. This led to the author's conclusion that the current prototype was already perceived helpful and that participants saw a lot of potential in features that are still in development or envisioned.

Research Question RQ2:

> **What support measures in particular could such a system provide to achieve the goals stated in Research Question RQ1 and how to realize them?**

The following cSpaces (or CoMem) support measures have been discussed:

- Context Spaces enabling users to actually **work *with* and *in* their different contexts** (Sec. 8.3)
- **Condensation and Summarization**:
  - PIMO Diary – Condensation Using Retrospectively Mined Contexts (Sec. 8.4.1)
  - Condensation by Flat Context Views (Sec. 8.4.2)
- **Temporal Reorganization, Fading Out and Resurfacing**:
  - Fading Out and Resurfacing in CoMem (Sec. 8.5.1)
  - cSpaces Context Overview (Sec. 8.5.2)
  - cSpaces Saving Last Focus (Sec. 8.5.3)
  - cSpaces Temporal Hiding and Reorganization (Sec. 8.5.4)
- **Permanent Reorganization**:
  - Preservation Prototype II: General Preservation in CoMem (Sec. 8.6.1)
  - cSpaces Automated Context Management (Sec. 8.6.2)

Research Question RQ3:

> **Are Context Spaces a suitable user interface for such a system?**

Participants of the cSpaces multi-month user study (see Sec. 8.3.4) have been asked this question (Q_{M31}) and additionally whether they could think of a better solution (Q_{M32}). They either *quite* or *strongly agreed* to the first one, while the second one was answered rather negatively. Results are depicted in Figure 8.43. Benefits shown in other studies of this PhD project implicitly support this assessment.

Research Question RQ4:

> **How to enable users to actually work *with* and *in* Context Spaces?**

Several cSpaces user interfaces have been presented – briefly in Section 3.3.3 and in more detail in Section 8.2.2. Apart from dashboards and sidebars, context injections are a novel solution leading to a near-transparent Semantic Desktop with presumably the highest user familiarity possible.

Research Question RQ5:

> **Is the Semantic Desktop a suitable ecosystem for realizing Self-organizing Context Spaces?**
> **a) What needs to be adapted or added?**
> **b) What can be learned from failed attempts (like Nepomuk-KDE) to establish the Semantic Desktop as a widespread tool in practice?**
> **c) How to bootstrap such a system, i.e. how to overcome (or at least reduce) the cold start problem of such systems?**

With its user activity tracking capabilities and integration into existing (office) applications, the Semantic Desktop is very well suited as an ecosystem for cSpaces. More details are given in Section 2.3.1, in particular mentioning other researchers that came to the same conclusion.
ad a) Context Spaces, the extension of PIMO (Sec. 3.3) and context injections (Sec. 8.2.2) are major adaptions, whereas Plug-outs and AEBI (Sec. 6.1.1) are minor (helpful but not mandatory) ones. Section 3.3.3 also compares cSpaces to traditional Semantic Desktops.
ad b) What can be learned from failed attempts like Nepomuk-KDE is addressed in Section 4.4.
ad c) An approach to bootstrap cSpaces (and KGs in general) has been presented in Section 5.3: the described process consists of three steps, each with its own tool: the *PIM Crawler*, the *Semantifier* and the *Contextifier*.

Research Question RQ6:

> **How to context-sensitively assess the value of information items and thus perform the "right" self-(re)organization measures?**

Section 7 presented the Information Value Assessment approaches *Memory Buoyancy* (short/medium-term) and *Preservation Value* (long-term). Especially for the former, an advanced context-sensitive version (MB Prototype II) has been presented (Sec. 7.2.2).

Research Question RQ7:

> **How could search interfaces for such a self-(re)organizing and especially forgetting-enabled information system look like?**

In Section 8.7, three prototypes for searching forgetting-enabled information systems (FIS) have been presented. Each of them improving the preceding one's capabilities.

Research Question RQ8:

> **Will users trust such a highly autonomous system and what can be done to facilitate trust in this scenario?**

An extensive online study on users' trust in such systems has been conducted (Sec. 8.8) showing that more explanations were not necessarily perceived as more transparency, and for different situations different levels of transparency were preferred. Explanations by an assistant like cSpaces should thus be adaptable. In addition, participants of the cSpaces multi-month user study were asked whether they would trust the envisioned system in its support decisions like hiding, condensation, reorganization etc. (Q_{M28}). As shown in Figure 8.55, except for one participant all others *quite* or *strongly agreed*. In the evaluation of FIS Search Prototype II, participants were also asked a similar question (Q_{S15}), whether would entrust an FIS with their data if they were given such search capabilities to retrieve them. Most of the participants also *quite* or *strongly agreed*.

10.1.2 Contributions

This section lists this thesis' contributions itemized by the following categories:

- **Level of maturity**: stable prototype, early prototype, pre-prototypical (i.e. code snippets), position paper or early idea.
- **Data acquisition types**: quantitative, qualitative, informal or none.
- **Duration and type of studies**: computational, long-term usage, medium-term usage, short-term usage or inquiry only.
- **Level of control**: high (laboratory conditions), medium (e.g. software-guided or experimenter present) or low (participants on their own).
- **Participant types**: dataset (e.g. given, previously collected, generated), unknown (i.e. strangers), customers, known (non-team members) or members of the team.
- **Number of participants**: the number of participants in a user study or the number of contributors to a dataset (if known).
- **Related publications**: publications by the author related to the contribution.

Additionally, if the author of this thesis has not been among the main contributors for a certain topic (for example in cases of being the supervisor of a student thesis), the actual main contributor is stated.

The contributions are listed in Tables 10.1 to 10.3.

Topic (cue)	Main contributor	Section	Level of maturity	Data acquisition	Duration and type of study	Level of control	Participants	Number	Related publications
Survey on Memory Inhibition		3.1	–	none					[373]
ATIC KG Store & Index		5.2.2	early prot.	informal	long-term usage	low	cust./kn./team	~20	
ATIC & CoMem Search		5.2.3	stable prot.	informal	long-term usage	low	cust./kn./team	~20	
PIM Crawler	M. Schröder	5.3.1	stable prot.	informal	long-term usage	low	team	~10	[325]
Semantifier	M. Schröder	5.3.2	stable prot.	informal	long-term usage	low	team	~10	[325]
Contextifier	D. Heim	5.3.3	early prot.	qual.	short-term usage	low	known/team	14	[159, 160]
Plug-outs & AEBI		6.1.1	stable prot.	informal	long-term usage	low	cust./kn./team	~15	[187, 185]
Student Software Toolkit		6.1.1	stable prot.	informal	long-term usage	low	known/team	~10	
Privacy Preservation		6.1.2	–	qual.	inquiry only		team	7	
IT-RTC-OBNER Prototype I		6.2.2	stable prot.	quant.	computational	high	dataset	~2.5M[a]	[186]
IT-RTC-OBNER Prototype II		6.2.4	stable prot.	informal	long-term usage	low	cust./kn./team	~15	
				informal	medium-term usage	low	customers	~5	
Activity Stream Pre-analyzer		6.3.1	early prot.	quant.	computational	high	dataset	5	
Activity Stream Context Miner	J. Chwalek	6.3.2	early prot.	quant.	computational	high	dataset	5	[65]
Memory Buoyancy Prototype I	S. Schwarz	7.2.1	stable prot.	quant.	computational	high	dataset	~10	[376]
				informal	long-term usage	low	team	~10	[182, 240, 188, 244]

Table 10.1: Overview of Contributions (Part 1 of 3).

Main contributor: For the sake of readability, this column is only filled if the author was not among the main contributors.

Levels of maturity: *stable prototype / early prototype / pre-prototypical (code snippets) / position paper / early idea.*

Data acquisition types: *quantitative / qualitative / informal / none.*

Duration and type of studies: *computational / long-term usage / medium-term usage / short-term usage / inquiry only.*

Levels of control: *high (laboratory conditions) / medium (e.g. software-guided or experimenter present) / low (participants on their own).*

Participant types: *dataset (e.g. given, previously collected, generated) / unknown (i.e. strangers) / customers / known (non-team members) / team.*

[a] As of November 2016 (date of downloaded dataset [Sw122]), the German Wikipedia had about 2.5 million registered users of which 18.6 thousand were considered "active" at the time (i.e. they made at least one edit in the last 30 days) [400]. The amount of 2.5M is decreased by the number of registered users that did not edit an article however, and it is increased by the amount of anonymous editors. A guess by the author is that the former number is comparably low, but the latter might be considerable. However, since both numbers are not given in the statistics, the exact number of editors is unknown to the author.

Topic (cue)	Main contributor	Section	Level of maturity	Data acquisition	Duration and type of study	Level of control	Participants	Number	Related publications
Memory Buoyancy Prototype II	J. Chwalek	7.2.2	early prot.	qual.	computational	high	dataset		[63, 182]
Memory Buoyancy Future Ideas		7.2.3	early idea	none					
Preservation Prototype I (Photos)		7.3	stable prot.	qual.	med.-term usage	med.	unknown	10	[240, 188, 245, 246, 244]
User Study Framework		8.1	stable prot.	informal	med.-term usage	low	known	~5	
Context Sidebars		8.2	stable prot.	qual.	med.-term usage	low	team	7	
PIMO (CoMem) Desktop Application		8.2	stable prot.	informal	long-term usage	low	cust./kn./team	~15	
Context Injections		8.2	early prot.	qual.	short-term usage	low	team	7	[187]
ConTextMarker Prototype I		8.3.1	early prot.	informal	short-term usage	low	un./kn./team	~110	[183, 126]
ConTextMarker Prototype II		8.3.1	early prot.	none					
ConTextMarker Prototype III		8.3.1	pre-prot.	none					
Cognitive Offloading Effects		8.3.2	early prot.	quant.	short-term usage	high	unknown	48	[126]
Cognitive Offloading fNIRS Exp.		8.3.2	early prot.	quant.	short-term usage	high	unknown	26[a]	[129][b]
Effects on Task Resumption		8.3.3	early prot.	quant.	computational	high	dataset	51	[183]
cSpaces Unified Browser Study		8.3.5	early prot.	none					
Condensation by PIMO Diary		8.4.1	stable prot.	informal	long-term usage	low	team	~10	[184, 242]
Condensation by Flat Context Views		8.4.2	pre-prot.	qual.	inquiry only		team	7	
Fading Out and Resurfacing (CoMem)	S. Schwarz	8.5.1	stable prot.	informal	long-term usage	low	team	~10	[182, 240, 244]

Table 10.2: Overview of Contributions (Part 2 of 3).

Main contributor: For the sake of readability, this column is only filled if the author was not among the main contributors.

Levels of maturity: *stable prototype / early prototype / pre-prototypical (code snippets) / position paper / early idea*.

Data acquisition types: *quantitative / qualitative / informal / none*.

Duration and type of studies: *computational / long-term usage / medium-term usage / short-term usage / inquiry only*.

Levels of control: *high (laboratory conditions) / medium (e.g. software-guided or experimenter present) / low (participants on their own)*.

Participant types: *dataset (e.g. given, previously collected, generated) / unknown (i.e. strangers) / customers / known (non-team members) / team*.

[a] Study started as a part of this thesis but could only be completed during its review process (after the submission but before the doctoral viva to be precise).

[b] Paper published during this thesis' review process (after the submission but before the doctoral viva to be precise).

Topic (cue)	Main contributor	Section	Level of maturity	Data acquisition	Duration and type of study	Level of control	Participants	Number	Related publications
cSpaces Context Overview		8.5.2	pre-prot.	qual.	inquiry only		team	7	
cSpaces Saving Last Focus		8.5.3	pre-prot.	qual.	inquiry only		team	7	
cSpaces Temporal Hiding & Reorganization		8.5.4	early prot.	qual.	inquiry only		team	7	
cSpaces Intelligent Folder Injections		8.5.4	early port.	qual.	inquiry only		team	7	
Preservation Prototype II (CoMem in General)		8.6.1	stable prot.	informal	long-term usage	low	team	~10	[240, 407, 244]
cSpaces Automated Context Management		8.6.2	pre-prot.	qual.	inquiry only		team	7	
FIS Search Prototype I	S. Schwarz	8.7.1	stable prot.	informal	long-term usage	low	team	~10	[182, 240, 244]
FIS Search Prototype II	R. Novik[a]	8.7.2	early prot.	qual.	short-term usage	med.	kn./team	7	[274]
FIS Search Prototype III		8.7.3	pre-prot.	none					
Trust in Highly Autonomous Agents	J. Knabe	8.8	–	qual.	inquiry only		unknown	140	[383, 207]
			early prot.	qual.	med.-term usage	low	team	7	
supSpaces – cSpaces in IT Support		9.2.1	early prot.	qual.	short-term usage	med.	known	1[b]	[188]
PIMO Bot – cSpaces Chat Bots		9.2.2	early prot.	informal	long-term usage	low	team	~15	
cSpaces as Part of a DTO for Enterprise Mod.	U. Riss	9.3.1	pos. paper	none					[290]
Contextual States to Improve Recommend.	M. Bakhshizadeh	9.3.2	pos. paper	none					[21]
Inhibitory Spreading to Improve PID		9.3.3	pos. paper	none					[373]
Real-time Anytime Disambiguation	D. Heim	9.4	early idea	none					

Table 10.3: Overview of Contributions (Part 3 of 3).
Main contributor: For the sake of readability, this column is only filled if the author was not among the main contributors.
Levels of maturity: *stable prototype / early prototype / pre-prototypical (code snippets) / position paper / early idea*.
Data acquisition types: *quantitative / qualitative / informal / none*.
Duration and type of studies: *computational / long-term usage / medium-term usage / short-term usage / inquiry only*.
Levels of control: *high (laboratory conditions) / medium (e.g. software-guided or experimenter present) / low (participants on their own)*.
Participant types: *dataset (e.g. given, previously collected, generated) / unknown (i.e. strangers) / customers / known (non-team members) / team*.

[a] R. Koch née Novik – using birth name here to avoid confusion with regard to Novik [274]
[b] structured interview with domain expert reviewing the prototype

10.2 Conclusion

In conclusion, a self-(re)organizing and especially forgetting-enabled support system for information management and knowledge work has been realized in this PhD project. Its different features vary in maturity: some are already in practical use (also in industry), others are in earlier prototypical states, while a last group only contains ideas, some of them well elaborated (e.g. in position papers), others only roughly sketched. Different evaluation strategies have been applied ranging from mere data-driven experiments to various user studies. Some of them were rather short-term with controlled laboratory conditions, others less controlled but spanning several months. Different benefits of working with such a system could be quantified, e.g. cognitive offloading effects and reduced task switching/resumption time. Other benefits were gathered qualitatively, e.g. tidiness of the information sphere and its better alignment with the user's mental model. The presented approach has been shown to hold a lot of potential. In some aspects, however, only first steps have been taken towards tapping it, e.g. several support measures can be further refined and automation further increased.

10.3 Outlook

Several topics for future work have already been mentioned throughout the thesis. Especially Chapter 9 presented further potential and application scenarios. One may also find future tasks for every step/aspect of the cSpaces interaction cycle, for example:

- Further enhance the ATIC KG Store, in particular with regard to large-scale industry performance (Sec. 5.2).
- Develop more plug-outs to enhance user activity tracking (see Sec. 6.1.1).
- Enhance the IT-RTC-OBNER approach with disambiguation capabilities (see Sec. 9.4).
- Improve the Stream Context Miner. In particular, explore whether its "unknown" classification can be used as an indicator for creating a new context (i.e. context spawning) (see Sec. 6.3.2).
- Further optimize Memory Buoyancy, especially with regard to Big Personal Data and corporate scenarios (see Sec. 7.2.3).
- Conduct the postponed cSpaces Unified Browser Study (see Sec. 8.3.5).
- Fully develop those support measures that are still in an early state or envisioned by the end of this PhD project (see the respective sections of Chapter 8).

From the author's perspective, the last item is the most important one. Starting a second phase of the multi-month user study with a richer set of support measures is desirable. In addition, transferring the whole cSpaces idea to corporate scenarios, in which too extensive user activity tracking is not possible, would also be an interesting challenge. Allowing users to customize the amount of tracking but still provide beneficial support would be a goal in this regard. A solution could be that users subtly model certain aspects instead of being tracked. Dropping an item into a certain context was such an example showing that a single click or drag may already mean quite a lot for the system (see explanation of second item of *Motivation behind Context Spaces* in Sec. 3.2).

Part IV

Appendices

A

Additional Details on Experiments and Studies

Mainly for the sake of completeness, this chapter contains additional details on conducted experiments and studies.

A.1 Used Personal Information Management Ontology

This section complements Section 2.3.2 about the *Personal Information Model (PIMO)*.

In all cSpaces experiments, a general Personal Information Management (PIM) ontology was used. Thus, classes like persons, organizations, projects, events, tasks, topics, etc. were available for users to express relations on their data. A user's PIMO was either bootstrapped with their own data (see Sec. 5.3) or pre-filled with a few instances that seemed useful and obvious for a given scenario (e.g. food, beverage or locations for planning a barbecue party). The used PIM ontology was basically an exported excerpt [Sw110] from DFKI SDS' current CoMem [87] prototype productively used in daily work (see Sec. 2.3.3). It emerged from the *PIMO ontology* [305–307] envisioned and realized in the EPOS [110] and NEPOMUK [267] projects (see Sec. 4.1). After the export, some cSpaces-related extensions were made by the author, especially with respect to the modeling of contexts – see "`pimo:context`" properties in particular (Table A.4).

All resources and classes, literal and object properties as well as sub and inverse property relations are listed in Tables A.1 to A.8. Used prefixes are as follows (in Turtle[1] [23] syntax):

```
@prefix rdf:  <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix xsd:  <http://www.w3.org/2001/XMLSchema#>.
```

Note that strings starting with "`pimo:`" are actually not in prefix notation but are used as URIs in CoMem exactly as they are (although not complying with URI syntax rules specified in Berners-Lee et al. [34, 35]). cSpaces nevertheless uses the same URIs in order to stay compatible to CoMem. Additionally, for the sake of reduced table widths, the author has replaced all occurrences of the string "`informationElement`" with "`infoElem`" in "`pimo:`" URIs.

[1] Terse RDF Triple Language [23]

resource/class	**label**	**subclass of**
rdf:Alt	Alt	rdfs:Container
rdf:Bag	Bag	rdfs:Container
rdf:first	first	
rdf:List	List	rdfs:Resource
rdf:nil	nil	
rdf:object	object	
rdf:predicate	predicate	
rdf:Property	Property	rdfs:Resource
rdf:rest	rest	
rdf:Seq	Seq	rdfs:Container
rdf:Statement	Statement	rdfs:Resource
rdf:subject	subject	
rdf:type	type	
rdf:value	value	
rdf:XMLLiteral	XMLLiteral	rdfs:Literal
rdfs:Class	Class	rdfs:Resource
rdfs:comment	comment	
rdfs:Container	Container	rdfs:Resource
rdfs:ContainerMembershipProperty	ContainerMembershipProperty	rdf:Property
rdfs:Datatype	Datatype	rdfs:Class
rdfs:domain	domain	
rdfs:isDefinedBy	isDefinedBy	
rdfs:label	label	
rdfs:Literal	Literal	rdfs:Resource
rdfs:member	member	
rdfs:range	range	
rdfs:Resource	Resource	
rdfs:seeAlso	seeAlso	
rdfs:subClassOf	subClassOf	
rdfs:subPropertyOf	subPropertyOf	
xsd:anyURI	XMLSchema:anyURI	rdfs:Literal
xsd:boolean	XMLSchema:boolean	rdfs:Literal
xsd:date	XMLSchema:date	rdfs:Literal
xsd:dateTime	XMLSchema:dateTime	rdfs:Literal
xsd:decimal	XMLSchema:decimal	rdfs:Literal
xsd:double	XMLSchema:double	rdfs:Literal
xsd:duration	XMLSchema:duration	rdfs:Literal
xsd:float	XMLSchema:float	rdfs:Literal
xsd:int	XMLSchema:int	rdfs:Literal
xsd:integer	XMLSchema:integer	rdfs:Literal
xsd:string	XMLSchema:string	rdfs:Literal
pimo:context#Context	Context	pimo:thing#Collection
pimo:core#Address	Address	rdfs:Resource
pimo:core#EmailAddress	Email Address	pimo:core#Address
pimo:core#Ontology	Ontology	rdfs:Resource
pimo:core#prefLabel	nao:prefLabel	

Table A.1: Used PIM Ontology: Resources and Classes (Part 1 of 2).

resource/class	label	subclass of
pimo:gis#GeoLocation	GeoLocation	pimo:core#Address
pimo:gis#GeoLocationWithRadius	GeoLocationWithRadius	pimo:gis#GeoLocation
pimo:infoElem#Email	Email	pimo:infoElem#InfoElem
pimo:infoElem#File	File	pimo:infoElem#InfoElem
pimo:infoElem#Folder	Folder	pimo:infoElem#InfoElem
pimo:infoElem#InfoElem	Information Element	rdfs:Resource
pimo:infoElem#SharedspaceFile	SharedspaceFile	pimo:infoElem#InfoElem
pimo:infoElem#Webpage	Webpage	pimo:infoElem#WebResource
pimo:infoElem#WebResource	WebResource	pimo:infoElem#InfoElem
pimo:stats#UsageStats	UsageStats	rdfs:Resource
pimo:task#Task	Task	pimo:thing#Job
pimo:thing#Agent	Agent	pimo:thing#Thing
pimo:thing#Building	Building	pimo:thing#Location
pimo:thing#City	City	pimo:thing#Location
pimo:thing#Collection	Collection	pimo:thing#Thing
pimo:thing#Conference	Conference	pimo:thing#SocialEvent
pimo:thing#Contract	Contract	pimo:thing#Document
pimo:thing#Country	Country	pimo:thing#Location
pimo:thing#Document	Document	pimo:thing#Media
pimo:thing#Email	Email	pimo:thing#Document pimo:thing#Message
pimo:thing#Event	Event	pimo:thing#Thing
pimo:thing#Group	Group	pimo:thing#Agent
pimo:thing#Image	Image	pimo:thing#Media
pimo:thing#Job	Job	pimo:thing#Thing
pimo:thing#Location	Location	pimo:thing#Thing
pimo:thing#Media	Media	pimo:thing#Thing
pimo:thing#Meeting	Meeting	pimo:thing#SocialEvent
pimo:thing#Message	Message	pimo:thing#Thing
pimo:thing#Note	Note	pimo:thing#Document
pimo:thing#Organization	Organization	pimo:thing#Group
pimo:thing#Person	Person	pimo:thing#Agent
pimo:thing#Project	Project	pimo:thing#Job
pimo:thing#Room	Room	pimo:thing#Location
pimo:thing#SocialEvent	SocialEvent	pimo:thing#Event
pimo:thing#Software	Software	pimo:thing#Tool
pimo:thing#State	State	pimo:thing#Location
pimo:thing#Thing	Thing	rdfs:Resource
pimo:thing#Tool	Tool	pimo:thing#Thing
pimo:thing#Topic	Topic	pimo:thing#Thing
pimo:thing#Video	Video	pimo:thing#Media
pimo:thing#Webpage	Webpage	pimo:thing#Document
pimo:thing#Website	Website	pimo:thing#Media

Table A.2: Used PIM Ontology: Resources and Classes (Part 2 of 2).

property	domain	range
rdfs:comment	rdfs:Resource	rdfs:Literal
rdfs:label	rdfs:Resource	rdfs:Literal
pimo:core#associatedDate	rdfs:Resource	rdfs:Literal
pimo:core#createdBy	rdfs:Resource	rdfs:Literal
pimo:core#hasPrefImageSharedspaceFileAsLiteral	rdfs:Resource	rdfs:Literal
pimo:core#hasVersion	pimo:core#Ontology	rdfs:Literal
pimo:core#modificationTime	rdfs:Resource	rdfs:Literal
pimo:core#modifiedBy	rdfs:Resource	rdfs:Literal
pimo:core#prefLabel	rdfs:Resource	rdfs:Literal
pimo:core#propagates	rdf:Property	rdfs:Literal
pimo:gis#point	pimo:gis#GeoLocation	rdfs:Literal
pimo:gis#radius	pimo:gis#GeoLocationWithRadius	rdfs:Literal
pimo:infoElem#creationDate	pimo:infoElem#InfoElem	rdfs:Literal
pimo:infoElem#emailSubject	pimo:infoElem#Email	rdfs:Literal
pimo:infoElem#fileExtension	pimo:infoElem#File	rdfs:Literal
pimo:infoElem#sentDate	pimo:infoElem#Email	xsd:dateTime
pimo:infoElem#webpageTitle	pimo:infoElem#Webpage	rdfs:Literal
pimo:stats#stackPosition	pimo:stats#UsageStats	rdfs:Literal
pimo:stats#usageInFiltersCount	pimo:stats#UsageStats	rdfs:Literal
pimo:task#activeDuration	pimo:task#Task	rdfs:Literal
pimo:task#activeInterval	pimo:task#Task	rdfs:Literal
pimo:task#hasPriority	pimo:task#Task	rdfs:Literal
pimo:task#hasTaskType	pimo:task#Task	rdfs:Literal
pimo:task#hotness	pimo:stats#UsageStats	rdfs:Literal
pimo:task#isFinished	pimo:task#Task	rdfs:Literal
pimo:task#isHiddenInCalendar	pimo:thing#Event	rdfs:Literal
pimo:task#isUnfinishedLeafTask	pimo:task#Task	rdfs:Literal
pimo:task#lastSeen	pimo:task#Task	rdfs:Literal
pimo:task#mustBeFinishedByMillis	pimo:task#Task	rdfs:Literal
pimo:task#recentlyActiveInterval	pimo:task#Task	rdfs:Literal
pimo:task#remindedUserLastTime	pimo:task#Task	rdfs:Literal
pimo:task#remindsUserAtMillis	pimo:task#Task	rdfs:Literal
pimo:task#stackPosition	pimo:stats#UsageStats	rdfs:Literal
pimo:task#willNotStartBeforeMillis	pimo:task#Task	rdfs:Literal
pimo:thing#acronym	pimo:thing#Thing	rdfs:Literal
pimo:thing#associatedDate	pimo:thing#Media	rdfs:Literal
pimo:thing#copiesLabelFromOccurrence	pimo:thing#Thing	rdfs:Literal
pimo:thing#endsAtMillis	pimo:thing#Event	rdfs:Literal
pimo:thing#firstName	pimo:thing#Person	xsd:string
pimo:thing#height	pimo:thing#Image	rdfs:Literal
pimo:thing#imageAnalysis_suggestions	pimo:thing#Image	rdfs:Literal
pimo:thing#imageAnalysis_xml	pimo:thing#Image	rdfs:Literal
pimo:thing#imagequality_score	pimo:thing#Image	rdfs:Literal
pimo:thing#isPhotoTakeAtDate	pimo:thing#Image	rdfs:Literal
pimo:thing#lastName	pimo:thing#Person	xsd:string
pimo:thing#phoneNumber	pimo:thing#Agent	rdfs:Literal
pimo:thing#plainTextContent	rdfs:Resource	rdfs:Literal
pimo:thing#roomNumber	pimo:thing#Room	rdfs:Literal
pimo:thing#sentDate	pimo:thing#Message	xsd:dateTime
pimo:thing#skypeName	pimo:thing#Person	rdfs:Literal
pimo:thing#startsAtMillis	pimo:thing#Event	rdfs:Literal
pimo:thing#width	pimo:thing#Image	rdfs:Literal
pimo:thing#wikiText	pimo:thing#Thing	rdfs:Literal
pimo:thing#wikiTextState	pimo:thing#Thing	rdfs:Literal

Table A.3: Used PIM Ontology: Literal Properties.

property	domain	range
rdf:first	rdf:List	rdfs:Resource
rdf:object	rdf:Statement	rdfs:Resource
rdf:predicate	rdf:Statement	rdfs:Resource
rdf:rest	rdf:List	rdf:List
rdf:subject	rdf:Statement	rdfs:Resource
rdf:type	rdfs:Resource	rdfs:Class
rdf:value	rdfs:Resource	rdfs:Resource
rdfs:domain	rdf:Property	rdfs:Class
rdfs:isDefinedBy	rdfs:Resource	rdfs:Resource
rdfs:member	rdfs:Resource	rdfs:Resource
rdfs:range	rdf:Property	rdfs:Class
rdfs:seeAlso	rdfs:Resource	rdfs:Resource
rdfs:subClassOf	rdfs:Class	rdfs:Class
rdfs:subPropertyOf	rdf:Property	rdf:Property
pimo:context#characterizes	pimo:thing#Thing	pimo:context#Context
pimo:context#contains	pimo:context#Context	rdfs:Resource
pimo:context#hasNucleus	pimo:context#Context	pimo:thing#Thing
pimo:context#hasSelectedContext	pimo:thing#Person	pimo:context#Context
pimo:context#hasSubContext	pimo:context#Context	pimo:context#Context
pimo:context#hasSuperContext	pimo:context#Context	pimo:context#Context
pimo:context#hasVisibleContextCollection	pimo:thing#Person	pimo:thing#Collection
pimo:context#isAbout	pimo:context#Context	pimo:thing#Thing
pimo:context#isContainedIn	rdfs:Resource	pimo:context#Context
pimo:context#isNucleusOf	pimo:thing#Thing	pimo:context#Context
pimo:context#isSelectedContextOf	pimo:context#Context	pimo:thing#Person
pimo:context#isSubContextOf	pimo:context#Context	pimo:context#Context
pimo:context#isSuperContextOf	pimo:context#Context	pimo:context#Context
pimo:context#isVisibleContextCollectionOf	pimo:thing#Collection	pimo:thing#Person
pimo:core#inverseProperty	rdf:Property	rdf:Property
pimo:gis#hasGeoLocation	pimo:thing#Thing	pimo:gis#GeoLocation
pimo:gis#isGeoLocationOf	pimo:gis#GeoLocation	pimo:thing#Thing
pimo:infoElem#hasBccRecipient	pimo:infoElem#Email	pimo:core#EmailAddress
pimo:infoElem#hasCcRecipient	pimo:infoElem#Email	pimo:core#EmailAddress
pimo:infoElem#hasRecipient	pimo:infoElem#Email	pimo:core#EmailAddress
pimo:infoElem#hasSender	pimo:infoElem#Email	pimo:core#EmailAddress
pimo:infoElem#isBccRecipientOf	pimo:core#EmailAddress	pimo:infoElem#Email
pimo:infoElem#isCcRecipientOf	pimo:core#EmailAddress	pimo:infoElem#Email
pimo:infoElem#isRecipientOf	pimo:core#EmailAddress	pimo:infoElem#Email
pimo:infoElem#isSenderOf	pimo:core#EmailAddress	pimo:infoElem#Email
pimo:stats#hasUsageStats	pimo:thing#Thing	pimo:stats#UsageStats
pimo:stats#isUsageStatsOf	pimo:stats#UsageStats	pimo:thing#Thing
pimo:task#displayedAfter	pimo:task#Task	pimo:task#Task
pimo:task#displayedBefore	pimo:task#Task	pimo:task#Task
pimo:task#hasRootTask	pimo:task#Task	pimo:task#Task
pimo:task#hasSubTask	pimo:task#Task	pimo:task#Task
pimo:task#hasSuperTask	pimo:task#Task	pimo:task#Task
pimo:task#isDueOfTask	pimo:thing#Event	pimo:task#Task
pimo:task#isStartOfTask	pimo:thing#Event	pimo:task#Task
pimo:task#mustBeFinishedBy	pimo:task#Task	pimo:thing#Event
pimo:task#remindsUserAt	pimo:task#Task	pimo:thing#Event
pimo:task#remindsUserOf	pimo:thing#Event	pimo:task#Task
pimo:task#willNotStartBefore	pimo:task#Task	pimo:thing#Event

Table A.4: Used PIM Ontology: Resource/Object Properties (Part 1 of 3).

property	domain	range
pimo:thing#attends	pimo:thing#Agent	pimo:thing#SocialEvent
pimo:thing#containsThing	pimo:thing#Collection	pimo:thing#Thing
pimo:thing#creates	pimo:thing#Agent	pimo:thing#Media
pimo:thing#depicts	pimo:thing#Image	pimo:thing#Thing
pimo:thing#does	pimo:thing#Agent	pimo:thing#Job
pimo:thing#funds	pimo:thing#Agent	pimo:thing#Project
pimo:thing#hasAcceptedTopic	pimo:thing#Thing	pimo:thing#Thing
pimo:thing#hasAddress	pimo:thing#Agent	pimo:core#Address
pimo:thing#hasAggregatedTopic	pimo:thing#Thing	pimo:thing#Thing
pimo:thing#hasAttendee	pimo:thing#SocialEvent	pimo:thing#Agent
pimo:thing#hasCachedOccurrence	pimo:thing#Thing	pimo:infoElem#InfoElem
pimo:thing#hasDeclinedTopic	pimo:thing#Thing	pimo:thing#Thing
pimo:thing#hasEmailAddress	pimo:thing#Agent	pimo:core#EmailAddress
pimo:thing#hasGroundingOccurrence	pimo:thing#Thing	pimo:infoElem#InfoElem
pimo:thing#hasHomepage	pimo:thing#Website	pimo:infoElem#Webpage
pimo:thing#hasImage	rdfs:Resource	pimo:thing#Image
pimo:thing#hasInferredTopic	pimo:thing#Thing	pimo:thing#Thing
pimo:thing#hasLocation	pimo:thing#Thing	pimo:thing#Location
pimo:thing#hasManager	pimo:thing#Group	pimo:thing#Person
pimo:thing#hasMember	pimo:thing#Group	pimo:thing#Agent
pimo:thing#hasNote	rdfs:Resource	pimo:thing#Note
pimo:thing#hasOrganizer	pimo:thing#SocialEvent	pimo:thing#Agent
pimo:thing#hasPart	pimo:thing#Thing	pimo:thing#Thing
pimo:thing#hasPhotoTakenAtLocation	pimo:thing#Location	pimo:thing#Image
pimo:thing#hasPrefImage	rdfs:Resource	pimo:thing#Image
pimo:thing#hasSubcollection	pimo:thing#Collection	pimo:thing#Collection
pimo:thing#hasSubLocation	pimo:thing#Location	pimo:thing#Location
pimo:thing#hasSubTopic	pimo:thing#Topic	pimo:thing#Topic
pimo:thing#hasSuggestedTopic	pimo:thing#Thing	pimo:thing#Thing
pimo:thing#hasSuperTopic	pimo:thing#Topic	pimo:thing#Topic
pimo:thing#hasTopic	pimo:thing#Thing	pimo:thing#Thing
pimo:thing#hasWebsite	pimo:thing#Agent	pimo:thing#Website
pimo:thing#isAcceptedTopicOf	pimo:thing#Thing	pimo:thing#Thing
pimo:thing#isAddressOf	pimo:core#Address	pimo:thing#Agent
pimo:thing#isAggregatedTopicOf	pimo:thing#Thing	pimo:thing#Thing
pimo:thing#isCachedOccurrenceOf	pimo:infoElem#InfoElem	pimo:thing#Thing
pimo:thing#isContainedIn	pimo:thing#Thing	pimo:thing#Collection
pimo:thing#isCreatedBy	pimo:thing#Media	pimo:thing#Agent
pimo:thing#isDeclinedTopicOf	pimo:thing#Thing	pimo:thing#Thing
pimo:thing#isDepictedIn	pimo:thing#Thing	pimo:thing#Image
pimo:thing#isDoneBy	pimo:thing#Job	pimo:thing#Agent
pimo:thing#isEmailAddressOf	pimo:core#EmailAddress	pimo:thing#Agent
pimo:thing#isFundedBy	pimo:thing#Project	pimo:thing#Agent
pimo:thing#isGroundingOccurrenceOf	pimo:infoElem#InfoElem	pimo:thing#Thing
pimo:thing#isHomepageOf	pimo:infoElem#Webpage	pimo:thing#Website
pimo:thing#isImageOf	pimo:thing#Image	rdfs:Resource
pimo:thing#isInferredTopicOf	pimo:thing#Thing	pimo:thing#Thing
pimo:thing#isLocationOf	pimo:thing#Location	pimo:thing#Thing
pimo:thing#isManagedBy	pimo:thing#Job	pimo:thing#Person
pimo:thing#isManagerOf	pimo:thing#Person	pimo:thing#Group
pimo:thing#isMemberOf	pimo:thing#Agent	pimo:thing#Group
pimo:thing#isMentionedInWikiText	pimo:thing#Thing	pimo:thing#Thing
pimo:thing#isNoteOf	pimo:thing#Note	rdfs:Resource
pimo:thing#isPartOf	pimo:thing#Thing	pimo:thing#Thing
pimo:thing#isPhotoTakenAtLocation	pimo:thing#Image	pimo:thing#Location

Table A.5: Used PIM Ontology: Resource/Object Properties (Part 2 of 3).

property	domain	range
pimo:thing#isPrefImageOf	pimo:thing#Image	rdfs:Resource
pimo:thing#isReceivedBy	pimo:thing#Message	pimo:thing#Agent
pimo:thing#isRelated	pimo:thing#Thing	pimo:thing#Thing
pimo:thing#isSentBy	pimo:thing#Message	pimo:thing#Agent
pimo:thing#isSubcollectionOf	pimo:thing#Collection	pimo:thing#Collection
pimo:thing#isSubLocationOf	pimo:thing#Location	pimo:thing#Location
pimo:thing#isSuggestedTopicOf	pimo:thing#Thing	pimo:thing#Thing
pimo:thing#isTopicOf	pimo:thing#Thing	pimo:thing#Thing
pimo:thing#isWebsiteOf	pimo:thing#Website	pimo:thing#Agent
pimo:thing#manages	pimo:thing#Person	pimo:thing#Job
pimo:thing#organizes	pimo:thing#Agent	pimo:thing#SocialEvent
pimo:thing#receives	pimo:thing#Agent	pimo:thing#Message
pimo:thing#sends	pimo:thing#Agent	pimo:thing#Message
pimo:thing#wikiTextMentions	pimo:thing#Thing	pimo:thing#Thing

Table A.6: Used PIM Ontology: Resource/Object Properties (Part 3 of 3).

property	subproperty of
rdfs:isDefinedBy	rdfs:seeAlso
pimo:core#prefLabel	rdfs:label
pimo:task#hasSubTask	pimo:thing#hasPart
pimo:task#hasSuperTask	pimo:thing#isPartOf
pimo:thing#attends	pimo:thing#isPartOf
pimo:thing#containsThing	pimo:thing#hasPart
pimo:thing#hasAttendee	pimo:thing#hasPart
pimo:thing#hasEmailAddress	pimo:thing#hasAddress
pimo:thing#hasHomepage	pimo:thing#hasGroundingOccurrence
pimo:thing#hasManager	pimo:thing#hasMember
pimo:thing#hasMember	pimo:thing#hasPart
pimo:thing#hasSubLocation	pimo:thing#hasPart
pimo:thing#isContainedIn	pimo:thing#isPartOf
pimo:thing#isEmailAddressOf	pimo:thing#isAddressOf
pimo:thing#isHomepageOf	pimo:thing#isGroundingOccurrenceOf
pimo:thing#isManagedBy	pimo:thing#isDoneBy
pimo:thing#isManagerOf	pimo:thing#isMemberOf
pimo:thing#isMemberOf	pimo:thing#isPartOf
pimo:thing#isSubLocationOf	pimo:thing#isPartOf
pimo:thing#manages	pimo:thing#does

Table A.7: Used PIM Ontology: Sub-property Relations.

property	inverse property
pimo:context#contains	pimo:context#isContainedIn
pimo:context#hasNucleus	pimo:context#isNucleusOf
pimo:context#hasSelectedContext	pimo:context#isSelectedContextOf
pimo:context#hasSubContext	pimo:context#isSubContextOf
pimo:context#hasSuperContext	pimo:context#isSuperContextOf
pimo:context#isAbout	pimo:context#characterizes
pimo:context#isVisibleContextCollectionOf	pimo:context#hasVisibleContextCollection
pimo:core#inverseProperty	pimo:core#inverseProperty
pimo:gis#isGeoLocationOf	pimo:gis#hasGeoLocation
pimo:infoElem#hasBccRecipient	pimo:infoElem#isBccRecipientOf
pimo:infoElem#isCcRecipientOf	pimo:infoElem#hasCcRecipient
pimo:infoElem#isRecipientOf	pimo:infoElem#hasRecipient
pimo:infoElem#isSenderOf	pimo:infoElem#hasSender
pimo:stats#isUsageStatsOf	pimo:stats#hasUsageStats
pimo:task#displayedAfter	pimo:task#displayedBefore
pimo:task#hasSubTask	pimo:task#hasSuperTask
pimo:task#isDueOfTask	pimo:task#mustBeFinishedBy
pimo:task#remindsUserAt	pimo:task#remindsUserOf
pimo:task#willNotStartBefore	pimo:task#isStartOfTask
pimo:thing#creates	pimo:thing#isCreatedBy
pimo:thing#depicts	pimo:thing#isDepictedIn
pimo:thing#funds	pimo:thing#isFundedBy
pimo:thing#hasAcceptedTopic	pimo:thing#isAcceptedTopicOf
pimo:thing#hasAggregatedTopic	pimo:thing#isAggregatedTopicOf
pimo:thing#hasAttendee	pimo:thing#attends
pimo:thing#hasCachedOccurrence	pimo:thing#isCachedOccurrenceOf
pimo:thing#hasHomepage	pimo:thing#isHomepageOf
pimo:thing#hasManager	pimo:thing#isManagerOf
pimo:thing#hasNote	pimo:thing#isNoteOf
pimo:thing#hasPart	pimo:thing#isPartOf
pimo:thing#hasPhotoTakenAtLocation	pimo:thing#isPhotoTakenAtLocation
pimo:thing#hasSubcollection	pimo:thing#isSubcollectionOf
pimo:thing#hasSuperTopic	pimo:thing#hasSubTopic
pimo:thing#hasTopic	pimo:thing#isTopicOf
pimo:thing#isAddressOf	pimo:thing#hasAddress
pimo:thing#isContainedIn	pimo:thing#containsThing
pimo:thing#isDeclinedTopicOf	pimo:thing#hasDeclinedTopic
pimo:thing#isDoneBy	pimo:thing#does
pimo:thing#isEmailAddressOf	pimo:thing#hasEmailAddress
pimo:thing#isGroundingOccurrenceOf	pimo:thing#hasGroundingOccurrence
pimo:thing#isImageOf	pimo:thing#hasImage
pimo:thing#isInferredTopicOf	pimo:thing#hasInferredTopic
pimo:thing#isLocationOf	pimo:thing#hasLocation
pimo:thing#isMemberOf	pimo:thing#hasMember
pimo:thing#isPrefImageOf	pimo:thing#hasPrefImage
pimo:thing#isRelated	pimo:thing#isRelated
pimo:thing#isSubLocationOf	pimo:thing#hasSubLocation
pimo:thing#isSuggestedTopicOf	pimo:thing#hasSuggestedTopic
pimo:thing#isWebsiteOf	pimo:thing#hasWebsite
pimo:thing#manages	pimo:thing#isManagedBy
pimo:thing#organizes	pimo:thing#hasOrganizer
pimo:thing#receives	pimo:thing#isReceivedBy
pimo:thing#sends	pimo:thing#isSentBy
pimo:thing#wikiTextMentions	pimo:thing#isMentionedInWikiText

Table A.8: Used PIM Ontology: Inverse Property Relations.

A.2 Additional Details on Contextifier

This section contains an additional screenshot and survey results complementing the main section on the *Contextifier* (Sec. 5.3.3).

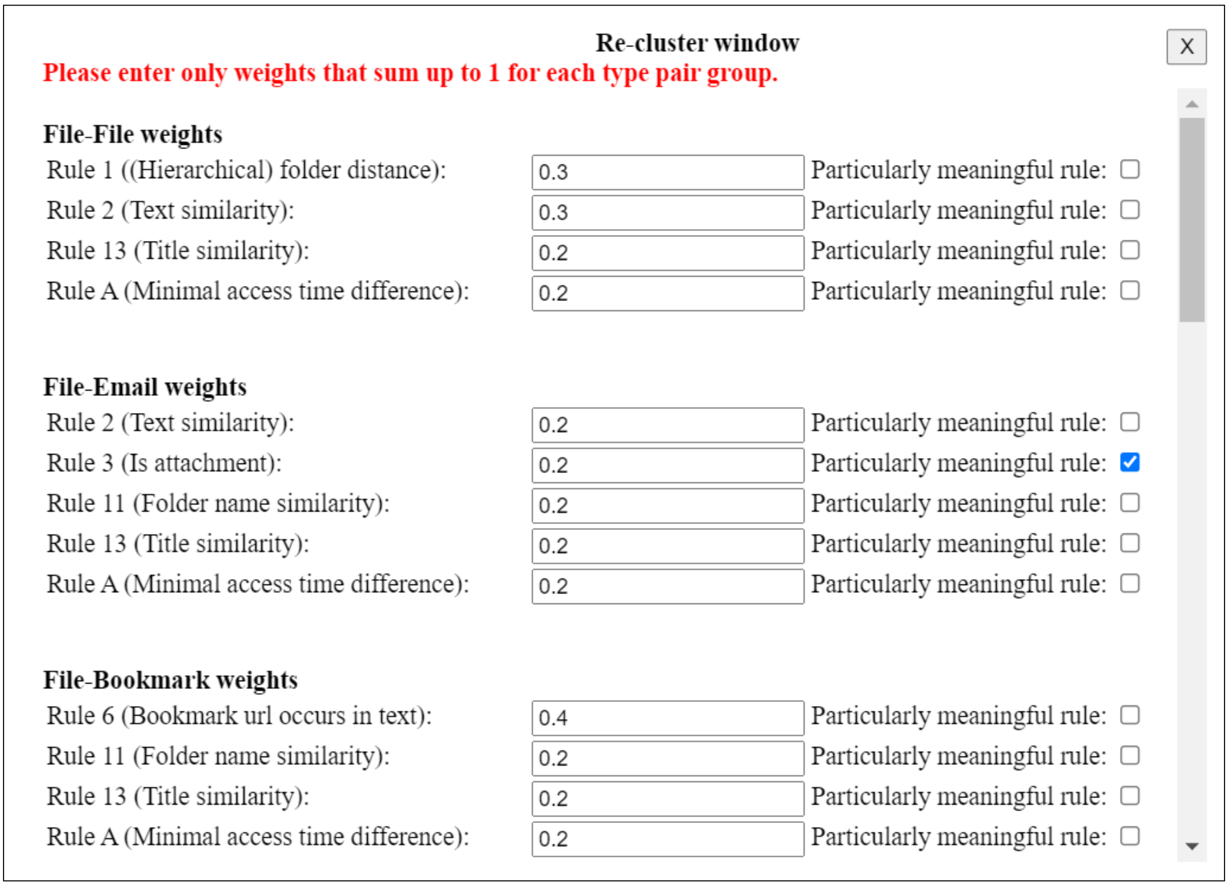


Figure A.1: Contextifier: GUI to Adjust Rule Weights Realized in Heim [159].
Note: A *particularly meaningful rule* basically boosts two items' relationship confidence.
(Image source: Heim [159], whitespace compacted for better readability.)

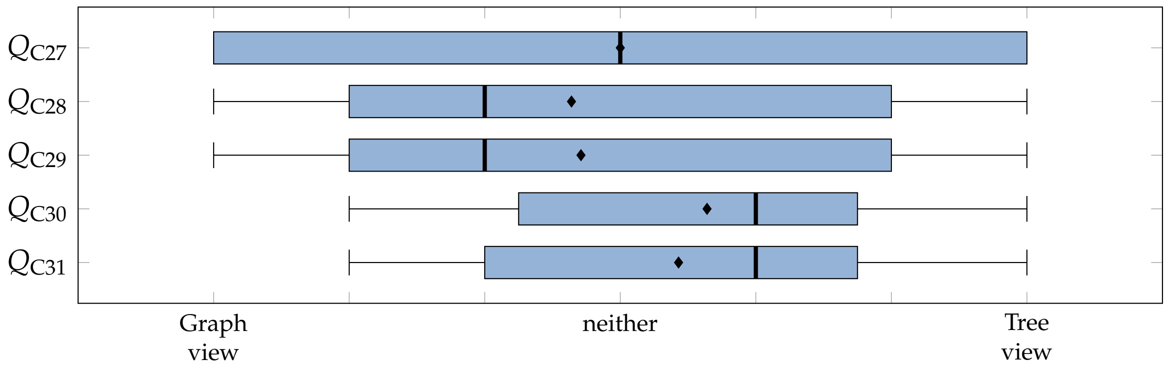


Figure A.2: Contextifier: Survey Result for Comparison of Graph View and Tree View (see Fig. 5.8) as Reported in Heim [159].

Q_{C27}: Which view do you like better overall?
Q_{C28}: Using context modification operations is easier in this view.
Q_{C29}: Navigation through the different contexts and documents is easier in this view.
Q_{C30}: This view displays the results more comprehensible.
Q_{C31}: This view has higher information content.

(n=14; details on box plot semantics in Sec. B.2.)

A.3 Additional Details on FIS Search Prototype II

This section contains an additional screenshot complementing the main section on the *FIS Search Prototype II* (Sec. 8.7.2).

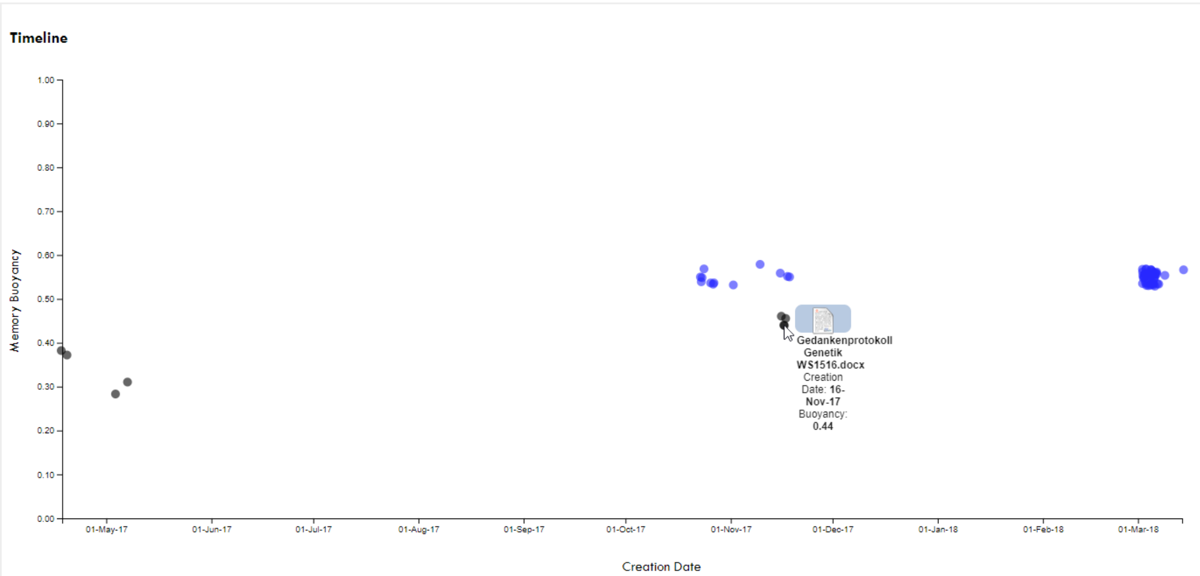


Figure A.3: FIS Search Prototype II: Timeline View as Realized in Novik [274].
(Image source: Novik [274], whitespace compacted for better readability.)

A.4 Additional Details on Multi-month User Study

The final interview of the cSpaces multi-month user study consisted of 45 items. 43 of the them asked for assessments on a 7-point Likert scale [226]. Two items allowed for free text feedback: Q_{M17} asked participants to describe to what extent they integrated cSpaces into their daily work. Q_{M45} was the questionnaire's last item and asked for arbitrary feedback regarding the study. The field also offered the possibility to give feedback to any previous item, which was used by many participants to justify or explain their assessments or give hints at possible future improvements. In total, verbal feedback amounted to 5,521 words. For privacy and non-disclosure reasons, the feedback is not printed here. However, the author strived for incorporating it into the interpretation and summarization of the results also including a few direct quotations.

The study is mainly addressed in Section 8.3.4, whereas items are split according to topics and discussed in corresponding sections. An overview of where to find each item is given in Table A.9.

The remainder of this section presents the full document of the final interview questionnaire.

Items	Figure	Topic
Q_{M1}–Q_{M3}	8.14	Contextual Notes
Q_{M4}	8.37	Context Overview
Q_{M5}	8.28	cSpaces GUI Switches
Q_{M6}–Q_{M7}	8.38	Last Focus
Q_{M8}	8.33	Flat Context Views
Q_{M9}–Q_{M10}	8.42	Intelligent Folder Injections
Q_{M11}–Q_{M14}	8.46	Automated Context Management
Q_{M15}–Q_{M16}	6.3	Privacy Preservation
Q_{M17}		(free text item)
Q_{M18}–Q_{M19}	8.26	Idea of Context Spaces
Q_{M20}–Q_{M27}	8.27	Working with and in Context Spaces
Q_{M28}–Q_{M29}	8.55	Trust and Explanations
Q_{M30}–Q_{M32}	8.43	Self-organizing Context Spaces
Q_{M33}–Q_{M44}	8.29	Usability of cSpaces
Q_{M45}		(free text item)

Table A.9: Multi-month User Study: Mapping of Items to Figures.

cSpaces User Study – Final Interview 2021

Dear participant,

thank you very much for taking part in the cSpaces user study so far. This interview concludes this year's part of the study. As you may know, the study served two purposes: first, I needed to evaluate several features of my system. Second, I needed to get deeper insights (data) that will help me in developing further features.

You surely realized that not everything we developed for cSpaces over the years found its way into the current prototype. There are several reasons for this: for some aspects we already had completed one or even two dedicated evaluations, so we prioritized other features over them. And as the year 2021 was slowly coming to an end, we postponed some features to 2022 in order to avoid introducing breaking bugs by enabling unfinished features in a hurry. Nevertheless, some of these features are important when talking about the overall vision of the system. For the rest of the interview I will therefore refer to both: the **current prototype** and the **envisioned prototype**. Please read later questions carefully, since both prototypes may explicitly be addressed – sometimes the same question will even be asked twice: first for the current and right after that for the envisioned prototype. For each question addressing the envisioned prototype, please try to imagine how the additional feature would alter – hopefully improve – the system and let me know your impressions. By the way, there is a **field for free feedback at the end of the interview**, so whenever you have additional feedback to any of the questions, please mention it there (possibly referring to a certain question explicitly, e.g. "@Q13: …").

Ok, let us start with a **tour of envisioned features not yet in the current prototype** – some of them are rather minor, others are more important, but the order in which they are presented does not reflect this. There will already be questions throughout the tour – actually, due to the many screenshots presented, the tour will cover more pages than the rest of the interview ☺ If you wonder why you are asked to rate features that you have not experienced yet, let me please make you aware that your impressions are very valuable for me: With cSpaces, some of you have used related or similar features for up to 5 months now, so I assume your impressions and judgments are more underpinned than those of other people that only experienced the system for just an hour or not at all.

Additional Features of the Envisioned Prototype: Debug View & Tagging

Although not unimportant but nevertheless lowest in priority was a **debug view** – similar like the ones you know from CoMem:

Deutsches Forschungszentrum für Künstliche Intelligenz

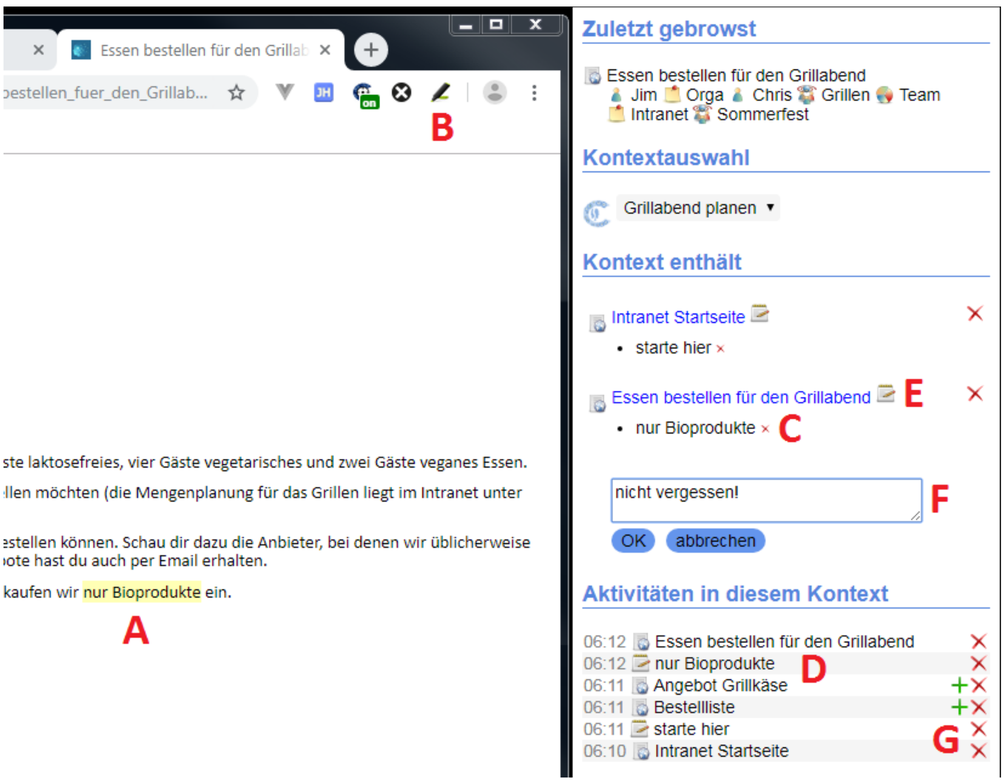
Property	Value
URI	pimo:13478324324:334
ID	0815
Preferred Label	Deutsches Forschungszentrum für Künstliche Intelligenz
Label	DFKI German Research Center for AI
Type	Organization
hasLocation	Kaiserslautern
hasMember	Andreas Dengel Heiko Maus

Figure A.4: Multi-month User Study: Final Interview Questionnaire (Page 1 of 16).

Additionally, **tagging all kinds of information elements** in the sidebar needs to be re-enabled (like you know it from the CoMem Desktop App):

Additional Features of the Envisioned Prototype:
Note Taking

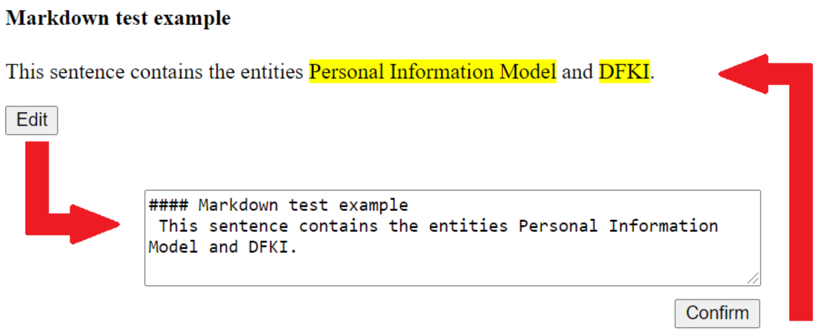
In our prototypes used 2019 at the University of Trier, we had a **manual note taking** feature as well as a feature to **convert highlighted text into a note** associated with the currently browsed website automatically:



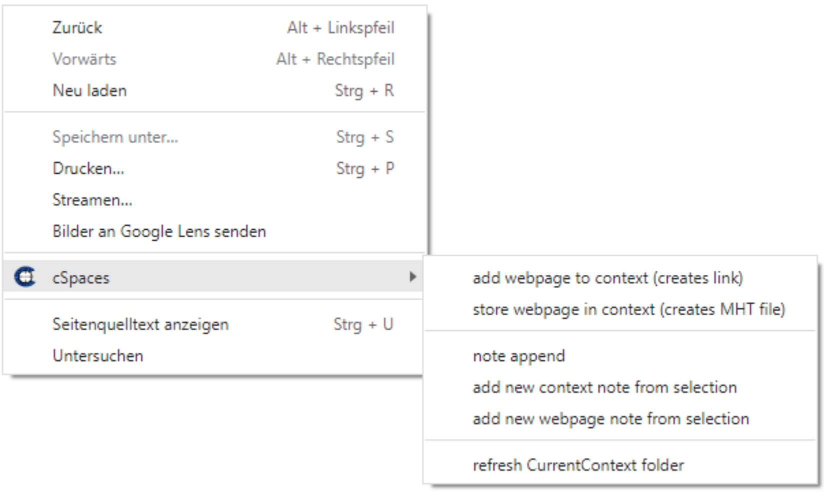
In the above screenshot this is illustrated for a web browser: you could highlight text (A) and click the **highlighter plug-in** (B) which converted the highlighted text into a note associated with the current website (C). If the website had not been “rebirthed” and added to the context yet, it would have been done at that point. Participants could also write manual notes by clicking the notepad icon next to each element in the context (E): a text area opened right below the item (F). We have an early prototype that improves this simple text editing area to also **support**

Figure A.5: Multi-month User Study: Final Interview Questionnaire (Page 2 of 16).

Markdown to allow for better text formatting etc. Even simple **highlighting of entities** found in **named entity recognition** performed on these notes is possible:



Another plug-in with further improvements is in development. It is designed as a **web browser right-click menu**:

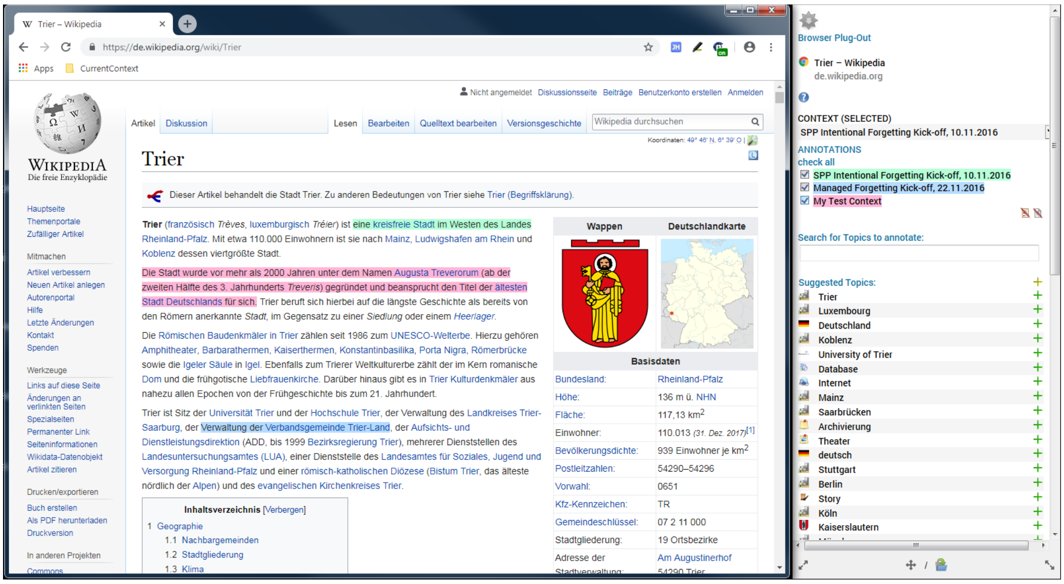


The plug-in is in early state. Like in the "Trier prototype" above, users are able to add highlighted text of a website as a **note to the currently browsed website ("webpage note")** in that context – this time using a right-click menu instead of the highlighter button in the browser's tool bar. The note would by default only be visible in that context. A website could then have **different notes in different contexts** – I will come back to this idea in a moment. The user may, however, always convert a contextual note to a general one sticking to that website in any context. As an enhancement to the aforementioned "Trier prototype", there is an option to highlight text and add it as a **note to the current context in general ("context note")**. This is helpful if you would like to capture some information relevant for the whole context and not a single one of its contained items in particular. No matter which option you choose, you can then **append to that particular note ("note append")** subsequent clicks. That way it is easily possible to assemble a kind of summarizing note for a website by sequentially highlighting parts of it, for example. A last idea regarding this plug-in was to allow for **saving a complete website for offline use as a self-contained file (save as MHT file)**.

Coming back to the idea of a website (or any other information item) having different notes in different contexts. The following screenshot illustrates such a scenario. However, in this early prototype, highlighted text was not automatically converted to notes in the sidebar but remained being just highlighted sections shown when returning

Figure A.6: Multi-month User Study: Final Interview Questionnaire (Page 3 of 16).

to the webpage and selecting the respective context in the sidebar. Please do not be confused by the looks of the sidebar: it was an early prototype with quite different appearance.



Here are first questions: please tick an answer by writing “x” or “X” in one of the cells. Some questions, like the second one, additionally have the option to add individual text right below the line of your ticked answer.

1. The concept of contextual notes (vs. general ones) is easy to understand.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

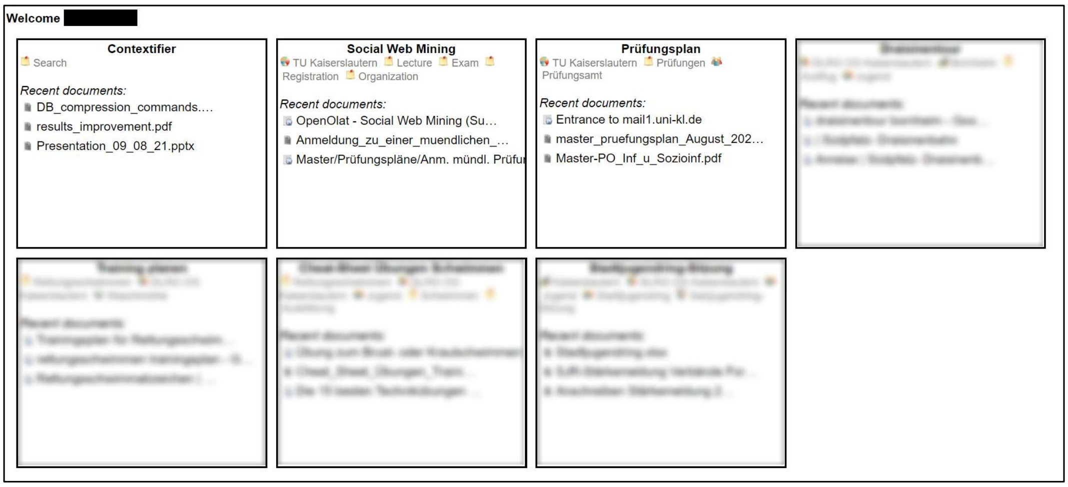
2. I have already seen contextual notes in existing software applications. (If this is the case, please name the applications in the field for individual text below.)						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree
<please replace this with your individual text if needed>						

3. Contextual notes would allow me to write things down exactly where I would like to have them.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

Figure A.7: Multi-month User Study: Final Interview Questionnaire (Page 4 of 16).

Additional Features of the Envisioned Prototype:
Get/keep an overview with a Welcome Screen / Context Overview

An important feature not yet enabled is a **welcome screen in your web browser when starting up the cSpaces app**. A prototypical version in development is given in the following screenshot:



Please note that this screenshot was consensually captured on one of the participants' computer. I nevertheless blurred some of the contexts to increase privacy a bit more. The idea of the **welcome screen (context overview)** is to provide an overview of those contexts with the highest memory buoyancy, thus users easily get an overview when resuming their work (after a break or at the beginning of a new work day, for example). Users can set the default amount of shown tiles according to their liking, e.g. 6, 9, 12 tiles – with a "more" button, users may always load the next X contexts in such a ranking. Each tile shows a **context summary** that could consist of **tags, high buoyancy items or last activities**. I will in a moment refer to another idea called "Last Focus" of a context which could also be shown for each context. **Clicking a context will bring up the sidebar with the context already selected** and you may start/resume your work with cSpaces as you know it.

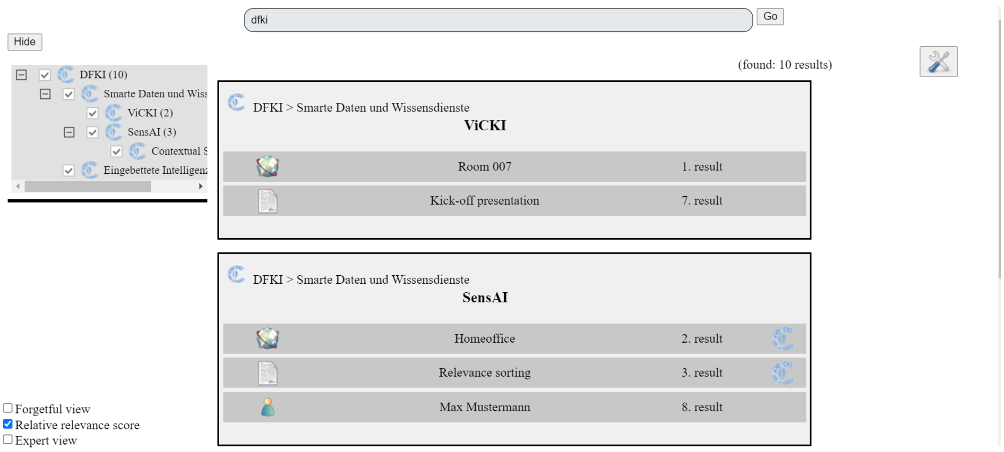
4. Context overview would be a useful feature to get/keep an overview of currently relevant contexts.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

Speaking of memory buoyancy (MB): in the current prototype, calculating MB has not been enabled, yet. The main reasons for this were that we already had two evaluations regarding MB (thus we reduced priority compared to completely non-evaluated features) and some parts are still missing in the current prototype to conduct a meaningful third evaluation. Having active MB calculation would however have resulted in different rankings across all GUIs you have seen. Please just remember that MB is one of the cornerstones of this approach and let us assume for now that its calculation works quite well – especially for a research prototype (as shown in earlier experiments). I will come back to memory buoyancy in a moment.

Figure A.8: Multi-month User Study: Final Interview Questionnaire (Page 5 of 16).

Additional Features of the Envisioned Prototype: Contextual & Forgetful Search

It is important to be able to also search in cSpaces – and since it is a rather autonomous agent (at least in the envisioned version), good search interfaces are even harder to create than regular ones already are. During the last years, we have already created two versions of a forgetful search. In a first version, users had the tendency to drag an MB slider to bring up further items already in the forgotten area of their personal knowledge graph since the things they were looking for were not among the search results. Dragging the slider was however not always the solution: the things they were looking for were not necessarily forgotten but rather the search query would have needed some rephrasing. Nevertheless, they could not be sure about it so they kept dragging the slider. A second version improved this situation by introducing a coverage meter and additionally showing and respecting contexts (divided into forgotten and still active ones) to allow for further drill-down. In a third version currently in development, we intend to further improve forgetful search by making contexts even more prominent and additionally taking context hierarchies into account. This version is in early development. Here are two screenshots of its current state: the first one as a plain contextual search and the second one with forgetting features enabled.



Forgetting-enabled version having a coverage meter and distinguishing between forgotten and still active contexts:

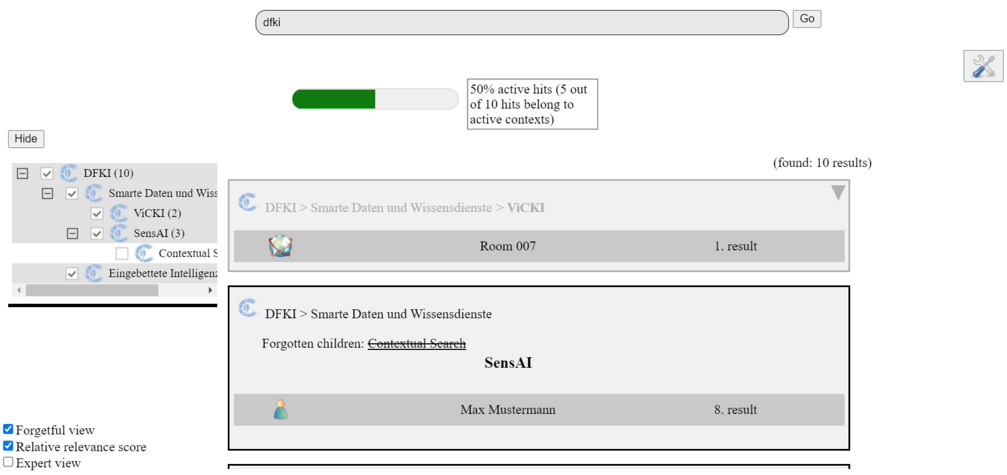
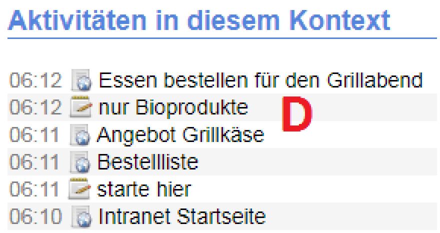


Figure A.9: Multi-month User Study: Final Interview Questionnaire (Page 6 of 16).

Additional Features of the Envisioned Prototype: “Sashboard” – Easily Switching between a Context’s GUI Representations

As you have realized the activity history in the current prototype shows every click. This was necessary for some of the evaluations. The 2019’s “Trier prototype” is more user-friendly: it **merges events and only shows the last browsing of a particular resource** and not each individual one. So each click brings the resource to the top of the activity list, removing older entries (see D).




(expand icon in the sidebar)

However, there are several occasions in which users need more space. So another idea was to allow the sidebar to be dragged at any time to become a full dashboard covering more parts of the screen – I metaphorically called this a **“sashboard”** referring to the mechanism of a sash window (see image on the left rotated by 90 degrees).

With this easy transition to a dashboard, users have more space when browsing the context’s content in more detail, writing (longer) notes or start a (full-fledged) search in that particular context (which we could already preset with contextual information). Needless to say that the dashboard basically holds the same information as the sidebar just with more available space to show it on screen, so certain views may be drawn in a less compact way, showing more details, etc. – selecting the same context again when switching views is not necessary of course: the dashboard is already set to show the same context as the sidebar.

5. Being able to easily switch between a context’s compact (sidebar) and more sprawling representation (dashboard) is a feature that would make my work more efficient.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

Additional Features of the Envisioned Prototype: Condensation by Capturing and Showing a Context’s Last Focus

You already know that **memory buoyancy (MB)** tries to bring up those information items which are presumably most relevant at the moment (while others that have lost relevance sink away). It is a form of information value assessment inspired by human memory and cognition that we have developed together with cognitive psychologists over the years. It is different from a **least recently used (LRU)** value: While for the last few clicks MB and LRU may have similar results, the differences become more evident the more time passes. MB is influenced by several factors like connections in the graph, the type of a certain element, start/end date of an event or task etc., while LRU only takes the last access date into account. For cSpaces, I always had two things in mind in this regard: MB as the main mechanism for ranking, hiding, reorganizing information items. But when returning to a context (resumption), it may nevertheless be helpful to show the user just those resources and activities that were last used or performed, respectively. Such a **“last focus” view** would be more driven by LRU than by MB and would serve as a complement.

Figure A.10: Multi-month User Study: Final Interview Questionnaire (Page 7 of 16).

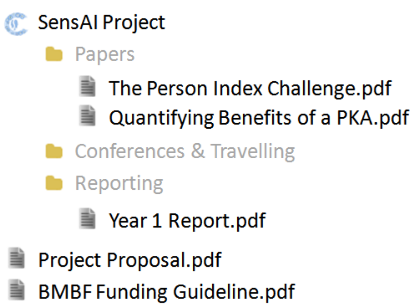
In **future versions**, we could enhance the scope of "last focus" to cover more aspects than just resources and activities, e.g. **open/active applications, the last people a user was in contact with (by chat, email, …) etc.**

6. I see a difference between Memory Buoyancy and Least Recent Used values.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

7. Looking at the last focus would help me remember more easily where I left off when returning to a context.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

Additional Features of the Envisioned Prototype: Condensation by having Flat Context Views

In one of this year's last updates, I introduced hierarchical contexts. This gives users more freedom to organize the content and structure of their context spaces and it also induces new potential for more helpful GUIs. One such GUI I had in mind for a long time is a **"flat" context view** similar to those found in software development environments.



Imagine browsing the "SensAI Project" context as depicted in the figure on the left. Having a flat view would not only show the direct content of that context (here: proposal and funding guideline) but **also a "most relevant selection" of each sub-context/folder** (here: two papers and year 1 report). Only if items in child contexts/folders are **above a certain memory buoyancy threshold** they make it into the flat view. In this example, no item of the "Conferences & Travelling" folder had a high enough buoyancy.
When returning to this context after some time of absence, a user is likely able to see the currently most relevant items right away – **borders between parent context and child contexts/folders are temporarily dissolved** – illustrated in this example by subfolders labeled in light gray.

This flat context view is meant as a **complementary condensation feature**: it is possibly enabled by default when returning to a context but can be disabled if a user browses deeper or explicitly turns it off. There could also be a threshold whether the feature should be turned on by default: rather empty contexts will less likely cognitively overwhelm the users when returning to them.

8. Flat Context Views are a feature that would help me being less overwhelmed when resuming a (non-sparsely filled) context.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

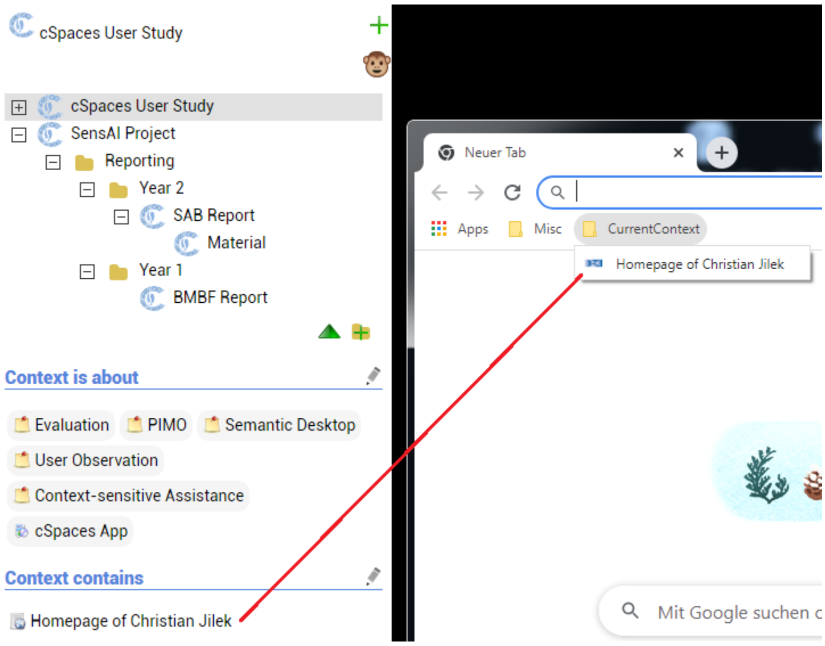
Figure A.11: Multi-month User Study: Final Interview Questionnaire (Page 8 of 16).

For the next two sections I recommend doing a little recap by watching a technical demo of cSpaces I created some years ago (video runtime: 7 min):

https://www.dfki.uni-kl.de/~jilek/demo/cspaces/demo.mp4

Additional Features of the Envisioned Prototype: Transparently Work in Context Spaces by having further Context Injections

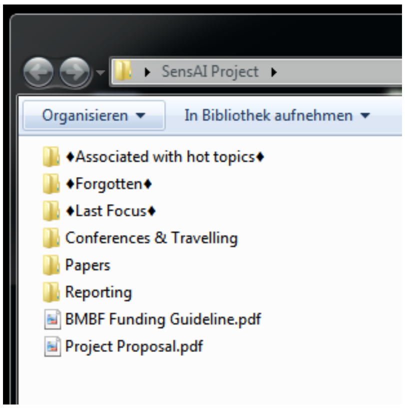
In the current prototype, **transparent context injection into the web browser** is already available: a filtered view (websites only) of the currently selected context is available as a **"Current Context" folder** in the browser's **bookmark bar**. Of course, sub-contexts/folders could be injected as well, this has so far been skipped for the sake of simplicity and development speed.



My intention was to also have all contexts **available in the mail client, the file system and possibly even the calendar** whereas the file system the most important of the three, since it has the highest impact: most software applications will at some point access persistent storage. For the file system and calendar the video above shows a prototypical realization. I will come back to these context injections in the section about "working with and in context spaces".

Figure A.12: Multi-month User Study: Final Interview Questionnaire (Page 9 of 16).

Additional Features of the Envisioned Prototype: (Self-)Reorganization by Intelligent Folder Injects (i.e. Virtual Context Subdivisions)



The video above also showed a feature I call **"Intelligent Folder Injects" (IFI)** which are more or less just **virtual subdivisions of contexts according to certain criteria**. They are primarily meant as a simple means to **access the results of temporal or permanent hiding, condensation or reorganization measures performed by the system**. In the technical demo video, two very simple criteria were used: creation and modification dates. However, with a knowledge graph, activity event stream and memory buoyancy at hand, we are able to realize more sophisticated IFIs. The figure on the left shows examples. The folder's main content (here: two pdf files and three sub-contexts/folders) could be just those items with a high buoyancy and all other items are hidden by default. Thus, if a user would like to see the rest of the content (i.e. the forgotten items), they could browse the **"Forgotten" IFI**.

IFIs are by the way highlighted with a **black diamond symbol** as their label's first and last character to set them apart from "real" sub-contexts/folders. The earlier mentioned **"Last Focus"** could be another IFI. And as a last example, imagine items in the current context that have a low *local* buoyancy (i.e. buoyancy within that particular context) but are tagged with topics, projects, persons, etc. that currently have a high *global* buoyancy (i.e. buoyancy outside that particular context). Those files could be interesting for the user even though not necessarily in the currently viewed context (but another one). Thus, an **"Associated with hot topics" IFI** could help to point the user to just those files.

As you may have learned from this example, **upon revisitation** of each context, cSpaces needs to calculate **this context's current appearance** ("incarnation"). This is done by a component I call the **incarnator**. Dependent on user settings (e.g. only show high buoyancy items by default, hide the rest), the folders actual content as well as all relevant IFIs are calculated (or at least injected – calculation can be lazy, i.e. actual calculation when entering a folder). Please note that all self-reorganization measures by the system **can be reverted** by the user and **everything is performed virtually**, not a single bit is actually moved on the harddisk ensuring **real-time performance**.

9. Intelligent Folder Injections are easy to understand.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

10. Intelligent Folder Injections are an easy way to work with a highly autonomous assistant like cSpaces and the results of its support decisions like hiding, condensation or reorganization.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

Additional Features of the Envisioned Prototype: (Self-)Reorganization by Automated Context Management

The last idea presented in this tour is **automated context management** consisting of four operations: **spawn, retract, split and merge**, which will be presented in more detail in the following:

Figure A.13: Multi-month User Study: Final Interview Questionnaire (Page 10 of 16).

spawn: You probably know that in a master thesis recently conducted with cSpaces, we were able to map observed user actions to contexts quite well. A certain percentage of actions, however, could not be mapped to a known context. A reason for these cases could be that **a completely new context had just been (or is in the process of being) created**. Whenever the system is in such a situation ("unknown context"), it could **create ("spawn") a new, empty context and start to associate activities (observed user actions) with it** (depending on the configuration we could also allow items, etc.). The **user may agree** to this new context – **explicitly** using an *accept* button or alike, **or implicitly** by adding things to it, renaming it, etc. (also dependent on given settings, e.g. should adding items be seen as implicit feedback of acceptance or not).

retract: Alternatively, the **user may disagree** and **direct the system to the existing context that he/she is actually in now**. Since this could also happen **quite some time later** (after several user activities or items have been added to the context, etc.), the system then needs to **merge ("retract") the wrongly assumed new context with the correct one given by the user**.

split: The system may always perform background calculations aiming at splitting large contexts for the sake of decluttering. For example, we had good experiences with the **clustering algorithm of PIMO Diary**. Using such an algorithm or alike, the system could come to the conclusion that a **context meanwhile consists of several sub-topic clusters**. It could present those clusters to the user and **ask whether this would be a good split**. The user may just **agree or give more detailed feedback**, e.g. by exchanging some items between the clusters. After that, the system **performs a context split** resulting in several smaller sub-contexts replacing the cluttered single one.

merge: The opposite operation is a **context merge**. For example, after a long time has passed and things dropped in relevance, it could be **not interesting anymore whether something was in a certain sub-context or not**. As a consequence, **child contexts could be merged with their parent** while only the most important items are kept and shown by default, whereas less important ones are hidden by default. Again, this could be done by asking the user and then performing the action or by simply just doing it and have an undo function for the user to correct it.

11. Spawning & Retraction are a feature I would find useful.

strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

12. Splitting is a feature I would find useful.

strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

13. Merging is a feature I would find useful.

strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

14. Regarding self-(re)organization in general, I would prefer *just doing* over *asking first*, i.e. the system performs the support measures it finds appropriate without asking and has an undo function to correct it, instead of asking for user feedback before doing something. (Please remember: all reorganization measures are virtual, nothing is physically moved on the harddisk.)

strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

Figure A.14: Multi-month User Study: Final Interview Questionnaire (Page 11 of 16).

Ok, that was the tour of additional features not yet available in the current prototype. Please also note that **we decided against an initial crawl of your system when installing cSpaces**. This would have been possible. As a consequence, all your files, mails, bookmarks etc. would have been known by and available in the system right away. So, when talking about the envisioned prototype, please remember that **in the "ultimate" vision, the assistant would possibly know everything available on your computer**.

Now, let us begin with the second part of the interview: your experience with the system so far. Please remember that questions may explicitly refer to the **current prototype** (the one you worked with in 2021) or the **envisioned prototype** (which additionally has the features presented in the tour).

User Activity Tracking & Privacy Preservation

As you know, having software agents tracking user activities is a privacy-sensitive topic. In this study, I chose a solution which never sends anything to a server – everything remains on your local system, for your use only. If something leaves your computer, it is typically sent by yourself, e.g. anonymized data of the study that you could review before sending and whose submission was not mandatory (nevertheless a much appreciated support ☺).

15. I feel comfortable with the current prototype's privacy preservation measures. (If this is not the case, please let us know your concerns in the field for individual text below.)						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree
<please replace this with your individual text if needed>						

16. I sometimes forgot that I am being observed by the system.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

Working with and in Context Spaces

For me, one of the most important aspects of the study was the possibility to have you **work with and in context spaces**. You could access your **contexts in the sidebar** and, after one of the last updates, they were **also injected into the web browser**. In the additional feature tour, you also saw **file system and calendar injection**. We also had a little prototype for **mail client injection**. By having your contexts in as many applications as possible, it would be possible to **work on only one contexts/folders hierarchy instead of several different ones** (in your file, mail and web browsing app). I hope you could – at least a bit – experience that feeling. If that is not the case, maybe give the web browser injection another try before answering the next questions.

There are several envisioned benefits of working with and in context spaces: First, the idea was that context spaces are **more aligned with your mental model** compared to using traditional applications to structure your information sphere. Second, self-(re)organizing and especially forgetful **knowledge assistants need a lot of contextual information to support a user**. How can we obtain this knowledge? By observing the user – but this is not always sufficient: there is presumably always a capture and interpretation gap. We need the user to help the software agent. Since we do not want to ask the user for explicit feedback too often (since this distracts them from their actual work, etc.), I rather tried to establish mechanisms that allow users to give implicit feedback and thus "model" their information sphere by subtle and tiny extra actions like dragging something to a certain context, add tags, etc. Ultimately this led to the idea of having people work with and in context spaces. Third, having this extra contextual

Figure A.15: Multi-month User Study: Final Interview Questionnaire (Page 12 of 16).

information, the assistant **can then provide support services that were not realistically possible before** – this serves as an additional incentive to work this way: users will have benefits when doing a tiny bit of extra work (i.e. doing the aforementioned dragging, tagging etc.). Ultimately such services hopefully lead to **reduced cognitive load** while working, **faster context switching/resumption**, etc.

17. Please describe how deeply integrated the cSpaces prototype was in your daily work, e.g. how much of your usual work did you perform with the system? Were there special activities you did with it? Was the sidebar "always" visible on screen or often minimized? etc.
<please replace this with your individual text>

18. The idea of (working with and in) Context Spaces is easy to understand.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

19. I have already seen something like Context Spaces in existing software applications. (If this is the case, please name the applications in the field for individual text below.)						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree
<please replace this with your individual text if needed>						

20. With the current prototype of cSpaces, information items (like files, bookmarks, etc.) are right where I would like to have them.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

21. With the envisioned prototype of cSpaces, information items (like files, bookmarks, etc.) would be right where I would like to have them.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

22. Context spaces as provided by the current prototype are more aligned with my mental model than with traditional file, mail and web browsing apps.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

23. Context spaces as provided by the envisioned prototype would be more aligned with my mental model than with traditional file, mail and web browsing apps.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

Figure A.16: Multi-month User Study: Final Interview Questionnaire (Page 13 of 16).

24. The current prototype of cSpaces enables faster context switches (i.e. it takes less time to resume another context) than with traditional file, mail and web browsing apps.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

25. The envisioned prototype of cSpaces would enable faster context switches (i.e. it takes less time to resume another context) than with traditional file, mail and web browsing apps.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

26. The current prototype of cSpaces helps in keeping my information sphere more tidied up.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

27. The envisioned prototype of cSpaces would help in keeping my information sphere more tidied up.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

28. I would trust the envisioned prototype of cSpaces in its hiding, condensation and reorganization decisions. (If this is not the case, please let us know your concerns in the field for individual text below.)						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree
<please replace this with your individual text if needed>						

29. The support measures as provided by the envisioned prototype of cSpaces need more (on-screen) explanations.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

30. Using Context Spaces in general (not necessarily the current prototypical implementation) to only view the current context (and its content) while hiding all others (and their content) reduces my cognitive load while working.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

31. Context Spaces in general (not necessarily the current prototypical implementation) are a suitable user interface to work with a highly autonomous assistant that is capable of self-(re)organization and forgetting.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

Figure A.17: Multi-month User Study: Final Interview Questionnaire (Page 14 of 16).

32. I could think of a better solution to subtly provide contextual meta-information to the assistant. (If this is the case, please let us know your ideas in the field for individual text below.)						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree
<please replace this with your individual text if needed>						

Usability Questions

Last not least, I would like to ask you about your impressions regarding the **current cSpaces prototype's usability**. As you will realize the following questions are standardized and intentionally not tailored to the current system.

33. Usefulness (1/3): It helps me be more effective.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

34. Usefulness (2/3): It helps me be more productive.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

35. Usefulness (3/3): It is useful.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

36. Ease of Use (1/3): It is easy to use.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

37. Ease of Use (2/3): It is user friendly.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

38. Ease of Use (3/3): It requires the fewest steps possible to accomplish what I want to do with it.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

39. Ease of Learning (1/3): I learned to use it quickly.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

Figure A.18: Multi-month User Study: Final Interview Questionnaire (Page 15 of 16).

40. Ease of Learning (2/3): I easily remember how to use it.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

41. Ease of Learning (3/3): It is easy to learn to use it.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

42. Satisfaction (1/3): I am satisfied with it.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

43. Satisfaction (2/3): I would recommend it to a friend.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

44. Satisfaction (3/3): It is fun to use.						
strongly disagree	quite disagree	slightly disagree	neither	slightly agree	quite agree	strongly agree

Free Feedback

Last not least (now really), here is the opportunity to **let me know anything else that you would like to provide as a feedback but did not match any of the questions so far**. Please remember the special situation you are in as the only ones having used the system for up to 5 months. **Any feedback – positive or negative – will very likely be helpful for me.** Thank you!

45. Free Feedback
<please replace this with your individual text if needed>

Thank you very much for taking the time to do this interview!

Please do not forget to send me this document (filled) and your anonymized cSpaces database*.

*) Anonymization is done using the **cSpaces anonymizer** available in the cSpaces deployment cloud: the file is called `cSpaces_Anonymizer_v1.zip` – just unpack it to your cSpaces installation directory and run `anonymize.bat` [`.sh` on Linux and Mac]. The anonymizer first copies the whole database, which may take several gigabytes of storage, and then performs the anonymization. After its completion, the file `anonymized_cSpaces.sqlite` in your installation directory is the one I need. **Hint:** the anonymized database will be much smaller than the original one (possibly even less than 1% of its original size) and its compressibility is quite high, so you may additionally zip it to make sending/uploading it even faster ☺

Figure A.19: Multi-month User Study: Final Interview Questionnaire (Page 16 of 16).

B Further General Remarks

This chapter complements the general remarks section (1.6) found in this thesis' introduction.

B.1 Remarks on Technical Protocol and Standard Citations

In this thesis, several technical protocols and standards are cited, e.g. HTTP, IMAP or RDF – just to name a few[1]. Many of them have evolved over decades, beginning with an initial version and followed by revisions, extensions, etc. This often led to long sequences of publications: As an example, consider the *Internet Message Access Protocol (IMAP)*. It is defined by several so-called *Requests for Comments (RFC)*, which are memoranda on Internet standards[2]. The most important ones in this case are presumably

- the initial RFC 1064 (Jul. 1988) [71] as well as
- RFC 1176 (Aug. 1990),
- RFC 1730 (Dec. 1994),
- RFC 2060 (Dec. 1996),
- RFC 3501 (Mar. 2003) and
- the currently most recent RFC 9051 (Aug. 2021) [248].

Since all these documents typically state to which (directly) preceding and succeeding documents they refer to, navigating such publication sequences is straightforward. Thus, when not referring to a particular version but the protocol/standard *in general*, the author decided to name the initial and the latest version while leaving out intermediate ones. This is mainly done for the sake of readability but serves three other purposes as well: citing the latest version allows to directly view the current state (1). Additionally citing the initial one shows the point in time a particular protocol/standard became available (2) while also giving credit to the original authors (3). Although authors of intermediate versions deserve credit as well, a trade-off solution had to be found between readability and providing proper references and background information. Coming back to the example above, citing IMAP in this thesis thus reads as "IMAP [71, 248]".

[1] Written-out forms and references were purposely omitted here. They are given each time one of these terms is used in its proper context for the first time.

[2] for example accessible online at `https://www.rfc-editor.org/`

B.2 Details on Used Box Plot Semantics

Box plots shown in this thesis use semantics originating from Tukey [378]. The box represents all data points from the lower (Q_1) to the upper quartile (Q_3), i.e. the mid 50% of the dataset. The distance between these quartiles, the innerquartile range IQR $= Q_3 - Q_1$, is used for defining the whiskers. Whiskers cover all data points within 1.5 times the IQR below Q_1 and above Q_3, respectively. They end at the last observed data point within each range. Values below or above, respectively, are considered outliers and are marked with a circle (○). The median (Q_2) is denoted by a thick black line (within the box) and the mean by a black diamond symbol (◆). The following figure illustrates these definitions using an example:

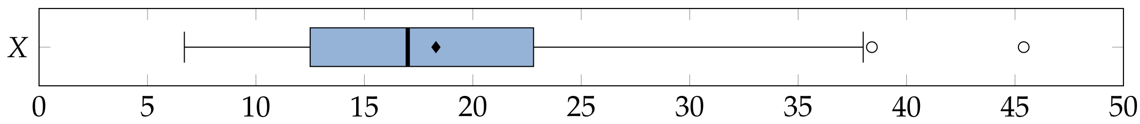


Figure B.1: Example to Illustrate Used Box Plot Semantics:

value	**symbol**	**meaning**
6.7	|	lower whisker: lowest observed data point within $Q_1 - 1,5 \cdot \text{IQR}$ and minimum observed data point in this example (no lower-end outliers)
12.5	|	lower quartile Q_1 (lower end of the box)
17.0	**|**	median (Q_2)
18.3	◆	mean $\bar{x}$
22.8	|	upper quartile Q_3 (upper end of the box)
38.0	|	upper whisker: highest observed data point within $Q_3 + 1,5 \cdot \text{IQR}$
38.4	○	first outlier
45.4	○	second outlier and maximum observed data point in this example

The author chose to denote the median by a *thick* black line to ensure its visibility in cases of the median coinciding with Q_1 or Q_3 (i.e. lying on one end of the box). In some cases, Q_1 additionally equals Q_3 making the median, the box and the whiskers collapse into a single line.

C
Software References and Dependencies

This chapter complements the Bibliography by software references – for further details please see Section 1.6. All links have been revisited by the author in August 2022 to ensure they still work. Thus, individual last access dates were omitted in the following. Readers may assume the **1st of August 2022 as a last access date for each reference** (if not stated otherwise, e.g. in cases of archived copies). Apart from typical information like authors and title, for each software reference the following details are provided:

- – If an artifact has been integrated into a software component developed in this thesis (**software dependency**), this is indicated by a special icon at the beginning of the reference. Note: Only the first, manually managed level of the dependency hierarchy is listed here, i.e. downstream dependencies were omitted.
- Version – Especially if an artifact is a dependency, its particular version is stated.
- Year or date – The year or date a given version has been released. If no particular version is stated, typically two years are given: the one of the initial release and another of the latest version. Depending on whether the software is explicitly stated to be discontinued, both numbers are either separated by a dash or slash. Examples are *iOS* [Sw25] (ongoing) and *Google Desktop* [Sw47] (discontinued).
- URL – A link to the artifact's major website is provided, which is often a *GitHub*[1] repository.
- MVN – An additional link to the *Maven Central Repository*[2] is given if the artifact is written in *Java* [Sw95] and has also been published there.
- [#] – If the artifact is associated with a specific paper, the respective reference is given, for example see [Sw115].

Sw1 Alfresco Software, *JLAN Embedded*, Version 6.4 (2018),
URL: https://www.alfresco.com/, MVN: https://artifacts.alfresco.com/nexus/content/repositories/public-snapshots/org/alfresco/alfresco-jlan-embed/

Sw2 Apache Software Foundation, *Commons CLI*, Version 1.4 (2017),
URL: https://commons.apache.org/proper/commons-cli/,
MVN: https://mvnrepository.com/artifact/commons-cli/commons-cli

Sw3 Apache Software Foundation, *Commons Configuration*, Version 1.9 (2012),
URL: https://commons.apache.org/proper/commons-configuration/, MVN: https://mvnrepository.com/artifact/commons-configuration/commons-configuration

[1] https://github.com/
[2] https://mvnrepository.com/repos/central

Sw4 Apache Software Foundation, *Commons IO*, Version 2.6 (2017),
URL: https://commons.apache.org/proper/commons-io/,
MVN: https://mvnrepository.com/artifact/commons-io/commons-io

Sw5 Apache Software Foundation, *Commons Lang*, Version 2.6 (2011),
URL: https://commons.apache.org/proper/commons-lang/,
MVN: https://mvnrepository.com/artifact/commons-lang/commons-lang

Sw6 Apache Software Foundation, *Commons Pool*, Version 1.6 (2012),
URL: https://commons.apache.org/proper/commons-pool/,
MVN: https://mvnrepository.com/artifact/commons-pool/commons-pool

Sw7 Apache Software Foundation, *Groovy: A powerful, dynamic language for the JVM*, Version 2.4.12 (2017), URL: https://groovy-lang.org/,
MVN: https://mvnrepository.com/artifact/org.codehaus.groovy/groovy-all

Sw8 Apache Software Foundation, *James, Protocols IMAP*, Version 3.0.1 (2017),
URL: https://james.apache.org/protocols/imap4.html, MVN: https://mvnrepository.com/artifact/org.apache.james.protocols/protocols-imap

Sw9 Apache Software Foundation, *Jena (Libraries POM)*, Version 3.6.0 (2017),
URL: https://jena.apache.org/,
MVN: https://mvnrepository.com/artifact/org.apache.jena/apache-jena-libs

Sw10 Apache Software Foundation, *Jena SPARQL Text Search*, Version 3.6.0 (2017),
URL: https://jena.apache.org/,
MVN: https://mvnrepository.com/artifact/org.apache.jena/jena-text

Sw11 Apache Software Foundation, *Jena SPARQL Text Search*, Version 4.5.0 (2022),
URL: https://jena.apache.org/,
MVN: https://mvnrepository.com/artifact/org.apache.jena/jena-text

Sw12 Apache Software Foundation, *Lucene (Core)*, Version 8.9.0 (2021),
URL: https://lucene.apache.org/,
MVN: https://mvnrepository.com/artifact/org.apache.lucene/lucene-core

Sw13 Apache Software Foundation, *Lucene (Queryparser)*, Version 8.9.0 (2021),
URL: https://lucene.apache.org/, MVN: https://mvnrepository.com/artifact/org.apache.lucene/lucene-queryparser

Sw14 Apache Software Foundation, *Lucene Common Analyzers*, Version 8.9.0 (2021),
URL: https://lucene.apache.org/, MVN: https://mvnrepository.com/artifact/org.apache.lucene/lucene-analyzers-common
(*GermanStemmer* based on Caumanns [55])

Sw15 Apache Software Foundation, *Maven Shade Plugin*, Version 2.4.3 (2016),
URL: https://maven.apache.org/plugins/maven-shade-plugin/, MVN: https://mvnrepository.com/artifact/org.apache.maven.plugins/maven-shade-plugin

Sw16 Apache Software Foundation, *OpenNLP Tools*, Version 1.9.0 (2018), URL: https://opennlp.apache.org/, MVN: https://mvnrepository.com/artifact/org.apache.opennlp/opennlp-tools

Sw17 Apache Software Foundation, *POI: Java API To Access Microsoft Format Files (Common)*, Version 4.1.2 (2020), URL: https://poi.apache.org/, MVN: https://mvnrepository.com/artifact/org.apache.poi/poi

Sw18 Apache Software Foundation, *POI: Java API To Access Microsoft Format Files, API Based On OPC and OOXML Schemas*, Version 4.1.2 (2020), URL: https://poi.apache.org/, MVN: https://mvnrepository.com/artifact/org.apache.poi/poi-ooxml

Sw19 Apache Software Foundation, *Tika (Core)*, Version 1.26 (2021), URL: https://tika.apache.org/, MVN: https://mvnrepository.com/artifact/org.apache.tika/tika-core

Sw20 Apache Software Foundation, *Tika Language Detection*, Version 1.26 (2021), URL: https://tika.apache.org/, MVN: https://mvnrepository.com/artifact/org.apache.tika/tika-langdetect

Sw21 Apache Software Foundation, *Tika Parser Modules*, Version 1.26 (2021), URL: https://tika.apache.org/, MVN: https://mvnrepository.com/artifact/org.apache.tika/tika-parsers

Sw22 Apache Software Foundation, *Tomcat Embedded (Core)*, Version 7.0.73 (2016), URL: https://tomcat.apache.org/, MVN: https://mvnrepository.com/artifact/org.apache.tomcat.embed/tomcat-embed-core

Sw23 Apache Software Foundation, *Tomcat Embedded Logging JULI*, Version 7.0.73 (2016), URL: https://tomcat.apache.org/, MVN: https://mvnrepository.com/artifact/org.apache.tomcat.embed/tomcat-embed-logging-juli

Sw24 Aperture Project, *Aperture: a Java framework for getting data and metadata* (2005/2010), URL: http://aperture.sourceforge.net/

Sw25 Apple Inc., *iOS* (2007/2022), URL: https://www.apple.com/ios/

Sw26 Apple Inc., *macOS* (2001/2022), URL: https://www.apple.com/macos/

Sw27 Apple Inc., *Siri* (2011/2022), part of various Apple operating systems, URL: https://www.apple.com/siri/

Sw28 BootstrapVue Project, *BootstrapVue*, Version 2.12.0 (2020), URL: https://github.com/bootstrap-vue/bootstrap-vue

Sw29 B. Brodski, *7-Zip-JBinding: free cross-platform java binding of 7-Zip (Core)*, Version 9.20-2.00beta (2015), URL: https://github.com/borisbrodski/sevenzipjbinding, MVN: https://mvnrepository.com/artifact/net.sf.sevenzipjbinding/sevenzipjbinding

Sw30 B. Brodski, *7-Zip-JBinding: free cross-platform java binding of 7-Zip, All-Platforms module*, Version 9.20-2.00beta (2015),
URL: https://github.com/borisbrodski/sevenzipjbinding,
MVN: https://mvnrepository.com/artifact/net.sf.sevenzipjbinding/sevenzipjbinding-all-platforms

Sw31 S. Capadisli and Read-Write Linked Data Team, *dokieli* (2013/2022),
URL: https://github.com/linkeddata/dokieli

Sw32 A. Chatterjee et al., *Nitrite: An in-memory, single-file based embedded NoSQL persistent document store*, Version 2.1.0 (2017),
URL: https://github.com/nitrite/nitrite-java,
MVN: https://mvnrepository.com/artifact/org.dizitart/nitrite

Sw33 W. W. Cohen, *Enron Email Dataset*, Version of 2015-05-07,
URL: https://www.cs.cmu.edu/~./enron/
(associated with Klimt and Yang [206])

Sw34 DBpedia Project, *DBpedia* (since 2007), URL: https://www.dbpedia.org/

Sw35 DBpedia Project, *German DBpedia Instance Types*, Version 3.9 (2015),
URL: https://downloads.dbpedia.org/3.9/de/instance_types_de.ttl.bz2

Sw36 C. Deckers et al., *DJ Native Swing: Web Browser, Flash Player, HTML editor, Media player for Swing*, Version 1.0.2.20120308 (2012),
URL: https://github.com/Chrriis/DJ-Native-Swing,
MVN: https://mvnrepository.com/artifact/chrriis/dj-nativeswing

Sw37 C. Deckers et al., *DJ Native Swing: Web Browser, Flash Player, HTML editor, Media player for Swing – SWT implementation*, Version 1.0.2.20120308 (2012),
URL: https://github.com/Chrriis/DJ-Native-Swing,
MVN: https://mvnrepository.com/artifact/chrriis/dj-nativeswing-swt

Sw38 Eclipse Foundation, *Eclipse Desktop & Web IDE* (2001/2022),
URL: https://www.eclipse.org/ide/

Sw39 Eclipse Foundation, *Eclipse SWT for Maven Users*, Version 4.5 (2015),
URL: https://github.com/maven-eclipse/maven-eclipse.github.io

Sw40 FasterXML LLC, *Jackson Databind*, Version 2.2.3 (2013),
URL: https://github.com/FasterXML/jackson, MVN: https://mvnrepository.com/artifact/com.fasterxml.jackson.core/jackson-databind

Sw41 FasterXML LLC, *Jackson Dataformat YAML*, Version 2.3.0 (2013),
URL: https://github.com/FasterXML/jackson-dataformats-text,
MVN: https://mvnrepository.com/artifact/com.fasterxml.jackson.dataformat/jackson-dataformat-yaml

Sw42 GATE Project, University of Sheffield, *GATE: General Architecture for Text Engineering (Core)*, Version 8.1 (2015), URL: https://gate.ac.uk/, MVN: https://mvnrepository.com/artifact/uk.ac.gate/gate-core

Sw43 Git Project, *Git* (2005/2022), URL: https://git-scm.com/

Sw44 Google LLC, *AngularJS: HTML enhanced for web apps*, Version 1.6.2 (2017), URL: https://angularjs.org/, https://github.com/angular/angular.js

Sw45 Google LLC, *Google Assistant* (2016/2022), URL: https://assistant.google.com/

Sw46 Google LLC, *Google Chrome* (2008/2022), URL: https://www.google.com/chrome/

Sw47 Google LLC, *Google Desktop* (2004–2011), originally online at http://desktop.google.com/ (not available anymore), archived copy of 2011-08-24, URL: https://web.archive.org/web/20110824073151/http://desktop.google.com:80/

Sw48 Google LLC, *Google Now*(2012–2015), succeeded by *Google Assistant* [Sw45]

Sw49 Google LLC, *Google Search* (since 1997), URL: https://www.google.com/

Sw50 Google LLC, *Gson: A Java serialization/deserialization library to convert Java Objects into JSON and back*, Version 2.8.0 (2016), URL: https://github.com/google/gson, MVN: https://mvnrepository.com/artifact/com.google.code.gson/gson

Sw51 Google LLC & Open Handset Alliance, *Android* (2008/2021), URL: https://www.android.com/

Sw52 M. Greenblatt and CEF Team, *Chromium Embedded Framework (CEF): A simple framework for embedding Chromium-based browsers in other applications*, Version of 2021-07-19, URL: https://github.com/chromiumembedded/cef

Sw53 M. Greenblatt and CEF Team, *Java Chromium Embedded Framework (JCEF): A simple framework for embedding Chromium-based browsers in other applications using the Java programming language*, Version of 2021-07-19, URL: https://github.com/chromiumembedded/java-cef

Sw54 J. Hedley, *Jsoup: Java HTML Parser*, Version 1.11.3 (2018), URL: https://jsoup.org/, MVN: https://mvnrepository.com/artifact/org.jsoup/jsoup

Sw55 M. Honnibal, I. Montani et al. (Explosion AI), *spaCy* (2015/2022), URL: https://github.com/explosion/spaCy

Sw56 H. Irawan et al., *WebDAV Servlet*, Version 2.0.2-fixlock (2013), URL: https://github.com/ceefour/webdav-servlet, MVN: https://mvnrepository.com/artifact/net.sf.webdav/webdav-servlet

Sw57 Jabsorb Project, *Jabsorb – A JSON-RPC ORB in Java and JavaScript*, Version 1.3.2 (2015), URL: https://code.google.com/archive/p/jabsorb/, MVN: https://mvnrepository.com/artifact/org.jabsorb/jabsorb

Sw58 Java CSV Project, *Java CSV Reader and Writer*, Version 2.0 (2008), URL: http://javacsv.sourceforge.net/, MVN: https://mvnrepository.com/artifact/net.sourceforge.javacsv/javacsv

Sw59 Java Native Access Project, *Java Native Access (JNA)*, Version 4.2.2 (2016), URL: https://github.com/java-native-access/jna, MVN: https://mvnrepository.com/artifact/net.java.dev.jna/jna

Sw60 Java Native Access Project, *Java Native Access (JNA) Platform*, Version 4.2.2 (2016), URL: https://github.com/java-native-access/jna, MVN: https://mvnrepository.com/artifact/net.java.dev.jna/jna-platform

Sw61 jQuery Team, *jQuery*, Version 3.1.1 (2016), URL: https://github.com/jquery/jquery

Sw62 JUnit Team, *JUnit*, Version 4.12 (2014), URL: https://github.com/junit-team/junit4, MVN: https://mvnrepository.com/artifact/junit/junit

Sw63 K. Kawaguchi et al., *COM4J: Type-safe Java/COM binding*, Version 2.1 (2014), URL: https://github.com/kohsuke/com4j, MVN: https://mvnrepository.com/artifact/org.jvnet.com4j/com4j

Sw64 K. Kawaguchi et al., *COM4J: Type-safe Java/COM binding*, *Code Generation Plugin*, Version 2.1 (2014), URL: https://github.com/kohsuke/com4j, MVN: https://mvnrepository.com/artifact/org.jvnet.com4j/maven-com4j-plugin

Sw65 KDE Community, *KDE Software Compilation 4 (KDE SC 4)* (2008–2017), URL: https://kde.org/

Sw66 C. Kohlschütter et al., *Boilerpipe: Boilerplate Removal and Fulltext Extraction From HTML Pages*, Version 1.1.0 (2010), URL: https://github.com/kohlschutter/boilerpipe, MVN: https://mvnrepository.com/artifact/de.l3s.boilerpipe/boilerpipe

Sw67 W. Kraaij, S. Koldijk and M. Sappelli, *The SWELL Knowledge Work Dataset for Stress and User Modeling Research* (2014), Data Archiving and Networked Services (DANS), URL: https://doi.org/10.17026/dans-x55-69zp (associated with Koldijk et al. [213])

Sw68 C. Kroells, *Thumbnailator – a thumbnail generation library for Java*, Version 0.4.8 (2014), URL: https://github.com/coobird/thumbnailator, MVN: https://mvnrepository.com/artifact/net.coobird/thumbnailator

Sw69 LanguageTooler GmbH, *LanguageTool*, Version 3.6 (2016), URL: https://github.com/languagetool-org/languagetool

Sw70 Legion of the Bouncy Castle, *Bouncy Castle Cryptography APIs Provider for JDK 1.5 and up*, Version 1.54 (2015), URL: https://www.bouncycastle.org/java.html, MVN: https://mvnrepository.com/artifact/org.bouncycastle/bcprov-jdk15on

Sw71 S. R. Lingala et al., *Zip4j – A Java library for zip files and streams*, Version 2.9.0 (2021), URL: https://github.com/srikanth-lingala/zip4j, MVN: https://mvnrepository.com/artifact/net.lingala.zip4j/zip4j

Sw72 MariaDB Foundation, *MariaDB Server: The open source relational database* (2009/2022), URL: https://mariadb.org/

Sw73 Mattermost Inc., *Mattermost* (2015/2022), URL: https://github.com/mattermost/mattermost-server

Sw74 P. N. Mendes, M. Jakob, J. Daiber et al., *DBpedia Spotlight* (2010/2013), URL: https://www.dbpedia-spotlight.org/demo/ (associated with Mendes et al. [249] and Daiber et al. [76])

Sw75 Metaweb Technologies Inc., *Freebase* (2007–2016), originally online at http://www.freebase.com/ (not available anymore), succeeded by *Wikidata* [Sw121]

Sw76 Micronode, *iCal4j: A Java library for reading and writing iCalendar (*.ics) files*, Version 2.1.5 (2017), URL: https://ical4j.github.io/, MVN: https://mvnrepository.com/artifact/org.mnode.ical4j/ical4j

Sw77 Microsoft Corporation, *Cortana* (2014/2021), URL: https://www.microsoft.com/en-us/cortana/

Sw78 Microsoft Corporation, *Microsoft Edge* (2015/2022), URL: https://www.microsoft.com/edge/

Sw79 Microsoft Corporation, *Microsoft Excel* (1987/2021), URL: https://www.microsoft.com/excel/

Sw80 Microsoft Corporation, *Microsoft Office* (1990/2021), URL: https://www.office.com/

Sw81 Microsoft Corporation, *Microsoft Office Assistant* (1996/2004), part of *Microsoft Office* [Sw80] (based on the *Lumière Project* [171])

Sw82 Microsoft Corporation, *Microsoft Outlook* (1997/2021), URL: https://www.microsoft.com/outlook/

Sw83 Microsoft Corporation, *Microsoft Windows* (1985/2022), URL: https://www.microsoft.com/windows/

Sw84 Microsoft Corporation, *Microsoft Word* (1983/2022), URL: https://www.microsoft.com/word/

Sw85 Microsoft Corporation, *Windows File Explorer* (1995/2022), formerly known as *Windows Explorer*, part of *Microsoft Windows* [Sw83]

Sw86 R. Moeller et al., *FST: fast java serialization drop in-replacement*, Version 2.56 (2017), URL: https://github.com/RuedigerMoeller/fast-serialization, MVN: https://mvnrepository.com/artifact/de.ruedigermoeller/fst

Sw87 Mozilla Foundation, *Mozilla Firefox* (2002/2022),
URL: https://www.mozilla.org/firefox/

Sw88 Mozilla Foundation, *Mozilla Thunderbird* (2003/2022),
URL: https://www.thunderbird.net/

Sw89 Mozilla Project, *Mozilla Interfaces* (e.g. part of *XULRunner SDK*), Version 1.9 (2015),
URL: https://ftp.mozilla.org/pub/xulrunner/releases/

Sw90 Mozilla Project, *Readability.js: A standalone version of the readability library used for Firefox Reader View*, Version 0.4.1 (2021),
URL: https://github.com/mozilla/readability

Sw91 D. Naber et al., *German part-of-speech dictionary*, Version 1.1 (2016),
URL: https://github.com/languagetool-org/german-pos-dict

Sw92 NEPOMUK Consortium, *NEPOMUK Semantic Desktop Prototype*, Final Build (March 2010), originally online at http://dev.nepomuk.semanticdesktop.org/ (not available anymore), archived copy of 2015-09-20,
URL: https://web.archive.org/web/20150920022727/http://dev.nepomuk.semanticdesktop.org/download/

Sw93 G. Nguyen, Š. Dlugolinský, M. Laclavík and M. Šeleng, Information and Knowledge oriented Technologies (IKT) Research Group, Slovak Academy of Sciences, *IKT Gazetteers*, Version 1.2 (2013), URL: http://ikt.ui.sav.sk/gazetteer/ (associated with Dlugolinský et al. [93] and Nguyen et al. [269])

Sw94 Noldus IT, *uLog* (2008/2022),
URL: https://www.noldus.com/human-behavior-research/products/ulog
(associated with van Drunen et al. [386])

Sw95 OpenHMS, Health Market Science Inc., *Jackcess: A pure Java library for reading from and writing to MS Access databases*, Version 2.1.2 (2015),
URL: https://jackcess.sourceforge.io/, MVN: https://mvnrepository.com/artifact/com.healthmarketscience.jackcess/jackcess

Sw96 OpenJDK Project, *Java Platform, Standard Edition (Java SE)* (2007/2022),
URL: https://openjdk.org/

Sw97 OpenJFX Project, *OpenJFX* (2011/2022), URL: https://github.com/openjdk/jfx

Sw98 Opera, *Opera Web Browser* (1995/2022), URL: https://www.opera.com/

Sw99 Oracle Corporation, *Java Servlet API*, Version 3.1.0 (2013),
URL: https://javaee.github.io/servlet-spec/,
MVN: https://mvnrepository.com/artifact/javax.servlet/javax.servlet-api

Sw100 Oracle Corporation, *MySQL* (1995/2022), URL: https://www.mysql.com/

Sw101 Oracle Corporation, *NoSQL Database Server (Berkley Database Java Edition – build and runtime support)*, Version 18.3.1 (2018),
URL: https://www.oracle.com/database/nosql/technologies/nosql/,
MVN: https://mvnrepository.com/artifact/com.sleepycat/je

Sw102 M. Piacentini, R. Ravanini, J. Miltner, P. Morgan, R. Peinthor, M. Kleusberg, J. Clift et al., *DB Browser for SQLite* (2003/2022),
URL: https://github.com/sqlitebrowser/sqlitebrowser

Sw103 Python Software Foundation, *Python* (1991/2022), URL: https://www.python.org/

Sw104 QOS.ch Sàrl, *Logback: The reliable, generic, fast and flexible logging framework for Java Classic Module*, Version 1.2.5 (2021), URL: https://github.com/qos-ch/logback,
MVN: https://mvnrepository.com/artifact/ch.qos.logback/logback-classic

Sw105 T. L. Saito et al., *SQLite JDBC Driver*, Version 3.20.1 (2017),
URL: https://github.com/xerial/sqlite-jdbc,
MVN: https://mvnrepository.com/artifact/org.xerial/sqlite-jdbc

Sw106 S. Schwarz, P. Mohr, L. Coyle, L. Sauermann, J. Gaugaz, G. Demartini and M.Kopycki, *UserContext* (2007–2009), Originally online at http://usercontext.opendfki.de/ (not available anymore). Archived copy of 2022-01-24. URL: https://web.archive.org/web/20220124110222/http://usercontext.opendfki.de/

Sw107 S. Schwarz, M. Plössl and A. Wagner, *Dragontalk – Talking with the (Mozilla) Dragon* (2005–2010), Originally online at http://dragontalk.opendfki.de/ (not available anymore). Archived copy of 2021-01-27. URL: https://web.archive.org/web/20210127101842/http://dragontalk.opendfki.de/

Sw108 S. Schwarz, L. Sauermann, J. Gaugaz, G. Demartini, *User Observation Hub (UOH)* (2007–2009), Originally online at
http://usercontext.opendfki.de/wiki/UserObservationHub (not available anymore). Archived copy of 2022-01-20.
URL: https://web.archive.org/web/20220120220306/http://usercontext.opendfki.de/wiki/UserObservationHub

Sw109 S. Schwarz, F. Steinmann and A. Lauer, *PimoScript*, Version of 2013-03-21, DFKI-internal code

Sw110 S. Schwarz, F. Steinmann, H. Maus et al., *PIMO Ontology Used in CoMem, exported as PimoScript [Sw109] code*, Version of 2017-04-13 (evolved from *PIMO Ontology* [307]; documentation available in Section A.1), DFKI-internal code

Sw111 SLF4J Project, *Simple Logging Facade for Java (SLF4J), API Module*, Version 1.7.25 (2017), URL: https://www.slf4j.org/,
MVN: https://mvnrepository.com/artifact/org.slf4j/slf4j-api

Sw112 SLF4J Project, *Simple Logging Facade for Java (SLF4J), JDK14 Binding*, Version 1.7.25 (2017), URL: https://www.slf4j.org/,
MVN: https://mvnrepository.com/artifact/org.slf4j/slf4j-jdk14

Sw113 Stanford Natural Language Processing Group, *Stanford CoreNLP*, Version 3.9.1 (2018), URL: https://stanfordnlp.github.io/CoreNLP/, MVN: https://mvnrepository.com/artifact/edu.stanford.nlp/stanford-corenlp

Sw114 L. Torvalds et al., *Linux* (1991/2022), URL: https://www.kernel.org/

Sw115 Ubiquitous Knowledge Processing Lab (UKP), TU Darmstadt, *Java-based Wiktionary Library (JWKTL)*, Version 1.0.1 (2014), URL: https://github.com/dkpro/dkpro-jwktl, MVN: https://mvnrepository.com/artifact/de.tudarmstadt.ukp.jwktl/jwktl (associated with Zesch et al. [405])

Sw116 Vue.js Project, *Vue.js: A progressive, incrementally-adoptable JavaScript framework for building UI on the web*, Version 2.6.10 (2019), URL: https://github.com/vuejs/vue

Sw117 vue-jstree Project, *vue-jstree: A Tree Plugin For Vue 2.0+*, Version of 2018-04-25, URL: https://github.com/zdy1988/vue-jstree

Sw118 vue-markdown Project, *vue-markdown: A Powerful and Highspeed Markdown Parser for Vue*, Version of 2019-01-24, URL: https://github.com/miaolz123/vue-markdown

Sw119 A. Wagner, S. Schwarz and H. Maus, *FireTag – Annotation Sidebar for Mozilla Firefox and Thunderbird* (2013–2015), URL: https://github.com/wagnerand/FireTag/

Sw120 L. Wang et al., *pdf2htmlEX – Convert PDF to HTML without losing text or format*, Version 0.14.6 (2015), URL: https://github.com/coolwanglu/pdf2htmlEX

Sw121 Wikimedia Foundation, *Wikidata* (since 2012), URL: https://www.wikidata.org/

Sw122 Wikimedia Foundation, Wikimedia Downloads, *German Wikipedia dump*, Version of 2016-11-01, URL: https://dumps.wikimedia.org/backup-index.html (Statistics about the German Wikipedia at the time can be found in Wikimedia Foundation [400].)

Sw123 Wikimedia Foundation, Wikimedia Downloads, *German Wiktionary dump*, Version of 2016-07-01, URL: https://dumps.wikimedia.org/backup-index.html

Sw124 M. Wilson, C. G. Hagerty, D. Daniels, J. Phillips and K. Tsujino, *Annozilla (Annotea on Mozilla)* (2001–2009), originally online at http://annozilla.mozdev.org/ (not available anymore), archived copy of 2020-02-20, URL: https://web.archive.org/web/20200220232643/http://annozilla.mozdev.org/

Sw125 Z. Zhang et al., *JATE*, Version 1.11 (2013), URL: https://code.google.com/archive/p/jatetoolkit/ (associated with Zhang et al. [406])

Bibliography

Remark on Provided Paper Links. In August 2022, a freely accessible version (e.g. open-access, an author's copy or preprint) for the majority of referenced publications could be found on arXiv[1], CiteSeerX[2], Internet Archive's Wayback Machine[3], ResearchGate[4] or the websites of the respective authors, universities, research institutes or publishers (other sources like Google Books[5] not considered here). Thus, **apart from very few exceptions, each provided URL directly links to a text document** reducing the reader's effort to get a first glimpse on the respective paper to a single mouse click. However, given the dynamic character of the World Wide Web, more and more dead links are likely to occur in the future. To ease finding alternative versions – or in cases of preprints and author's copies the final publication – additional identifiers like arXiv ID, DOI or ISBN number are provided for each reference (if available).

Remark on Online References. All online references have been revisited by the author in August 2022. Their content (i.e. text, figures, etc.) was still as intended to be referenced by this thesis. Thus, individual last access dates were omitted in the following. Readers may assume the **1st of August 2022 as a last access date for each online reference** (if not stated otherwise, e.g. in cases of archived copies).

Remark on Software References. Software references are listed in Appendix C, whereas further remarks can be found in Section 1.6.

[1] W. van der Aalst. "Structuring Behavior or Not, That is the Question." In: *The Art of Structuring: Bridging the Gap Between Information Systems Research and Practice*. Ed. by K. Bergener, M. Räckers and A. Stein. Springer International Publishing, 2019, pp. 221–226. DOI: 10.1007/978-3-030-06234-7_21. URL: http://www.padsweb.rwth-aachen.de/wvdaalst/publications/p1050.pdf (cit. on p. 220).

[2] A. Aamodt and E. Plaza. "Case-based reasoning: Foundational issues, methodological variations, and system approaches." In: *AI communications* 7.1 (1994), pp. 39–59. DOI: 10.3233/AIC-1994-7104. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.89.4054&rep=rep1&type=pdf (cit. on p. 215).

[1] https://arxiv.org/
[2] http://citeseerx.ist.psu.edu/
[3] https://web.archive.org/
[4] https://www.researchgate.net/
[5] https://books.google.com/

[3] C. Abela and C. Staff. "Behaviour Mining for Automatic Task-Keeping and Visualisations for Task-Refinding." In: *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval (CHIIR '16), Carrboro, North Carolina, USA*. ACM, 2016, pp. 23–32. DOI: 10.1145/2854946.2854966 (cit. on pp. 57, 117).

[4] C. Abela, C. Staff and S. Handschuh. "Automatic Task-Cluster Generation based on Document Switching and Revisitation." In: *Posters, Demos, Late-breaking Results and Workshop Proceedings of the 23rd Conference on User Modeling, Adaptation, and Personalization (UMAP 2015), Dublin, Ireland, June 29 - July 3, 2015*. Ed. by A. Cristea, J. Masthoff, A. Said and N. Tintarev. Vol. 1388. CEUR Workshop Proceedings. CEUR-WS.org, 2015. URL: http://ceur-ws.org/Vol-1388/DeCat2015-paper1.pdf (cit. on pp. 57, 117).

[5] C. Abela, C. Staff and S. Handschuh. "Collecting and Analysing Personal Information Management Data." In: *Proceedings of the First DIACHRON Workshop on Managing the Evolution and Preservation of the Data Web co-located with 12th European Semantic Web Conference (ESWC 2015)*. Ed. by J. Debattista, M. d'Aquin and C. Lange. Vol. 1377. CEUR Workshop Proceedings. CEUR-WS.org, 2015. URL: http://ceur-ws.org/Vol-1377/paper3.pdf (cit. on pp. 57, 91, 117, 142).

[6] S. Abney. "Partial parsing via finite-state cascades." In: *Natural Language Engineering* 2.4 (1996), pp. 337–344. DOI: 10.1017/S1351324997001599. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.30.566&rep=rep1&type=pdf (cit. on pp. 102, 108).

[7] ACTIVE Consortium. *ACTIVE: Enabling the Knowledge Powered Enterprise*. Project website. Originally online at http://active-project.eu/ (not available anymore). Archived copy of 2011-03-05. 2008. URL: https://web.archive.org/web/20110305113841/http://active-project.eu/. Also see https://cordis.europa.eu/project/id/215040 (cit. on pp. 56, 65).

[8] ADiWa Consortium. *ADiWa: Alliance Digital Warenfluss [Alliance Digital Product Flow]*. Project website. Originally online at http://www.adiwa.net/ (not available anymore). Archived copy of 2015-02-02. 2009. URL: https://web.archive.org/web/20150202200550/http://adiwa.net/ (cit. on pp. 52, 53).

[9] A. V. Aho and M. J. Corasick. "Efficient string matching: an aid to bibliographic search." In: *Communications of the ACM* 18.6 (1975), pp. 333–340. DOI: 10.1145/360825.360855. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.96.4671&rep=rep1&type=pdf (cit. on pp. 102, 108).

[10] E. M. Altmann and J. G. Trafton. "Task interruption: Resumption lag and the role of cues." In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 26. 26. 2004. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.4.5438&rep=rep1&type=pdf (cit. on p. 166).

[11] C. Amrit, A. B. J. M. Wijnhoven and D. Beckers. "Information Waste on the World Wide Web and Combating the Clutter." In: *Proceedings of the European Conference on Information Systems 2015 completed research papers*. 2015. DOI: 10.18151/7217266. URL: https://research.utwente.nl/files/232243236/Amrit2015information.pdf (cit. on p. 125).

[12] Apache Software Foundation. *Class IndexReader*. Lucene 9.1.0 core API (online documentation). 2022. URL: https://lucene.apache.org/core/9_1_0/core/org/apache/lucene/index/IndexReader.html (cit. on p. 77).

[13] APOSDLE Consortium. *APOSDLE: Advanced Process-Oriented Self-Directed Learning Environment*. Project website. Originally online at http://www.aposdle.tugraz.at/ (not available anymore). Archived copy of 2019-05-03. 2006. URL: https://web.archive.org/web/20190503192712/http://www.aposdle.tugraz.at/. Also see https://cordis.europa.eu/project/id/027023 (cit. on p. 56).

[14] K. Apostolidis, F. Markatopoulou, C. Tzelepis, V. Mezaris and I. Patras. "Multimedia Processing Essentials." In: *Personal Multimedia Preservation: Remembering or Forgetting Images and Video*. Ed. by V. Mezaris, C. Niederée and R. H. Logie. Springer, 2018, pp. 47–98. DOI: 10.1007/978-3-319-73465-1_3 (cit. on pp. 30, 57).

[15] M. H. Ashcraft. "Cognitive arithmetic: A review of data and theory." In: *Cognition* 44.1 (1992), pp. 75–106. DOI: 10.1016/0010-0277(92)90051-I. URL: https://www.researchgate.net/profile/Mark-Ashcraft/publication/222340465_Ashcraft_M_H_Cognitive_arithmetic_a_review_of_data_and_theory_Cognition_44_75-106/links/5a2aee1ca6fdccfbbf8520fd/Ashcraft-M-H-Cognitive-arithmetic-a-review-of-data-and-theory-Cognition-44-75-106.pdf (cit. on p. 163).

[16] M. H. Ashcraft and E. P. Kirk. "The relationships among working memory, math anxiety, and performance." In: *Journal of Experimental Psychology: General* 130.2 (2001), pp. 224–237. DOI: 10.1037/0096-3445.130.2.224. URL: https://www.apa.org/news/press/releases/xge1302224.pdf (cit. on p. 163).

[17] J. Attard and R. Brennan. "A Semantic Data Value Vocabulary Supporting Data Value Assessment and Measurement Integration." In: *Proceedings of the 20th International Conference on Enterprise Information Systems - Volume 2: ICEIS*. INSTICC. SciTePress, 2018, pp. 133–144. DOI: 10.5220/0006777701330144. URL: https://doras.dcu.ie/22982/1/iceis_attardandBrennan.pdf (cit. on p. 125).

[18] J. Attard and R. Brennan. "Challenges in Value-Driven Data Governance." In: *On the Move to Meaningful Internet Systems. OTM 2018 Conferences, Confederated International Conferences: CoopIS, C&TC, and ODBASE 2018, Valletta, Malta, October 22-26, 2018, Proceedings, Part II*. Ed. by H. Panetto, C. Debruyne, H. A. Proper, C. A. Ardagna, D. Roman and R. Meersman. Springer, 2018, pp. 546–554. DOI: 10.1007/978-3-030-02671-4_33. URL: https://core.ac.uk/download/pdf/195384462.pdf (cit. on p. 126).

[19] J. Attard and R. Brennan. "DaVe: A Semantic Data Value Vocabulary to Enable Data Value Characterisation." In: *Enterprise Information Systems, 20th International Conference, ICEIS 2018, Funchal, Madeira, Portugal, March 21-24, 2018, Revised Selected Papers*. Ed. by S. Hammoudi, M. Śmiałek, O. Camp and J. Filipe. Springer, 2019, pp. 239–261. DOI: 10.1007/978-3-030-26169-6_12. URL: https://www.researchgate.net/profile/Rob-Brennan/publication/334722723_DaVe_A_Semantic_Data_Value_Vocabulary_to_Enable_Data_Value_Characterisation/links/5d529f424585153040709a4a/DaVe-A-Semantic-Data-Value-Vocabulary-to-Enable-Data-Value-Characterisation.pdf (cit. on p. 125).

[20] S. Bahri, N. Zoghlami, M. Abed and J. M. R. S. Tavares. "BIG DATA for Healthcare: A Survey." In: *IEEE Access* 7 (2019), pp. 7397–7408. DOI: 10.1109/ACCESS.2018.2889180. URL: https://ieeexplore.ieee.org/iel7/6287639/8600701/08585021.pdf (cit. on p. 15).

[21] M. Bakhshizadeh, C. Jilek, H. Maus and A. Dengel. "Leveraging Context-aware Recommender Systems for Improving Personal Knowledge Assistants by Introducing Contextual States." In: *Proceedings of the "Lernen, Wissen, Daten, Analysen – Learning, Knowledge, Data, Analytics" 2021 Workshops (LWDA 2021): FGWM, KDML, FGWI-BIA, and FGIR*. Ed. by T. Seidl, M. Fromm and S. Obermeier. Vol. 2993. CEUR Workshop Proceedings. CEUR-WS.org, 2021. URL: http://ceur-ws.org/Vol-2993/paper-01.pdf (cit. on pp. 211, 218, 227).

[22] X. Bao, J. L. Herlocker and T. G. Dietterich. "Fewer Clicks and Less Frustration: Reducing the Cost of Reaching the Right Folder." In: *Proceedings of the 11th International Conference on Intelligent User Interfaces (IUI '06), Sydney, Australia*. ACM, 2006, pp. 178–185. DOI: 10.1145/1111449.1111490. URL: https://web.engr.oregonstate.edu/~tgd/publications/IUI06-FolderPredictor.pdf (cit. on p. 56).

[23] D. Beckett, T. Berners-Lee, E. Prud'hommeaux and G. Carothers. *RDF 1.1 Turtle – Terse RDF Triple Language*. W3C Recommendation. W3C, Feb. 2014. URL: https://www.w3.org/TR/turtle/ (cit. on p. 233).

[24] S. L. Beilock and T. H. Carr. "When High-Powered People Fail: Working Memory and "Choking Under Pressure" in Math." In: *Psychological Science* 16.2 (2005), pp. 101–105. DOI: 10.1111/j.0956-7976.2005.00789.x. URL: https://hpl.uchicago.edu/files/2018/07/Psychological-Science2005-1b1qwwq.pdf (cit. on p. 163).

[25] O. Ben-Kiki, C. Evans and B. Ingerson. *YAML Ain't Markup Language (YAML) 1.0, Final Draft*. Online. Jan. 2004. URL: https://yaml.org/spec/1.0/ (cit. on pp. xxxi, 146).

[26] Y. Benn, O. Bergman, L. Glazer, P. Arent, I. D. Wilkinson, R. Varley and S. Whittaker. "Navigating through digital folders uses the same brain structures as real world navigation." In: *Scientific reports* 5 (2015), p. 14719. DOI: 10.1038/srep14719. URL: https://www.nature.com/articles/srep14719.pdf (cit. on p. 44).

[27] O. Bergman, R. Beyth-Marom and R. Nachmias. "The Project Fragmentation Problem in Personal Information Management." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, 2006, pp. 271–274. DOI: 10.1145/1124772.1124813. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.100.6405&rep=rep1&type=pdf (cit. on pp. 17, 37).

[28] O. Bergman, R. Beyth-Marom and R. Nachmias. "The user-subjective approach to personal information management systems design: Evidence and implementations." In: *Journal of the American Society for Information Science and Technology* 59.2 (2008), pp. 235–246. DOI: 10.1002/asi.20738. URL: https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.20738 (cit. on pp. 17, 18, 59, 65, 73, 82, 141).

[29] O. Bergman, R. Beyth-Marom, R. Nachmias, N. Gradovitch and S. Whittaker. "Improved search engines and navigation preference in personal information management." In: *ACM Transactions on Information Systems* 26.4 (Oct. 2008), pp. 1–24. DOI: 10.1145/1402256.1402259. URL: https://www.researchgate.net/profile/Steve-Whittaker/publication/220515581_Improved_Search_Engines_and_Navigation_Preference_in_Personal_Information_Management/links/00b7d516cc928cbd93000000/Improved-Search-Engines-and-Navigation-Preference-in-Personal-Information-Management.pdf (cit. on p. 17).

[30] O. Bergman, S. Whittaker, M. Sanderson, R. Nachmias and A. Ramamoorthy. "The effect of folder structure on personal file navigation." In: *Journal of the American Society for Information Science and Technology* 61.12 (Dec. 2010), pp. 2426–2441. DOI: 10.1002/asi.21415. URL: https://www.researchgate.net/profile/Mark-Sanderson-3/publication/220432870_The_Effect_of_Folder_Structure_on_Personal_File_Navigation/links/5a170ff3aca272df0808a6b9/The-Effect-of-Folder-Structure-on-Personal-File-Navigation.pdf (cit. on p. 17).

[31] T. Berners-Lee. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*. Harper San Francisco, Sept. 1999. ISBN: 1402842937 (cit. on pp. xxxi, 25).

[32] T. Berners-Lee. *Linked Data*. Design Issues. W3C, July 2006. URL: https://www.w3.org/DesignIssues/LinkedData.html (cit. on pp. xxx, 29, 48).

[33] T. Berners-Lee, R. Fielding and H. Frystyk. *Hypertext Transfer Protocol – HTTP/1.0*. RFC 1945. RFC Editor, May 1996. DOI: 10.17487/RFC1945. URL: https://www.rfc-editor.org/rfc/rfc1945.html (cit. on pp. xxx, 154).

[34] T. Berners-Lee, R. Fielding and L. Masinter. *Uniform Resource Identifiers (URI): Generic Syntax*. RFC 2396. RFC Editor, Aug. 1998. DOI: 10.17487/RFC2396. URL: https://www.rfc-editor.org/rfc/rfc2396.html (cit. on pp. xxxi, 233).

[35] T. Berners-Lee, R. Fielding and L. Masinter. *Uniform Resource Identifier (URI): Generic Syntax*. RFC 3986. RFC Editor, Jan. 2005. DOI: 10.17487/RFC3986. URL: https://www.rfc-editor.org/rfc/rfc3986.html (cit. on pp. xxxi, 75, 233).

[36] T. Berners-Lee, J. Hendler and O. Lassila. "The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities." In: *Scientific American* 284.5 (2001), pp. 34–43. URL: http://web.archive.org/web/20040309192648/http://www.sciam.com/print_version.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21 (cit. on p. 19).

[37] T. Bienz and R. Cohn. *Portable Document Format Reference Manual (Adobe PDF 1.0)*. Addison-Wesley, 1993. ISBN: 0-201–62628–4. URL: https://opensource.adobe.com/dc-acrobat-sdk-docs/pdfstandards/pdfreference1.0.pdf (cit. on p. xxxi).

[38] R. A. Bjork. "Retrieval inhibition as an adaptive mechanism in human memory." In: *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (1989), pp. 309–330. URL: https://bjorklab.psych.ucla.edu/wp-content/uploads/sites/13/2016/07/RBjork_1989.pdf (cit. on p. 161).

[39] S. Bloehdorn, O. Görlitz, S. Schenk and M. Völkel. "TagFS – Tag Semantics for Hierarchical File Systems." In: *Proceedings of the 6th International Conference on Knowledge Management (I-KNOW 06), Graz, Austria*. 2006. URL: https://www.researchgate.net/profile/Max-Voelkel/publication/240789787_TagFS_Tag_Semantics_for_Hierarchical_File_Systems/links/5c6a8827a6fdcc404eb79f12/TagFS-Tag-Semantics-for-Hierarchical-File-Systems.pdf (cit. on pp. 53, 61).

[40] A. Bogomolny. *Modular Arithmetic*. Cut the Knot – Interactive Mathematics Miscellany and Puzzles (Website). 1996. URL: https://www.cut-the-knot.org/blue/Modulo.shtml (cit. on p. 163).

[41] W. Bokhove, B. Hulsebosch, B. van Schoonhoven, M. Sappelli and K. Wouters. "User Privacy in Applications for Well-being and Well-working: Requirements and Approaches for User Controlled Privacy." In: (2012). Ed. by M. Weyn. URL: https://repository.ubn.ru.nl/bitstream/handle/2066/103869/103869.pdf (cit. on p. 56).

[42] L. Bourtoule, V. Chandrasekaran, C. A. Choquette-Choo, H. Jia, A. Travers, B. Zhang, D. Lie and N. Papernot. "Machine Unlearning." In: *2021 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2021, pp. 141–159. DOI: 10.1109/SP40001.2021.00019. URL: https://www.researchgate.net/profile/David-Lie-4/publication/337855762_Machine_Unlearning/links/617993aea767a03c14be4702/Machine-Unlearning.pdf (cit. on p. 59).

[43] R. J. Brachman, R. E. Fikes and H. J. Levesque. "Krypton: A Functional Approach to Knowledge Representation." In: *Computer* 16.10 (Oct. 1983), pp. 67–73. DOI: 10.1109/MC.1983.1654200 (cit. on p. 20).

[44] S. Braun, A. Schmidt and C. Hentschel. "Semantic Desktop Systems for Context Awareness – Requirements and Architectural Implications." In: *1st Workshop on Architecture, Design, and Implementation of the Semantic Desktop (SemDesk Design), 4th European Semantic Web Conference (ESWC 2007), Innsbruck, Austria*. 2007. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.80.1263&rep=rep1&type=pdf (cit. on pp. 21, 64).

[45] T. Bray. *The JavaScript Object Notation (JSON) Data Interchange Format*. RFC 8259. RFC Editor, Dec. 2017. DOI: 10.17487/RFC8259. URL: https://www.rfc-editor.org/rfc/rfc8259.html (cit. on pp. xxx, 25).

[46] R. Brennan, J. Attard and M. Helfert. "Management of Data Value Chains, a Value Monitoring Capability Maturity Model." In: *Proceedings of the 20th International Conference on Enterprise Information Systems - Volume 2: ICEIS*. INSTICC. SciTePress, 2018, pp. 573–584. DOI: 10.5220/0006684805730584. URL: https://pdfs.semanticscholar.org/4bb3/9c2a00c8a95476d8effbe0304aa02b506da6.pdf (cit. on p. 126).

[47] R. Brennan, J. Attard, P. Petkov, T. Nagle and M. Helfert. "Exploring Data Value Assessment: A Survey Method and Investigation of the Perceived Relative Importance of Data Value Dimensions." In: *Proceedings of the 21st International Conference on Enterprise Information Systems - Volume 1: ICEIS*. INSTICC. SciTePress, 2019, pp. 200–207. DOI: 10.5220/0007723402000207. URL: https://pdfs.semanticscholar.org/e566/26c76b06d66758fb79c206c5ed733ecafbd8.pdf (cit. on p. 126).

[48] J. Broekstra, A. Kampman and F. van Harmelen. "Sesame: A Generic Architecture for Storing and Querying RDF and RDF Schema." In: *The Semantic Web – ISWC 2002, Proceedings of the First International Semantic Web Conference, Sardinia, Italy, June 2002*. Ed. by I. Horrocks and J. Hendler. Springer Berlin Heidelberg, 2002, pp. 54–68. DOI: 10.1007/3-540-48005-6_7. URL: https://link.springer.com/content/pdf/10.1007/3-540-48005-6_7.pdf (cit. on p. 71).

[49] G. Buscher, A. Dengel, L. van Elst and F. Mittag. "Generating and Using Gaze-Based Document Annotations." In: *CHI '08 Extended Abstracts on Human Factors in Computing Systems (CHI EA '08), Florence, Italy*. ACM, 2008, pp. 3045–3050. DOI: 10.1145/1358628.1358805 (cit. on p. 53).

[50] V. Bush. "As We May Think." In: *Atlantic Monthly* 176.1 (Mar. 1945), pp. 641–649. DOI: 10.1145/227181.227186. URL: https://www.ias.ac.in/article/fulltext/reso/005/11/0094-0103 (cit. on pp. 21, 51).

[51] Y. Cao and J. Yang. "Towards Making Systems Forget with Machine Unlearning." In: *2015 IEEE Symposium on Security and Privacy*. IEEE, 2015, pp. 463–480. DOI: 10.1109/SP.2015.35. URL: https://www.ieee-security.org/TC/SP2015/papers-archived/6949a463.pdf (cit. on p. 59).

[52] S. K. Card, T. P. Moran and A. Newell. *The Psychology of Human-Computer Interaction*. CRC Press, 1983. ISBN: 9780203736166. DOI: 10.1201/9780203736166 (cit. on pp. xxx, 10, 36).

[53] S. K. Card, G. G. Robertson and J. D. Mackinlay. "The Information Visualizer, an Information Workspace." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '91), New Orleans, Louisiana, USA*. ACM, 1991, pp. 181–186. DOI: 10.1145/108844.108874. URL: https://dl.acm.org/doi/pdf/10.1145/108844.108874 (cit. on p. 97).

[54] J. Carroll and G. Klyne. *Resource Description Framework (RDF): Concepts and Abstract Syntax*. W3C Recommendation. W3C, Feb. 2004. URL: https://www.w3.org/TR/2004/REC-rdf-concepts-20040210/ (cit. on pp. xxxi, 17, 19).

[55] J. Caumanns. *A fast and simple stemming algorithm for German words*. Tech. rep. TR B 99-16. Center für Digitale Systeme, Freie Universität Berlin, 1999. URL: https://refubium.fu-berlin.de/bitstream/handle/fub188/18405/tr-b-99-16.pdf?sequence=1. (Implementation available as *GermanStemmer* in *Apache Lucene* [Sw14].) (Cit. on pp. 109, 264).

[56] P. Chamberlain. "Knowledge is not everything." In: *Design for Health* 4.1 (2020), pp. 1–3. DOI: 10.1080/24735132.2020.1731203. URL: https://www.tandfonline.com/doi/epub/10.1080/24735132.2020.1731203?needAccess=true (cit. on pp. 14, 15).

[57] D. D. Chamberlin and R. F. Boyce. "SEQUEL: A Structured English Query Language." In: *Proceedings of the 1974 ACM SIGFIDET (Now SIGMOD) Workshop on Data Description, Access and Control*. SIGFIDET '74. ACM, 1974, pp. 249–264. ISBN: 9781450374156. DOI: 10.1145/800296.811515. URL: https://dl.acm.org/doi/pdf/10.1145/800296.811515 (cit. on pp. xxxi, 95).

[58] J. Y. C. Chen, K. Procci, M. Boyce, J. Wright, A. Garcia and M. Barnes. *Situation awareness-based agent transparency*. Tech. rep. ARL-TR-6905. U.S. Army Research Laboratory, Aberdeen Proving Ground, MD 21005-5425, Apr. 2014. URL: https://apps.dtic.mil/sti/pdfs/ADA600351.pdf (cit. on pp. 205, 207).

[59] M. Chen, S. Mao and Y. Liu. “Big Data: A Survey.” In: *Mobile Networks and Applications* 19.2 (Jan. 2014), pp. 171–209. DOI: 10.1007/s11036-013-0489-0. URL: http://cs.unibo.it/~danilo.montesi/CBD/Articoli/SurveyBigData.pdf (cit. on p. 15).

[60] A. Cheyer, J. Park and R. Giuli. “IRIS: Integrate. Relate. Infer. Share.” In: *Proceedings of the ISWC 2005 Workshop on The Semantic Desktop - Next Generation Information Management & Collaboration Infrastructure. Galway, Ireland, November 6, 2005*. Ed. by S. Decker, J. Park, D. Quan and L. Sauermann. Vol. 175. CEUR Workshop Proceedings. CEUR-WS.org, 2005, pp. 59–73. URL: http://ceur-ws.org/Vol-175/17_park_iris_final.pdf (cit. on p. 55).

[61] M. Chin. *Microsoft kicks Cortana out of the boot experience for Windows 11*. Online. TheVerge.com, June 2021. URL: https://www.theverge.com/2021/6/24/22548899/microsoft-windows-11-cortana-experience-taskbar (cit. on p. 55).

[62] P. A. Chirita, S. Costache, J. Gaugaz and W. Nejdl. “Desktop Context Detection Using Implicit Feedback.” In: *SIGIR 2006 Workshop on Personal Information Management, Seattle WA, USA, 10-11.08.06*. 2006. URL: http://pim.ischool.washington.edu/pim06/files/chirita-paper.pdf (cit. on p. 53).

[63] J. Chwalek. “Advanced Memory Buoyancy for Forgetful Information Systems.” Bachelor’s thesis. TU Kaiserslautern (in cooperation with DFKI), 2018 (cit. on pp. 95, 120, 132, 133, 212, 226).

[64] J. Chwalek. “Context Mining.” Student research project complemented by seminar paper. TU Kaiserslautern (in cooperation with DFKI), 2019 (cit. on pp. 81, 84, 95).

[65] J. Chwalek. “Context Mining on User Activity Streams.” Master’s thesis. TU Kaiserslautern (in cooperation with DFKI), 2021 (cit. on pp. 95, 117, 119–124, 212, 225).

[66] Q. Clark. *Update to the Update*. Microsoft Docs, Blogs Archive, What’s in Store: WinFS Team Blog. Microsoft Corporation, June 2006. URL: https://docs.microsoft.com/en-us/archive/blogs/winfs/update-to-the-update (cit. on p. 55).

[67] E. F. Codd. “A Relational Model of Data for Large Shared Data Banks.” In: *Communications of the ACM* 13.6 (June 1970), pp. 377–387. DOI: 10.1145/362384.362685. URL: https://dl.acm.org/doi/pdf/10.1145/362384.362685 (cit. on pp. xxxi, 71).

[68] J. Cohen. *Statistical power analysis for the behavioral sciences*. Academic Press, 1977. ISBN: 9781483276489 (cit. on p. 164).

[69] S. Costache, J. Gaugaz, E. Ioannou, C. Niederée and W. Nejdl. “Detecting Contexts on the Desktop Using Bayesian Networks.” In: 2010. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.297.1970&rep=rep1&type=pdf (cit. on p. 84).

[70] F. Crestani. "Application of Spreading Activation Techniques in Information Retrieval." In: *Artificial Intelligence Review* 11.6 (1997), pp. 453–482. DOI: 10.1023/A:1006569829653. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.105.5594&rep=rep1&type=pdf (cit. on pp. 35, 128, 218, 219).

[71] M. Crispin. *Interactive Mail Access Protocol: Version 2*. RFC 1064. RFC Editor, July 1988. DOI: 10.17487/RFC1064. URL: https://www.rfc-editor.org/rfc/rfc1064.html (cit. on pp. xxx, 81, 153, 261).

[72] D. Crockford. *The application/json Media Type for JavaScript Object Notation (JSON)*. RFC 4627. RFC Editor, July 2006. DOI: 10.17487/RFC4627. URL: https://www.rfc-editor.org/rfc/rfc4627.html (cit. on pp. xxx, 25).

[73] R. Cyganiak, D. Wood and M. Lanthaler. *RDF 1.1 Concepts and Abstract Syntax*. W3C Recommendation. W3C, Feb. 2014. URL: https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/ (cit. on pp. xxxi, 17, 19).

[74] C. Daboo. *CardDAV: vCard Extensions to Web Distributed Authoring and Versioning (WebDAV)*. RFC 6352. RFC Editor, Aug. 2011. DOI: 10.17487/RFC6352. URL: https://www.rfc-editor.org/rfc/rfc6352.html (cit. on pp. xxix, 29, 153).

[75] C. Daboo, B. Desruisseaux and L. Dusseault. *Calendaring Extensions to WebDAV (CalDAV)*. RFC 4791. RFC Editor, Mar. 2007. DOI: 10.17487/RFC4791. URL: https://www.rfc-editor.org/rfc/rfc4791.html (cit. on pp. xxix, 29, 153, 196).

[76] J. Daiber, M. Jakob, C.Hokamp and P. N. Mendes. "Improving Efficiency and Accuracy in Multilingual Entity Extraction." In: *Proceedings of the 9th International Conference on Semantic Systems*. I-SEMANTICS '13. ACM, Sept. 2013, pp. 121–124. DOI: 10.1145/2506182.2506198. URL: http://jodaiber.github.io/doc/entity.pdf. (Implementation available as *DBpedia Spotlight* [Sw74].) (Cit. on p. 269).

[77] T. H. Davenport. *Thinking for a Living: How to Get Better Performances and Results from Knowledge Workers*. Harvard Business School Press, 2005. ISBN: 1-59139-423-6 (cit. on p. 16).

[78] F. Dawson and D. Stenerson. *Internet Calendaring and Scheduling Core Object Specification (iCalendar)*. RFC 2445. RFC Editor, Nov. 1998. DOI: 10.17487/RFC2445. URL: https://www.rfc-editor.org/rfc/rfc2445.html (cit. on p. 81).

[79] R. De La Briandais. "File Searching Using Variable Length Keys." In: *Papers Presented at the the March 3-5, 1959, Western Joint Computer Conference, IRE-AIEE-ACM '59 (Western), San Francisco, California*. ACM, 1959, pp. 295–298. DOI: 10.1145/1457838.1457895. URL: https://dl.acm.org/doi/pdf/10.1145/1457838.1457895 (cit. on p. 102).

[80] S. Decker and M. Frank. *The Social Semantic Desktop*. DERI Technical Report 2004-05-02. Digital Enterprise Research Institute (DERI), May 2004. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.105.1289&rep=rep1&type=pdf (cit. on pp. xxix, xxxi, 4–6, 17, 18, 20, 25, 51, 91).

[81] A. Dengel. “Knowledge technologies for the social semantic desktop.” In: *Knowledge Science, Engineering and Management*. Springer, 2007, pp. 2–9. DOI: 10.1007/978-3-540-76719-0_2. URL: https://www.researchgate.net/profile/Andreas-Dengel/publication/221143596_Knowledge_Technologies_for_the_Social_Semantic_Desktop/links/54631b820cf2cb7e9da67741/Knowledge-Technologies-for-the-Social-Semantic-Desktop.pdf (cit. on p. 52).

[82] A. Dengel. “Einleitung [Introduction].” In: *Semantische Technologien: Grundlagen – Konzepte – Anwendungen [Semantic Technologies: Foundations – Concepts – Applications]*. Ed. by A. Dengel. Spektrum Akademischer Verlag, 2012, pp. 3–19. ISBN: 978-3-8274-2664-2. DOI: 10.1007/978-3-8274-2664-2_1 (cit. on p. 13).

[83] A. Dengel. “Wissensrepräsentation [Knowledge Representation].” In: *Semantische Technologien: Grundlagen – Konzepte – Anwendungen [Semantic Technologies: Foundations – Concepts – Applications]*. Ed. by A. Dengel. Spektrum Akademischer Verlag, 2012, pp. 21–72. ISBN: 978-3-8274-2664-2. DOI: 10.1007/978-3-8274-2664-2_2 (cit. on p. 20).

[84] A. Dengel, A. Abecker, A. Bernardi, L. van Elst, H. Maus, S. Schwarz and M. Sintek. “Konzepte zur Gestaltung von Unternehmensgedächtnissen [Design Concepts for Organizational Memories].” In: *KI – Künstliche Intelligenz, German Journal of Artificial Intelligence* 16.1 (2002), pp. 5–11. URL: https://www.dfki.uni-kl.de/~elst/papers/KI-Heft_OM-Prinzipien-draft.pdf (cit. on pp. 15, 16).

[85] A. Dengel and H. Maus. “Personalisierte Wissensdienste: Das Unternehmen denkt mit [Personalized Knowledge Services: The Enterprise Thinks for Itself].” In: *IM+io* 3 (Sept. 2018), pp. 46–49. URL: https://www.aws-institut.de/im-io/kuenstliche-intelligenz/personalisierte-wissensdienste-das-unternehmen-denkt-mit/ (cit. on pp. 24, 151).

[86] A. Dengel and H. Maus. “Ein ’Informationsbutler’ – mit Talent für smarte Daten [An ’Information Butler’ – with a Talent for Smart Data].” In: *Digitus: Digital Business, Strategien, Technologien, Wirtschaft, Unternehmen & Industrie 4.0 (DIGITUS) [Digitus: Digital Business, Strategies, Technologies, Economy, Companies & Industry 4.0 (DIGITUS)]* 1 (Feb. 2019), pp. 22–27. URL: https://digitusmagazin.de/2019/02/ein-informationsbutler-mit-talent-fuer-smarte-daten/ (cit. on pp. 24, 151, 213).

[87] A. Dengel, H. Maus, S. Schwarz, C. Jilek, M. Schröder, M. Schulze, A. Lauer, R. Koch and E. Baitemirov. *CoMem: a group-wide corporate memory ecosystem*. Web documentation/showcase. Smart Data & Knowledge Services (SDS) Research Department, DFKI, 2018. URL: https://comem.ai/ (cit. on pp. xxix, 21, 24, 26, 49, 50, 54, 63, 64, 233, 319).

[88] A. R. Dengel. “Six Thousand Words about Multi-Perspective Personal Document Management.” In: *2006 10th IEEE International Enterprise Distributed Object Computing Conf. Workshops (EDOCW’06)*. 2006, pp. 62–62. DOI: 10.1109/EDOCW.2006.63 (cit. on pp. 18, 52, 73, 82, 141).

[89] B. Desruisseaux. *Internet Calendaring and Scheduling Core Object Specification (iCalendar)*. RFC 5545. RFC Editor, Sept. 2009. DOI: 10.17487/RFC5545. URL: https://www.rfc-editor.org/rfc/rfc5545.html (cit. on p. 81).

[90] D. Devaurs, A. S. Rath and S. N. Lindstaedt. "Exploiting the user interaction context for automatic task detection." In: *Applied Artificial Intelligence* 26.1-2 (2012), pp. 58–80. DOI: 10.1080/08839514.2012.629522. URL: https://www.tandfonline.com/doi/pdf/10.1080/08839514.2012.629522 (cit. on pp. 56, 116).

[91] A. Dey and B. S. Prukayastha. "Named Entity Recognition using Gazetteer Method and N-gram Technique for an Inflectional Language: A Hybrid Approach." In: *International Journal of Computer Applications* 84.9 (2013). URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.402.4329&rep=rep1&type=pdf (cit. on p. 100).

[92] A. Dimou, M. Vander Sande, P. Colpaert, R. Verborgh, E. Mannens and R. Van de Walle. "RML: A Generic Language for Integrated RDF Mappings of Heterogeneous Data." In: *Proceedings of the Workshop on Linked Data on the Web co-located with the 23rd International World Wide Web Conference (WWW 2014), Seoul, Korea, April 8, 2014*. Ed. by C. Bizer, T. Heath, S. Auer and T. Berners-Lee. Vol. 1184. CEUR Workshop Proceedings. CEUR-WS.org, 2014. URL: http://ceur-ws.org/Vol-1184/ldow2014_paper_01.pdf (cit. on p. 89).

[93] Š. Dlugolinskỳ, G. Nguyen, M. Laclavík and M. Šeleng. "Character gazetteer for Named Entity Recognition with linear matching complexity." In: *3rd World Congress on Information and Communication Technologies (WICT)*. IEEE. 2013, pp. 361–365. DOI: 10.1109/WICT.2013.7113096. URL: http://ikt.ui.sav.sk/archive/vega15/AEC14_Dlugolinsky_Giang_gazetteer.pdf. (Implementation available as *IKT Gazetteers* [Sw93].) (Cit. on pp. 100, 109–113, 270).

[94] X. Dong, A. Halevy, E. Nemes, S. Sigurdsson and P. Domingos. "SEMEX: Toward On-the-fly Personal Information Integration." In: *Workshop on Information Integration on the Web (IIWEB)*. 2004. URL: https://pal.sri.com/wp-content/uploads/publications/calo/2004/semex_iiweb.pdf (cit. on p. 55).

[95] J. Dorbolo. *Dr. Tech: Learning Curves Ahead*. Weblog. Jan. 2016. URL: https://jondorbolo.wordpress.com/2016/01/27/dr-tech-learning-curves-ahead/ (cit. on p. 14).

[96] L. Dragan and S. Decker. "Knowledge Management on the Desktop." In: *Knowledge Engineering and Knowledge Management*. Springer, 2012, pp. 373–382. DOI: 10.1007/978-3-642-33876-2_33 (cit. on pp. 20–22, 51, 57, 63, 64, 81, 92).

[97] A. N. Dragunov, T. G. Dietterich, K. Johnsrude, M. McLaughlin, L. Li and J. L. Herlocker. "TaskTracer – A Desktop Environment to Support Multi-Tasking Knowledge Workers." In: *Proceedings of the 10th International Conference on Intelligent User Interfaces (IUI '05), San Diego, California, USA*. ACM, 2005, pp. 75–82. DOI: 10.1145/1040830.1040855. URL: https://web.engr.oregonstate.edu/~tgd/publications/iui2005-tasktracer.pdf (cit. on pp. 56, 116).

[98] S. Dumais, E. Cutrell, J. J. Cadiz, G. Jancke, R. Sarin and D. C. Robbins. "Stuff I've Seen: A System for Personal Information Retrieval and Re-use." In: *Proceedings of 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval (SIGIR'03)*. ACM, 2003, pp. 72–79. DOI: 10.1145/860435.860451. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.440.5787&rep=rep1&type=pdf (cit. on p. 55).

[99] L. Dusseault. *HTTP Extensions for Web Distributed Authoring and Versioning (WebDAV)*. RFC 4918. RFC Editor, June 2007. DOI: 10.17487/RFC4918. URL: https://www.rfc-editor.org/rfc/rfc4918.html (cit. on pp. xxxi, 25, 29, 153).

[100] H. Ebbinghaus. *Memory: A contribution to experimental psychology*. Trans. by H. A. Ruger and C. E. Bussenius. (Translated from *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie*, Duncker & Humblot, Leipzig, 1885). Columbia University Teachers College Press, 1913. DOI: 10.1037/10011-000. URL: https://archive.org/details/memorycontributi00ebbiuoft/ (cit. on p. 183).

[101] ECMA International. *ECMA-262, 1st edition, June 1997 – ECMAScript: A general purpose, cross-platform programming language*. ECMA Standards. ECMA International, June 1997. URL: https://www.ecma-international.org/wp-content/uploads/ECMA-262_1st_edition_june_1997.pdf. Also see https://www.ecma-international.org/publications-and-standards/standards/ecma-262/ (cit. on pp. xxx, 95).

[102] ECMA International. *ECMA-262, 12th edition, June 2021 – ECMAScript 2021 Language Specification*. ECMA Standards. ECMA International, June 2021. URL: https://www.ecma-international.org/wp-content/uploads/ECMA-262_12th_edition_june_2021.pdf. Also see https://www.ecma-international.org/publications-and-standards/standards/ecma-262/ (cit. on pp. xxx, 95).

[103] T. Eiter and G. Kern-Isberner. "A Brief Survey on Forgetting from a Knowledge Representation and Reasoning Perspective." In: *KI – Künstliche Intelligenz, German Journal of Artificial Intelligence - Organ des Fachbereichs "Künstliche Intelligenz" der Gesellschaft für Informatik e.V.* 33.1 (Mar. 2019), pp. 9–33. DOI: 10.1007/s13218-018-0564-6. URL: http://www.kr.tuwien.ac.at/staff/eiter/et-archive/forgetting_ki_aam.pdf (cit. on p. 59).

[104] B. Eldesouky, M. Bakry, H. Maus and A. Dengel. "Supporting early contextualization of textual content in digital documents on the Web." In: *13th International Conference on Document Analysis and Recognition (ICDAR), Tunis, Tunisia, August 23-26, 2015*. 2015, pp. 1071–1075. DOI: 10.1109/ICDAR.2015.7333926 (cit. on p. 29).

[105] B. Eldesouky, M. Bakry, H. Maus and A. Dengel. "Seed, an End-User Text Composition Tool for the Semantic Web." In: *The Semantic Web – ISWC 2016, Proceedings of the 15th International Semantic Web Conference, Kobe, Japan, October 17–21, 2016*. Ed. by P. Groth, E. Simperl, A. Gray, M. Sabou, M. Krötzsch, F. Lecue, F. Flöck and Y. Gil. Springer, 2016, pp. 218–233. DOI: 10.1007/978-3-319-46523-4_14. URL: https://link.springer.com/content/pdf/10.1007/978-3-319-46523-4_14.pdf (cit. on p. 29).

[106] B. Eldesouky, H. Maus, S. Schwarz and A. Dengel. "Seed, a Natural Language Interface to Knowledge Bases." In: *Human Interface and the Management of Information. Information and Knowledge Design, 17th International Conference, HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015, Proceedings*. Springer, 2015, pp. 280–290. DOI: 10.1007/978-3-319-20612-7_27 (cit. on p. 29).

[107] T. Ellwart and A. Kluge. "Psychological Perspectives on Intentional Forgetting: An Overview of Concepts and Literature." In: *KI – Künstliche Intelligenz, German Journal of Artificial Intelligence - Organ des Fachbereichs "Künstliche Intelligenz" der Gesellschaft für Informatik e.V.* 33.1 (Mar. 2019), pp. 79–84. DOI: 10.1007/s13218-018-00571-0 (cit. on p. 58).

[108] T. Ellwart, A.-S. Ulfert, C. H. Antoni, J. Becker, C. Frings, K. Göbel, G. Hertel, A. Kluge, S. M. Meeßen, C. Niessen, C. Nohe, D. M. Riehle, Y. Runge, U. Schmid, A. Schüffler, M. Siebers, S. Sonnentag, T. Tempel, M. T. Thielsch and W. Wehrt. "Intentional Forgetting in Socio-Digital Work Systems." In: *AIS Transactions on Enterprise Systems* 4.1 (May 2019). DOI: 10.30844/aistes.v4i1.16. URL: https://www.aes-journal.com/index.php/ais-tes/article/view/16/14 (cit. on p. 58).

[109] envia Mitteldeutsche Energie AG. *enviaM – Ihr regionaler Energieversorger*. Corporate website. 2022. URL: https://www.enviam.de/ (cit. on pp. 26, 89, 134, 151, 212, 213).

[110] EPOS Project Team. *EPOS – Evolving Personal to Organizational Knowledge Spaces*. Project Website. 2003. URL: https://www.dfki.uni-kl.de/epos/. Also see https://www.dfki.de/web/forschung/projekte-publikationen/projekte-uebersicht/projekt/epos (cit. on pp. 23, 52, 233).

[111] V. Ermolayev, C. Ruiz, M. Tilly, E. Jentzsch, J. M. Gomez-Perez and W.-E. Matzke. "A Context Model for Knowledge Workers." In: *Proceedings of the Second Workshop on Context, Information and Ontologies (CIAO-2010), Lisbon, Portugal, October 11, 2010*. Ed. by V. Ermolayev, J. M. Gomez-Perez, P. Haase and P. Warren. Vol. 626. CEUR Workshop Proceedings. CEUR-WS.org, 2010. URL: http://ceur-ws.org/Vol-626/regular2.pdf (cit. on pp. 56, 91, 116).

[112] European Commission. *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance)*. 2016. URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj (cit. on pp. xxx, 57, 59).

[113] A. Eustace. *A fall spring-clean*. Google Official Blog, Blogspot.com. Sept. 2011. URL: https://googleblog.blogspot.com/2011/09/fall-spring-clean.html (cit. on pp. 55, 64).

[114] J. D. Fernández, A. Llaves and O. Corcho. "Efficient RDF Interchange (ERI) Format for RDF Data Streams." In: *The Semantic Web – ISWC 2014, 13th International Semantic Web Conference, Riva del Garda, Italy, October 19-23, 2014. Proceedings, Part II*. Ed. by P. Mika, T. Tudorache, A. Bernstein, C. Welty, C. Knoblock, D. Vrandečić, P. Groth, N. Noy, K. Janowicz and C. Goble. Springer, 2014, pp. 244–259. DOI: 10.1007/978-3-319-11915-1_16. URL: https://link.springer.com/content/pdf/10.1007/978-3-319-11915-1_16.pdf (cit. on p. 79).

[115] M. Ferrari and V. Quaresima. "A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application." In: *NeuroImage* 63.2 (2012), pp. 921–935. DOI: 10.1016/j.neuroimage.2012.03.049. URL: https://iss.com/resources/pdf/publications/fnirs-history.pdf (cit. on p. 166).

[116] I. Fette and A. Melnikov. *The WebSocket Protocol*. RFC 6455. RFC Editor, Dec. 2011. DOI: 10.17487/RFC6455. URL: https://www.rfc-editor.org/rfc/rfc6455.html (cit. on pp. 148, 154).

[117] R. Fielding and J. Reschke. *Hypertext Transfer Protocol (HTTP/1.1): Message Syntax and Routing*. RFC 7230. RFC Editor, June 2014. DOI: 10.17487/RFC7230. URL: https://www.rfc-editor.org/rfc/rfc7230.html (cit. on pp. xxx, 154).

[118] R. T. Fielding. "Architectural styles and the design of network-based software architectures." PhD thesis. University of California, 2000. URL: https://www.ics.uci.edu/~fielding/pubs/dissertation/fielding_dissertation.pdf (cit. on pp. xxxi, 25, 71, 148).

[119] G. Flouris, I. Fundulaki, M. Michou and G. Antoniou. "Controlling Access to RDF Graphs." In: *Future Internet - FIS 2010, Third Future Internet Symposium, Berlin, Germany, September 20-22, 2010. Proceedings*. Ed. by A. J. Berre, A. Gómez-Pérez, K. Tutschku and D. Fensel. Springer Berlin Heidelberg, 2010, pp. 107–117. DOI: 10.1007/978-3-642-15877-3_12. URL: https://www.researchgate.net/profile/Maria_Michou/publication/221142755_Controlling_Access_to_RDF_Graphs/links/5481bb610cf263ee1adfe25d.pdf (cit. on p. 74).

[120] ForgetIT Consortium. *ForgetIT: Concise Preservation by Combining Managed Forgetting and Contextualized Remembering*. Project website. 2013. URL: https://www.forgetit-project.eu/. Also see https://cordis.europa.eu/project/id/600826 (cit. on pp. 4, 23, 30, 33, 54, 57, 134, 135, 181, 183, 319).

[121] T. Franz, A. Scherp and S. Staab. "Are Semantic Desktops Better?: Summative Evaluation Comparing a Semantic Against a Conventional Desktop." In: *Proceedings of the 5th International Conference on Knowledge Capture*. K-CAP '09. Redondo Beach, California, USA: ACM, 2009, pp. 1–8. ISBN: 978-1-60558-658-8. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.692.75&rep=rep1&type=pdf (cit. on pp. 21, 22, 53, 63, 64).

[122] T. Franz, S. Staab and R. Arndt. "The X-COSIM Integration Framework for a Seamless Semantic Desktop." In: *Proceedings of the 4th International Conference on Knowledge Capture (K-CAP '07), Whistler, BC, Canada*. ACM, 2007, pp. 143–150. DOI: 10.1145/1298406.1298433. URL: https://scholar.archive.org/work/3qkjccrgfbevdfewoadu75yxnq/access/wayback/https://userpages.uni-koblenz.de/~staab/Research/Publications/2007/x-cosim.pdf (cit. on p. 53).

[123] E. Fredkin. "Trie Memory." In: *Communications of the ACM* 3.9 (Sept. 1960), pp. 490–499. DOI: 10.1145/367390.367400. URL: https://dl.acm.org/doi/pdf/10.1145/367390.367400 (cit. on p. 102).

[124] FRODO Project Team. *FRODO – A Framework for Distributed Organizational Memories*. Project website. 2000. URL: https://www.dfki.uni-kl.de/frodo/ (cit. on p. 52).

[125] R. B. Fuller. *Critical Path*. Saint Martin's Press, New York, 1981. ISBN: 9780312174880 (cit. on pp. 14, 15).

[126] P. Gauselmann, Y. Runge, C. Jilek, C. Frings, H. Maus and T. Tempel. “A Relief from Mental Overload in a Digitalized World: How Context-sensitive User Interfaces Can Enhance Cognitive Performance.” In: *International Journal of Human-Computer Interaction* 39.1 (2023), pp. 140–150. DOI: 10.1080/10447318.2022.2041882 (cit. on pp. 33, 36, 141, 147, 163–165, 226).

[127] C. F. Gauß. *Disquisitiones arithmeticae*. Translated into English by A. A. Clark, Yale University Press, 1966. 1801. DOI: 10.1007/978-1-4939-7560-0. Original text (in Latin): http://resolver.sub.uni-goettingen.de/purl?PPN235993352 (cit. on p. 163).

[128] C. Geissler. “The role of the dorsolateral prefrontal cortex in executive functioning and action control: Insights from functional near-infrared spectroscopy.” PhD thesis. University of Trier, 2021 (cit. on p. 166).

[129] C. Geissler, P. Gauselmann, C. Jilek, H. Maus, C. Frings and T. Tempel. “A Functional Near-infrared Spectroscopy Study on the Prefrontal Correlates of Cognitive Offloading via a Personal Knowledge Assistant.” In: *Scientific Reports* 13.13938 (2023). DOI: 10.1038/s41598-023-39540-5 (cit. on pp. 166, 226).

[130] J. Gemmell, G. Bell and R. Lueder. “MyLifeBits: A Personal Database for Everything.” In: *Communications of the ACM* 49.1 (Jan. 2006), pp. 88–95. DOI: 10.1145/1107458.1107460. URL: https://www.microsoft.com/en-us/research/wp-content/uploads/2006/01/tr-2006-23.pdf (cit. on p. 55).

[131] J. Gemmell, G. Bell, R. Lueder, S. Drucker and C. Wong. “MyLifeBits: Fulfilling the Memex Vision.” In: *Proceedings of the Tenth ACM International Conference on Multimedia, MULTIMEDIA ’02, Juan-les-Pins, France*. ACM, 2002, pp. 235–238. DOI: 10.1145/641007.641053. URL: https://www.microsoft.com/en-us/research/wp-content/uploads/2002/12/mylifebitsmm02.pdf (cit. on p. 55).

[132] D. K. Gifford, P. Jouvelot, M. A. Sheldon and J. W. O’Toole Jr. “Semantic File Systems.” In: *ACM SIGOPS Operating Systems Review* 25.5 (1991), pp. 16–25. DOI: 10.1145/121133.121138. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.147.2488&rep=rep1&type=pdf (cit. on p. 61).

[133] K. Göbel, C. Niessen, D. Nandini and U. Schmid. “Die Rolle von Erklärungen in partnerschaftlichen KI-Systemen: Wem helfen Erklärungen in welchen Situationen wie gut? [The role of explanations in collaborative AI systems: Who do explanations help in which situations and how well?]” In: *Arbeit HumAIne Gestalten – 67. Frühjahrskongress der Gesellschaft für Arbeitswissenschaft e.V. [HumAIne Work Design – 67th Spring Congress of the Human Factors and Ergonomics Society] (GfA 2021), Ruhr-Universität Bochum, Online, March 3-5, 2021*. GfA-Press, 2021. ISBN: 978-3-936804-29-4 (cit. on p. 58).

[134] Y. Goland, E. Whitehead, A. Faizi, S. Carter and D. Jensen. *HTTP Extensions for Distributed Authoring – WEBDAV*. RFC 2518. RFC Editor, Feb. 1999. DOI: 10.17487/RFC2518. URL: https://www.rfc-editor.org/rfc/rfc2518.html (cit. on pp. xxxi, 25, 29, 153).

[135] J. M. Gomez-Perez, M. Grobelnik, C. Ruiz, M. Tilly and P. Warren. “Using Task Context to Achieve Effective Information Delivery.” In: *1st Workshop on Context, Information and Ontologies (CIAO ’09)*. Heraklion, Greece: ACM, 2009, 3:1–3:6. DOI: 10.1145/1552262.

1552265. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.494.600&rep=rep1&type=pdf (cit. on pp. 38, 56, 61, 62).

[136] D. Gonçalves. "Pseudo-desktop collections and PIM: The missing link." In: *ECIR 2011 Workshop on Evaluating Personal Search*. 2011, pp. 3–4. URL: http://web.ist.utl.pt/~daniel.j.goncalves/publications/2011/ecir.pdf (cit. on pp. 142, 143).

[137] V. M. González and G. Mark. "'Constant, constant, multi-tasking craziness': managing multiple working spheres." In: *SIGCHI Conference on Human factors in computing systems (CHI'04)*. Vienna, Austria: ACM, 2004, pp. 113–120. DOI: 10.1145/985692.985707. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.75.5354&rep=rep1&type=pdf (cit. on pp. 3, 17, 45, 120).

[138] J. Gordon and E. H. Shortliffe. "The Dempster-Shafer theory of evidence." In: *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project* (1984). Ed. by B. G. Buchanan and E. H. Shortliffe, pp. 272–292. URL: https://www.aaai.org/Papers/Buchanan/Buchanan15.pdf (cit. on pp. 137, 192).

[139] J. Gossman. *Introduction to Model/View/ViewModel pattern for building WPF apps*. Microsoft Docs, Blogs Archive, Tales from the Smart Client. Microsoft Corporation, Oct. 2005. URL: https://docs.microsoft.com/en-us/archive/blogs/johngossman/introduction-to-modelviewviewmodel-pattern-for-building-wpf-apps (cit. on pp. 148, 149).

[140] M. A. Greenwood, N. K. Tran, K. Apostolidis and V. Mezaris. "Keeping Information in Context." In: *Personal Multimedia Preservation: Remembering or Forgetting Images and Video*. Ed. by V. Mezaris, C. Niederée and R. H. Logie. Springer, 2018, pp. 131–182. DOI: 10.1007/978-3-319-73465-1_5 (cit. on p. 57).

[141] T. Groza, S. Handschuh, K. Moeller, G. Grimnes, L. Sauermann, E. Minack, C. Mesnage, M. Jazayeri, G. Reif and R. Gudjonsdottir. "The NEPOMUK Project - On the way to the Social Semantic Desktop." In: *Proceedings of I-Semantics' 07*. Ed. by T. Pellegrini and S. Schaffert. Journal of Universal Computer Science (JUCS), 2007, pp. 201–211. URL: http://siegfried-handschuh.net/pub/2007/nepomuk_isemantics2007.pdf (cit. on p. 53).

[142] R. Guha and D. Brickley. *RDF Vocabulary Description Language 1.0: RDF Schema*. W3C Recommendation. W3C, Feb. 2004. URL: https://www.w3.org/TR/2004/REC-rdf-schema-20040210/ (cit. on pp. 17, 20).

[143] R. Guha and D. Brickley. *RDF Schema 1.1*. W3C Recommendation. W3C, Feb. 2014. URL: https://www.w3.org/TR/2014/REC-rdf-schema-20140225/ (cit. on pp. 17, 20).

[144] V. Gupta, C. Jung, S. Neel, A. Roth, S. Sharifi-Malvajerdi and C. Waites. "Adaptive Machine Unlearning." In: *Advances in Neural Information Processing Systems*. Ed. by M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang and J. Wortman Vaughan. Vol. 34. Curran Associates, 2021, pp. 16319–16330. URL: https://proceedings.neurips.cc/paper/2021/file/87f7ee4fdb57bdfd52179947211b7ebb-Paper.pdf (cit. on p. 59).

[145] C. Gurrin, H. Lee and J. Hayes. "iForgot: A model of forgetting in robotic memories." In: *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2010, pp. 93–94. DOI: 10.1109/HRI.2010.5453255 (cit. on p. 59).

[146] C. Gurrin, A. F. Smeaton and A. R. Doherty. "LifeLogging: Personal Big Data." In: *Foundations and Trends in Information Retrieval* 8.1 (June 2014), pp. 1–125. DOI: 10.1561/1500000033 (cit. on pp. 59, 81).

[147] C. Gutierrez, C. Hurtado and A Vaisman. "Temporal RDF." In: *The Semantic Web: Research and Applications*. Ed. by A. Gómez-Pérez and J. Euzenat. Springer Berlin Heidelberg, 2005, pp. 93–107. DOI: 10.1007/11431053_7. URL: https://link.springer.com/content/pdf/10.1007/11431053_7.pdf (cit. on p. 73).

[148] K. Gyllstrom. "Enriching personal information management with document interaction histories." PhD thesis. University of North Carolina, 2009. URL: https://cdr.lib.unc.edu/downloads/6t053h383 (cit. on p. 56).

[149] K. Gyllstrom and E. R. Pedersen. "Not gone, but forgotten: Helping users re-find web pages by identifying those which are most likely to be lost." In: *Proceedings of the Workshop on Desktop Search at SIGIR* (2010). URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.308.6&rep=rep1&type=pdf (cit. on pp. 56, 125).

[150] K. Gyllstrom, C. Soules and A. Veitch. "Activity Put in Context: Identifying Implicit Task Context within the User's Document Interaction." In: *Proceedings of the Second International Symposium on Information Interaction in Context (IIiX '08), London, United Kingdom*. ACM, 2008, pp. 51–56. DOI: 10.1145/1414694.1414707. URL: http://www.cs.unc.edu/techreports/08-007.pdf (cit. on pp. 56, 116).

[151] K. Gyllstrom and D. Stotts. *Unsupervised Task Extraction from a Stream of Window Focus Events*. Tech. rep. TR06-023. University of North Carolina, 2006. URL: http://www.cs.unc.edu/techreports/06-023.pdf (cit. on pp. 56, 116).

[152] J. Haas, H. Maus, S. Schwarz and A. Dengel. "ConTask – Using Context-sensitive Assistance to Improve Task-oriented Knowledge Work." In: *Proceedings of the 12th International Conference on Enterprise Information Systems – Volume 2: ICEIS*. INSTICC. SciTePress, 2010, pp. 30–39. DOI: 10.5220/0002895400300039. URL: https://www.scitepress.org/Papers/2010/28954/28954.pdf (cit. on pp. 38, 54, 60, 65).

[153] E. Hall. *The application/mbox Media Type*. RFC 4155. RFC Editor, Sept. 2005. DOI: 10.17487/RFC4155. URL: https://www.rfc-editor.org/rfc/rfc4155.html (cit. on p. 81).

[154] V. Handa. *Nepomuk in 4.13 and beyond*. Online. KDE Mailing Lists, The Nepomuk Archives, Message 004858, Dec. 2013. URL: https://mail.kde.org/pipermail/nepomuk/2013-December/004858.html (cit. on pp. 5, 22, 63).

[155] O. Hartig, P.-A. Champin, G. Kellogg and A. Seaborne. *RDF-star and SPARQL-star*. W3C Community Group Specification. W3C, Dec. 2021. URL: https://w3c.github.io/rdf-star/cg-spec/ (cit. on p. 74).

[156] R. Hasan and R. Burns. "The life and death of unwanted bits: Towards proactive waste data management in digital ecosystems." In: *Third International Conference on Innovative Computing Technology (INTECH 2013)*. 2013, pp. 144–148. DOI: 10.1109/INTECH.2013.6653665. arXiv: 1106.6062v2. URL: https://arxiv.org/pdf/1106.6062v2.pdf (cit. on p. 125).

[157] M. A. Hearst. "Automatic Acquisition of Hyponyms from Large Text Corpora." In: *Proceedings of the 14th Conference on Computational Linguistics - Vol. 2*. Association for Computational Linguistics, 1992, pp. 539–545. DOI: 10.3115/992133.992154. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.36.701&rep=rep1&type=pdf (cit. on p. 102).

[158] L. V. Hedges. "Distribution Theory for Glass's Estimator of Effect size and Related Estimators." In: *Journal of Educational Statistics* 6.2 (1981), pp. 107–128. DOI: 10.3102/10769986006002107 (cit. on p. 171).

[159] D. Heim. "Contextifier: Retrospective Context Mining on (Big) Personal Data." Bachelor's thesis. TU Kaiserslautern (in cooperation with DFKI), 2021 (cit. on pp. 81, 84, 86–88, 95, 225, 241, 242).

[160] D. Heim, C. Jilek, H. Maus and A. Dengel. "A Retrospective Context Mining Approach for Bootstrapping Personal Knowledge Assistants." In: *Proceedings of the "Lernen, Wissen, Daten, Analysen – Learning, Knowledge, Data, Analytics" 2022 Workshops: FGWM, FGKD and FGDB (LWDA 2022), Hildesheim, Germany, October 5-7, 2022*. Ed. by P. Reuss, V. Eisenstadt, J. Schönborn and J. Schäfer. Vol. 3341. CEUR Workshop Proceedings. CEUR-WS.org, 2022. URL: https://ceur-ws.org/Vol-3341/WM-LWDA_2022_CRC_8910.pdf (cit. on pp. 69, 86, 88, 225).

[161] I. Heizer, P. Leach and D. Perry. *Common Internet File System Protocol (CIFS/1.0)*. Originally online at https://tools.ietf.org/html/draft-heizer-cifs-v1-spec-00 (not available anymore). Archived copy of 2019-08-08. Microsoft Corporation, June 1996. URL: https://web.archive.org/web/20190808115512/https://tools.ietf.org/html/draft-heizer-cifs-v1-spec-00 (cit. on pp. xxxi, 25, 153, 196).

[162] I. Herman. *Semantic Web Activity Statement*. W3C Activity Statements. W3C, 2013. URL: https://www.w3.org/2001/sw/Activity (cit. on p. 19).

[163] S. Hertling, M. Schröder, C. Jilek and A. Dengel. "Top-k Shortest Paths in Directed Labeled Multigraphs." In: *Semantic Web Challenges – Third SemWebEval Challenge at ESWC 2016, Heraklion, Crete, Greece, May 29 - June 2, 2016, Revised Selected Papers*. Ed. by H. Sack, S. Dietze, A. Tordai and C. Lange. Springer, 2016, pp. 200–212. DOI: 10.1007/978-3-319-46565-4_16 (cit. on pp. 80, 320).

[164] S. Hertling, M. Schröder, C. Jilek and A. Dengel. "Where is that Button again ?! – Towards a Universal GUI Search Engine." In: *Proceedings of the 9th International Conference on Agents and Artificial Intelligence (ICAART 2017), Porto, Portugal, February 24-26, 2017*. Vol. 2. SciTePress, 2017, pp. 217–227. DOI: 10.5220/0006201402170227. URL: https://www.scitepress.org/papers/2017/62014/62014.pdf (cit. on pp. 93, 320).

[165] F. Heylighen and C. Gershenson. "The Meaning of Self-organization in Computing." In: *IEEE Intelligent Systems* 18.4 (2003). URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.4.1576&rep=rep1&type=pdf (cit. on p. 43).

[166] D. R. Hipp and SQLite Team. *SQLite*. Online. Aug. 2000. URL: https://sqlite.org/ (cit. on pp. 77, 81, 82, 96).

[167] P. Hitzler, M. Krötzsch, B. Parsia, P. F. Patel-Schneider and S. Rudolph. *OWL 2 Web Ontology Language Primer (Second Edition)*. W3C Recommendation. W3C, Dec. 2012. URL: https://www.w3.org/TR/owl-primer (cit. on pp. 17, 20).

[168] A. Hogan, E. Blomqvist, M. Cochez, C. D'amato, G. De Melo, C. Gutierrez, S. Kirrane, J. E. L. Gayo, R. Navigli, S. Neumaier, A.-C. N. Ngomo, A. Polleres, S. M. Rashid, A. Rula, L. Schmelzeisen, J. Sequeda, S. Staab and A. Zimmermann. "Knowledge Graphs." In: *ACM Computing Surveys* 54.4 (July 2021). DOI: 10.1145/3447772. URL: https://dl.acm.org/doi/pdf/10.1145/3447772 (cit. on pp. xxx, 6).

[169] H. Holz, H. Maus, A. Bernardi and O. Rostanin. "A Lightweight Approach for Proactive, Task-Specific Information Delivery." In: *Proceedings of I-KNOW'05 – Special Track on Business Process Oriented Knowledge Infrastructures BPOKI'05. International Conference on Knowledge Management and New Media Technology (I-KNOW)*. 2005. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.88.7219&rep=rep1&type=pdf (cit. on p. 52).

[170] H. Holz, H. Maus, A. Bernardi and O. Rostanin. "From Lightweight, Proactive Information Delivery to Business Process-Oriented Knowledge Management." In: *Journal of Universal Knowledge Management. Special Issue on Knowledge Infrastructures for the Support of Knowledge Intensive Business Processes* 0.2 (2005), pp. 101–127. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.63.8027&rep=rep1&type=pdf (cit. on p. 52).

[171] E. Horvitz, J. Breese, D. Heckerman, D. Hovel and K. Rommelse. "The Lumière Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users." In: *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, Madison WI, July, 1998* (1998), pp. 256–265. URL: https://www.microsoft.com/en-us/research/wp-content/uploads/2016/11/lum.pdf. (Related to the *Microsoft Office Assistant* [Sw81].) (Cit. on pp. 55, 269).

[172] E. Horvitz, S. Dumais and P. Koch. "Learning predictive models of memory landmarks." In: *Proceedings of CogSci 2004: 26th Annual Meeting of the Cognitive Science Society*. 2004. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.65.9925&rep=rep1&type=pdf (cit. on pp. 55, 180).

[173] IBM Global Technology Services. *The toxic terabyte: How data-dumping threatens business efficiency*. Originally online at http://www-935.ibm.com/services/no/cio/leverage/levinfo_wp_gts_thetoxic.pdf (not available anymore). Archived copy of 2013-07-22. July 2006. URL: https://web.archive.org/web/20130722124236/http://www-935.ibm.com/services/no/cio/leverage/levinfo_wp_gts_thetoxic.pdf (cit. on p. 14).

[174] J. Indulska and P. Sutton. "Location Management in Pervasive Systems." In: *Proceedings of the Australasian Information Security Workshop Conference on ACSW Frontiers 2003 - Volume 21*. ACSW Frontiers '03. Adelaide, Australia: Australian Computer Society, 2003, pp. 143–151. ISBN: 1920682007. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.84.2579&rep=rep1&type=pdf (cit. on pp. 39, 92).

[175] International Organization for Standardization, ISO/IEC JTC 1/SC 32 Data management and interchange. *Information technology — Database languages — SQL — Part 1: Framework (SQL/Framework)*. ISO/IEC 9075-1:2016. Vernier, Geneva, Switzerland: International Organization for Standardization, Dec. 2016. URL: https://www.iso.org/standard/63555.html (cit. on pp. xxxi, 95).

[176] International Organization for Standardization, ISO/TC 171/SC 2 Document file formats, EDMS systems and authenticity of information. *Document management — Portable document format — Part 2: PDF 2.0*. ISO 32000-2:2020. Vernier, Geneva, Switzerland: International Organization for Standardization, Dec. 2020. URL: https://www.iso.org/standard/75839.html (cit. on p. xxxi).

[177] A. Jain and C. Farkas. "Secure Resource Description Framework: An Access Control Model." In: *Proceedings of the Eleventh ACM Symposium on Access Control Models and Technologies (SACMAT '06), Lake Tahoe, California, USA*. ACM, 2006, pp. 121–129. DOI: 10.1145/1133058.1133076. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.84.792&rep=rep1&type=pdf (cit. on p. 74).

[178] C. Jilek. "PIMO Diary: Diary Generation from Personal Information Models." Diploma thesis. TU Kaiserslautern (in cooperation with DFKI), 2014. URL: https://www.dfki.uni-kl.de/~jilek/files/2014/Jilek14_DA_PIMO_Diary.pdf (cit. on pp. 4, 33, 180, 320).

[179] C. Jilek. *supSpaces – Semantische Support-Wissensräume für agiles Wissensmanagement im 3rd-Level-Support [supSpaces – Semantic Support (Knowledge) Spaces for Agile Knowledge Management in 3rd-level Support]*. Poster presentation at KMU-innovativ: IKT Symposium, "Mittelstand: Digital. Innovativ. Vernetzt.", Hannover, Germany, October 10-11, 2016. 2016. URL: https://www.softwaresysteme.pt-dlr.de/de/fachtagung-2016.php. Poster available online: https://www.dfki.uni-kl.de/~jilek/files/2016/supSpaces_BMBF_KMU_innovativ_IKT_2016.pdf (cit. on pp. 25, 214, 321).

[180] C. Jilek. *Inflection-tolerant Ontology-based Named Entity Recognition for Real-time Applications*. Invited talk at 13th DBpedia Community Meeting, Leipzig, Germany, May 23, 2019. 2019. URL: https://wiki.dbpedia.org/meetings/Leipzig2019. Slides available online: https://www.dfki.uni-kl.de/~jilek/files/2018/JilekSchroederNovik+18_Inflection-Tolerant_OBNER_for_Real-Time_Applications_Slides_DBpedia19.pdf (cit. on pp. 214, 321).

[181] C. Jilek. *Gezieltes Vergessen: hin zum Computer, der sich selbst aufräumt [Managed Forgetting: Towards a Personal Computer that Tidies Up Itself]*. Invited talk at Sensorik X. HR-Expertenforum, Regensburg, Germany, October 14, 2020. 2020. URL: https://www.sensorik-bayern.de/sensorik-news/artikel/agiles-sprintlernen-managed-forgetting-und-working-out-loud/. Also see https://qd.bmbfcluster.de/de/hr-expertenforum-agiles-sprintlernen-managed-forgetting-und-working-out-loud-6346.php (cit. on pp. 214, 321).

[182] C. Jilek, J. Chwalek, S. Schwarz, M. Schröder, H. Maus and A. Dengel. "Advanced Memory Buoyancy for Forgetful Information Systems." In: *AIS Transactions on Enterprise Systems* 4.1 (May 2019). DOI: 10.30844/aistes.v4i1.11. arXiv: 1811.12177. URL:

https://www.aes-journal.com/index.php/ais-tes/article/view/11/11 (cit. on pp. 4, 13, 29, 69, 80, 91, 94, 98, 125–127, 130, 131, 133, 141, 184, 185, 200, 225–227).

[183] C. Jilek, P. Gauselmann, J. Chwalek, T. Tempel and A. Dengel. "Quantifying Benefits of a Personal Knowledge Assistant on Task Resumption." In: *Arbeit HumAIne Gestalten – 67. Frühjahrskongress der Gesellschaft für Arbeitswissenschaft e.V. [HumAIne Work Design – 67th Spring Congress of the Human Factors and Ergonomics Society] (GfA 2021), Ruhr-Universität Bochum, Online, March 3-5, 2021*. GfA-Press, 2021. ISBN: 978-3-936804-29-4 (cit. on pp. 141, 147, 169, 171, 226).

[184] C. Jilek, H. Maus, S. Schwarz and A. Dengel. "Diary Generation from Personal Information Models to Support Contextual Remembering and Reminiscence." In: *Workshop on Human Memory-Inspired Multimedia Organization and Preservation (HMMP 2015), Proceedings of the 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW 2015), Turin, Italy, June 29 - July 3, 2015*. IEEE, 2015, pp. 1–6. DOI: 10.1109/ICMEW.2015.7169753. URL: https://pimo.opendfki.de/wp9-pilot/diary/JilekMausSchwarz+15_PIMO_Diary_HMMP15.pdf (cit. on pp. 4, 82, 118, 141, 181, 226).

[185] C. Jilek, Y. Runge, C. Niederée, H. Maus, T. Tempel, A. Dengel and C. Frings. "Managed Forgetting to Support Information Management and Knowledge Work." In: *KI – Künstliche Intelligenz, German Journal of Artificial Intelligence - Organ des Fachbereichs "Künstliche Intelligenz" der Gesellschaft für Informatik e.V.* 33.1 (Mar. 2019), pp. 45–55. DOI: 10.1007/s13218-018-00568-9. arXiv: 1811.12155. URL: https://arxiv.org/pdf/1811.12155.pdf (cit. on pp. 3, 4, 8, 33, 69, 70, 91, 94, 98, 141, 184, 225).

[186] C. Jilek, M. Schröder, R. Novik, S. Schwarz, H. Maus and A. Dengel. "Inflection-Tolerant Ontology-Based Named Entity Recognition for Real-Time Applications." In: *2nd Conference on Language, Data and Knowledge (LDK 2019), Leipzig, Germany, May 20-23, 2019*. Ed. by M. Eskevich, G. de Melo, C. Fäth, J. P. McCrae, P. Buitelaar, C. Chiarcos, B. Klimek and M. Dojchinovski. Vol. 70. OpenAccess Series in Informatics (OASIcs). Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2019, 11:1–11:14. DOI: 10.4230/OASIcs.LDK.2019.11. arXiv: 1812.02119. URL: https://drops.dagstuhl.de/opus/volltexte/2019/10375/pdf/OASIcs-LDK-2019-11.pdf (cit. on pp. 91, 97, 99, 101, 103–105, 110–113, 225, 320).

[187] C. Jilek, M. Schröder, S. Schwarz, H. Maus and A. Dengel. "Context Spaces as the Cornerstone of a Near-Transparent and Self-Reorganizing Semantic Desktop." In: *The Semantic Web: ESWC 2018 Satellite Events – ESWC 2018 Satellite Events, Heraklion, Crete, Greece, June 3-7, 2018, Revised Selected Papers*. Ed. by A. Gangemi, A. L. Gentile, A. G. Nuzzolese, S. Rudolph, M. Maleshkova, H. Paulheim, J. Z. Pan and M. Alam. Springer, 2018, pp. 89–94. DOI: 10.1007/978-3-319-98192-5_17. arXiv: 1805.02181. URL: https://arxiv.org/pdf/1805.02181.pdf (cit. on pp. 4, 33, 48, 51, 91, 141, 153, 188, 225, 226).

[188] C. Jilek, S. Schwarz, H. Maus and A. Dengel. "Managed Forgetting, Data Condensation & Preservation in Application." In: *Third Workshop on Ubiquitous Technologies for Augmenting the Human Mind (WAHM 2016), Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct, Heidelberg, Germany, September 12-16, 2016*. UbiComp 2016. ACM, 2016, pp. 1046–1053. DOI: 10.1145/2968219.2968567.

URL: https://www.dfki.uni-kl.de/~jilek/files/2016/JilekSchwarzMaus+16_Managed_Forgetting_WAHM16.pdf (cit. on pp. 4, 51, 211, 225–227).

[189] W. Jones. *Keeping Found Things Found: The Study and Practice of Personal Information Management*. Morgan Kaufmann, 2008. ISBN: 978-0-12-370866-3. DOI: 10.1016/B978-0-12-370866-3.X5001-2 (cit. on pp. xxxi, 15).

[190] W. Jones, A. J. Phuwanartnurak, R. Gill and H. Bruce. "Don't Take My Folders Away! Organizing Personal Information to Get Ghings Done." In: *CHI '05 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '05. Portland, OR, USA: ACM, 2005, pp. 1505–1508. DOI: 10.1145/1056808.1056952. URL: http://kftf.ischool.washington.edu/docs/folders.pdf (cit. on p. 17).

[191] S. Josefsson. *The Base16, Base32, and Base64 Data Encodings*. RFC 3548. RFC Editor, July 2003. DOI: 10.17487/RFC3548. URL: https://www.rfc-editor.org/rfc/rfc3548.html (cit. on p. 156).

[192] S. Josefsson. *The Base16, Base32, and Base64 Data Encodings*. RFC 4648. RFC Editor, Oct. 2006. DOI: 10.17487/RFC4648. URL: https://www.rfc-editor.org/rfc/rfc4648.html (cit. on p. 156).

[193] T. Atkins Jr., E. J. Etemad and F. Rivoal. *CSS Snapshot 2021*. W3C Group Note. W3C, Dec. 2021. URL: https://www.w3.org/TR/CSS/ (cit. on pp. xxx, 95).

[194] JSON-RPC Working Group. *JSON-RPC 1.0 Specification*. Online. 2005. URL: https://www.jsonrpc.org/specification_v1 (cit. on pp. 71, 75, 148).

[195] JSON-RPC Working Group. *JSON-RPC 2.0 Specification*. Online. 2010. URL: https://www.jsonrpc.org/specification (cit. on pp. 71, 75, 148).

[196] H. Al-Jumaily, P. Martínez, J. L. Martínez-Fernández and E. Van der Goot. "A real time Named Entity Recognition system for Arabic text mining." In: *Language Resources and Evaluation* 46.4 (2012), pp. 543–563. DOI: 10.1007/s10579-011-9146-z. URL: http://e-archivo.uc3m.es/bitstream/handle/10016/20235/real_LRE_2012_ps.pdf?sequence=1 (cit. on p. 100).

[197] B. Jung, I. Yoon, H. Lim, F. A. Ramirez-Weber and D. Petkovic. "Annotizer: User-friendly WWW annotation system for collaboration in research and education environments." In: *Proceedings of the IASTED International Conference on Web Technologies, Applications and Services, WTAS*. 2006. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.86.9276&rep=rep1&type=pdf (cit. on p. 157).

[198] N. Kanhabua, C. Niederée and W. Siberski. "Towards Concise Preservation by Managed Forgetting: Research Issues and Case Study." In: *Proceedings of the 10th International Conference on Preservation of Digital Objects (iPres)*. 2013. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.723.1478&rep=rep1&type=pdf (cit. on pp. 33, 35).

[199] D. R. Karger, K. Bakshi, D. Huynh, D. Quan and V. Sinha. "Haystack: A customizable general-purpose information management tool for end users of semistructured data." In: *Proceedings of the Second Biennial Conference on Innovative Data Systems Research (CIDR 2005), January 4-7, 2005, Asilomar, CA*. 2005. URL: https://www.cidrdb.org/cidr2005/papers/P02.pdf (cit. on p. 57).

[200] KDE Community. *Baloo*. KDE Community Wiki. KDE e.V., 2014. URL: https://community.kde.org/Baloo (cit. on p. 63).

[201] M. Kiesel and S. Schwarz. "Skipforward: a lightweight ontology-based peer-to-peer recommendation system." In: *Proceedings of the Poster and Demonstration Session at the 7th International Semantic Web Conference (ISWC 2008), Karlsruhe, Germany, October 28, 2008*. Ed. by C. Bizer and A. Joshi. Vol. 401. CEUR Workshop Proceedings. CEUR-WS.org, 2008. URL: http://ceur-ws.org/Vol-401/iswc2008pd_submission_43.pdf (cit. on pp. 53, 72, 74).

[202] M. Kiesel, S. Schwarz, L. van Elst and G. Buscher. "Mymory: Enhancing a Semantic Wiki with Context Annotations." In: *The Semantic Web: Research and Applications – 5th European Semantic Web Conference, ESWC 2008, Tenerife, Canary Islands, Spain*. Springer Berlin Heidelberg, 2008, pp. 817–821. DOI: 10.1007/978-3-540-68234-9_65. URL: https://link.springer.com/content/pdf/10.1007/978-3-540-68234-9_65.pdf (cit. on p. 52).

[203] M. Kiesel, S. Schwarz, L. van Elst and G. Buscher. "Using Attention and Context Information for Annotations in a Semantic Wiki." In: *Proceedings of the 3rd Semantic Wiki Workshop (SemWiki 2008) at the 5th European Semantic Web Conference (ESWC 2008), Tenerife, Spain, June 2nd, 2008*. Ed. by C. Lange, S. Schaffert, H. Skaf-Molli and M. Völkel. Vol. 360. CEUR Workshop Proceedings. CEUR-WS.org, 2008. URL: http://ceur-ws.org/Vol-360/paper-10.pdf (cit. on pp. 52, 53).

[204] J. Kim and W. B. Croft. "Building pseudo-desktop collections." In: *Proceedings of the SIGIR 2009 Workshop on the Future of IR Evaluation*. 2009, pp. 39–40. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.160.1252&rep=rep1&type=pdf#page=49 (cit. on p. 142).

[205] J. Kim and W. B. Croft. "Retrieval Experiments Using Pseudo-desktop Collections." In: *Proceedings of the 18th ACM Conference on Information and Knowledge Management*. CIKM '09. ACM, 2009, pp. 1297–1306. DOI: 10.1145/1645953.1646117. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.367.145&rep=rep1&type=pdf (cit. on p. 142).

[206] B. Klimt and Y. Yang. "Introducing the Enron Corpus." In: *CEAS 2004 - First Conference on Email and Anti-Spam, July 30-31, 2004, Mountain View, California, USA*. 2004. URL: https://ceas.cc/papers-2004/168.pdf. (*Enron Email Dataset* [Sw33] available online.) (Cit. on pp. 55, 266).

[207] J. Knabe. "Reflecting Intelligent Systems: The Effects of Transparency on Trust, Intention to Use and Subjective Perceptions." Master's thesis. Goethe University Frankfurt am Main (in cooperation with DFKI), 2021 (cit. on pp. 205–207, 227).

[208] KnowMore Project Team. *The KnowMore Project (Knowledge Management for Learning Organizations) or Context-Aware, Proactive Delivery of Task-Specific Knowledge*. Project website. 1997. URL: https://www.dfki.uni-kl.de/frodo/knowmore.html (cit. on p. 52).

[209] S. Koldijk. “Automatic Recognition of Context and Stress to Support Knowledge Workers.” In: *Proceedings of the 30th European Conference on Cognitive Ergonomics (ECCE ’12), Edinburgh, United Kingdom*. ACM, 2012, pp. D20–D23. DOI: 10.1145/2448136.2448185. URL: http://www.cs.ru.nl/~skoldijk/Papers/ECCE_%20DoctoralConsortium_Koldijk_cr.pdf (cit. on pp. 56, 116).

[210] S. Koldijk. “Capturing Context and Mental State of Knowledge Workers.” In: *Proceedings of the 4th Information Interaction in Context Symposium (IIIX ’12), Nijmegen, The Netherlands*. ACM, 2012, p. 322. DOI: 10.1145/2362724.2362786. URL: http://cs.ru.nl/~skoldijk/Papers/IIIX_DoctoralConsortium_SaskiaKoldijk_fullPaper_cr.pdf (cit. on pp. 56, 116).

[211] S. Koldijk, G. Koot, M. Neerincx and W. Kraaij. “Privacy and User Trust in Context-Aware Systems.” In: *User Modeling, Adaptation, and Personalization*. Ed. by V. Dimitrova, T. Kuflik, D. Chin, F. Ricci, P. Dolog and G.-J. Houben. Springer International Publishing, 2014, pp. 134–145. DOI: 10.1007/978-3-319-08786-3_12. URL: http://cs.ru.nl/~skoldijk/Papers/Paper%20for%20UMAP%202014_camera_ready_final_2.pdf (cit. on pp. 56, 96, 143).

[212] S. Koldijk, M. Sappelli, S. Verberne, M. A. Neerincx and W. Kraaij. “The SWELL Knowledge Work Dataset for Stress and User Modeling Research.” In: *Proceedings of the 16th International Conference on Multimodal Interaction (ICMI ’14), Istanbul, Turkey*. ACM, 2014, pp. 291–298. DOI: 10.1145/2663204.2663257. URL: http://www.cs.ru.nl/~skoldijk/Papers/ICMI%202014%20paper_final_cr.pdf (cit. on p. 56).

[213] S. Koldijk, M. Sappelli, S. Verberne, M. A. Neerincx and W. Kraaij. “The SWELL Knowledge Work Dataset for Stress and User Modeling Research.” In: *Proceedings of the 16th International Conference on Multimodal Interaction (ICMI ’14), Istanbul, Turkey*. ACM, 2014, pp. 291–298. DOI: 10.1145/2663204.2663257. URL: http://www.cs.ru.nl/~skoldijk/Papers/ICMI%202014%20paper_final_cr.pdf. (*SWELL Knowledge Work Dataset* [Sw67] available online.) (Cit. on pp. 142, 268).

[214] S. Koldijk, M. van Staalduinen, M. Neerincx and W. Kraaij. “Real-Time Task Recognition Based on Knowledge Workers’ Computer Activities.” In: *Proceedings of the 30th European Conference on Cognitive Ergonomics (ECCE ’12), Edinburgh, United Kingdom*. ACM, 2012, pp. 152–159. DOI: 10.1145/2448136.2448170. URL: https://repository.ubn.ru.nl/bitstream/handle/2066/130373/postprint_2066_130373-20141015140235.pdf?sequence=1 (cit. on pp. 56, 116).

[215] W. Kraaij, S. Verberne, S. Koldijk, E. de Korte, S. van Dantzig, M. Sappelli, M. Shoaib, S. Bosems, R. Achterkamp, A. Bonomi, J. Schavemaker, B. Hulsebosch, T. Wabeke, M. Vollenbroek-Hutten, M. Neerincx and M. van Sinderen. “Personalized support for well-being at work: an overview of the SWELL project.” In: vol. 30. 3. Springer, 2020, pp. 413–446. DOI: 10.1007/s11257-019-09238-3. URL: https://link.springer.com/content/pdf/10.1007/s11257-019-09238-3.pdf (cit. on p. 56).

[216] M. Kraus. “PimoDAV: A multi-view calendar-endpoint for PIMO.” Bachelor’s thesis. TU Kaiserslautern (in cooperation with DFKI), 2016 (cit. on pp. 29, 30).

[217] L. Kunze, C. Beierle and I. J. Timm, eds. *KI – Künstliche Intelligenz, German Journal of Artificial Intelligence - Organ des Fachbereichs "Künstliche Intelligenz" der Gesellschaft für Informatik e.V.* 33.1 (Mar. 2019): *Intentional Forgetting*. URL: https://link.springer.com/journal/13218/volumes-and-issues/33-1 (cit. on p. 58).

[218] S. Łazaruk, M. Kaczmarek, J. Dzikowski, O. Tokarchuk and W. Abramowicz. "Towards the Semantic Web – Incentivizing Semantic Annotation Creation Process." In: *International Conference on Knowledge Engineering and Knowledge Management*. Springer. 2012, pp. 282–291. DOI: 10.1007/978-3-642-33876-2_25 (cit. on pp. 39, 45).

[219] G. Leban and M. Grobelnik. "Displaying Email-Related Contextual Information Using Contextify." In: *Proceedings of the ISWC 2010 Posters & Demonstrations Track: Collected Abstracts, Shanghai, China, November 9, 2010*. Ed. by A. Polleres and H. Chen. Vol. 658. CEUR Workshop Proceedings. CEUR-WS.org, 2010, pp. 181–184. URL: http://ceur-ws.org/Vol-658/paper533.pdf (cit. on p. 56).

[220] F. R. León, R. Lundgren, A. Huapaya, I. Sinai and V. Jennings. "Challenging the Courtesy Bias Interpretation of Favorable Clients' Perceptions of Family Planning Delivery." In: *Evaluation Review* 31.1 (2007), pp. 24–42. DOI: 10.1177/0193841X06289044. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.845.3731&rep=rep1&type=pdf (cit. on p. 176).

[221] V. I. Levenshtein. "Binary codes capable of correcting deletions, insertions, and reversals." In: *Soviet Physics Doklady*. Vol. 10. 8. (Translated from Doklady Akademii Nauk SSSR, Vol. 163, No. 4, pp. 845-848, August, 1965). 1966, pp. 707–710. URL: https://nymity.ch/sybilhunting/pdf/Levenshtein1966a.pdf (cit. on pp. 109, 110).

[222] B. J. Levy and M. C. Anderson. "Inhibitory processes and the control of memory retrieval." In: *Trends in cognitive sciences* 6.7 (2002), pp. 299–305. DOI: 10.1016/S1364-6613(02)01923-X. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.668.3692&rep=rep1&type=pdf (cit. on pp. 35, 218).

[223] J. R. Lewis. "IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use." In: *International Journal of Human-Computer Interaction* 7.1 (1995), pp. 57–78. DOI: 10.1080/10447319509526110. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.584.6610&rep=rep1&type=pdf (cit. on pp. 85, 87).

[224] W. Lezius. *Die Wortklassensysteme von Morphy (Vollständiges Klassensystem, großes und kleines Tag Set) [The word class systems of Morphy (Complete class system, big and small tag set)]*. Tech. rep. FB 2 – Kognitive Psychologie, Universität Paderborn, Aug. 1998. URL: https://morphy.wolfganglezius.de/content/2-download/wklassen.pdf (cit. on p. 107).

[225] S. Li, L. D. Xu and S. Zhao. "The internet of things: a survey." In: *Information Systems Frontiers*. Vol. 17. 2. Springer, 2015, pp. 243–259. DOI: 10.1007/s10796-014-9492-7. URL: https://www.academia.edu/download/47649087/2015_Li_-_The_internet_of_things-a_survey.pdf (cit. on p. 53).

[226] R. Likert. "A technique for the measurement of attitudes." In: *Archives of Psychology* 22.140 (1932), pp. 5–55 (cit. on pp. 85, 147, 244).

[227] S. Lindstaedt and H. Mayer. “A Storyboard of the APOSDLE Vision.” In: *Innovative Approaches for Learning and Knowledge Sharing*. Ed. by W. Nejdl and K. Tochtermann. Springer Berlin Heidelberg, 2006, pp. 628–633. DOI: 10.1007/11876663_64. URL: https://www.know-center.at/filemaker/pdf_publikationen/A%20Storyboard%20of%20the%20APOSDLE%20Vision.pdf (cit. on p. 56).

[228] T.-Y. Liu. “Learning to Rank for Information Retrieval.” In: *Foundations and Trends in Information Retrieval* 3.3 (2009), pp. 225–331. DOI: 10.1561/1500000016 (cit. on p. 3).

[229] R. H. Logie. “Preserving and Forgetting in Human and Digital Memory.” Keynote speech at Human Memory-Inspired Multimedia Organization and Preservation Workshop at 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW 2015), Turin, Italy, June 29 - July 3, 2015. 2015. URL: https://www.forgetit-project.eu/en/get-involved/hmmp15/ (cit. on p. 3).

[230] R. H. Logie, M. Wolters and E. Niven. “Preserving and Forgetting in the Human Brain.” In: *Personal Multimedia Preservation: Remembering or Forgetting Images and Video*. Ed. by V. Mezaris, C. Niederée and R. H. Logie. Springer, 2018, pp. 9–45. DOI: 10.1007/978-3-319-73465-1_2 (cit. on pp. 57, 126).

[231] I Lu J.and Holubová. “Multi-Model Databases: A New Journey to Handle the Variety of Data.” In: *ACM Computing Surveys* 52.3 (June 2019). DOI: 10.1145/3323214. URL: https://www.cs.helsinki.fi/u/jilu/documents/Multi_model_Databases__A__New_Journey_to_Handle_the_Variety_of_DataFinal.pdf (cit. on pp. 74, 211).

[232] A. M. Lund. “Measuring usability with the USE questionnaire.” In: *Usability interface* 8(2) (2001), pp. 3–6. URL: https://www.researchgate.net/profile/Arnold_Lund/publication/230786746_Measuring_Usability_with_the_USE_Questionnaire/links/56e5a90e08ae98445c21561c/Measuring-Usability-with-the-USE-Questionnaire.pdf (cit. on pp. 176, 177, 203).

[233] J. Luttenberger. “Computern das Vergessen beibringen [Teaching Computers to Forget].” In: *Die Rheinpfalz* (July 27, 2023). URL: https://www.rheinpfalz.de/lokal/kaiserslautern_artikel,-das-dfki-erforscht-wie-man-computern-das-vergessen-beibringen-kann-_arid,5535812.html (cit. on pp. 214, 321).

[234] Managed Forgetting Project Team. *Managed Forgetting*. Project website. 2016. URL: http://www.spp1921.de/projekte/p4.html.en (cit. on pp. 33, 35, 54, 58, 162, 213, 319).

[235] H. B. Mann and D. R. Whitney. “On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other.” In: *The Annals of Mathematical Statistics* 18.1 (Mar. 1947), pp. 50–60. DOI: 10.1214/aoms/1177730491. URL: https://projecteuclid.org/journalArticle/Download?urlId=10.1214%2Faoms%2F1177730491 (cit. on p. 171).

[236] C. D. Manning, P. Raghavan and H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008. DOI: 10.1017/CBO9780511809071 (cit. on p. 35).

[237] H. Maus. “Workflow Context as a Means for Intelligent Information Support.” In: *Modeling and Using Context: 3rd International and Interdisciplinary Conference (CONTEXT 2001)*. Vol. 2116. LNAI. Springer, 2001, pp. 261–264. DOI: 10.1007/3-540-44607-9_20. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.23.9363&rep=rep1&type=pdf (cit. on pp. 52, 69, 70).

[238] H. Maus. "Workflow-Kontext zur Realisierung prozessorientierter Assistenz in Organisational Memories [Workflow Context to Realize Process-oriented Assistance in Organizational Memories]." PhD thesis. TU Kaiserslautern, 2007. ISBN: 978-3-86624-267-8 (cit. on pp. 53, 60, 69).

[239] H. Maus and C. Jilek. *ForgetIT: Concise Preservation by Combining Managed Forgetting and Contextualized Remembering*. Demo presentation at CeBIT 2015, Hannover, Germany, March 16-20, 2015. 2015. Demo slides available online: https://www.dfki.uni-kl.de/~jilek/files/2015/ForgetIT_CeBIT_2015.pdf (cit. on pp. 181, 214, 321).

[240] H. Maus, C. Jilek and S. Schwarz. "Remembering and Forgetting for Personal Preservation." In: *Personal Multimedia Preservation: Remembering or Forgetting Images and Video*. Ed. by V. Mezaris, C. Niederée and R. H. Logie. Springer, 2018, pp. 233–277. DOI: 10.1007/978-3-319-73465-1_7 (cit. on pp. 4, 13, 23, 25, 27, 28, 31, 91, 125, 138, 141, 184, 185, 195, 225–227).

[241] H. Maus, S. Schwarz and A. Dengel. "Weaving Personal Knowledge Spaces into Office Applications." In: *Integration of Practice-Oriented Knowledge Technology: Trends and Prospectives*. Ed. by M. Fathi. Springer, 2013, pp. 71–82. DOI: 10.1007/978-3-642-34471-8_6. URL: https://pimo.opendfki.de/wp9-pilot/intro/MausSchwarzDengel12_submitted.pdf (cit. on pp. xxx, 21, 22, 25, 54, 63, 64, 71).

[242] H. Maus, S. Schwarz, B. Eldesouky, C. Jilek, M. Wolters and B. Loğoğlu. *ForgetIT Deliverable D9.3: Use Cases & Mock-up Development*. Deliverable. ForgetIT Consortium, Feb. 2015. URL: https://www.forgetit-project.eu/fileadmin/fm-dam/deliverables/ForgetIT_WP9_D9.3.pdf (cit. on pp. 4, 23, 28, 181, 226).

[243] H. Maus, S. Schwarz, J. Haas and A. Dengel. "ConTask: Context-Sensitive Task Assistance in the Semantic Desktop." In: *Enterprise Information Systems*. Springer Berlin Heidelberg, 2011, pp. 177–192. DOI: 10.1007/978-3-642-19802-1_13. URL: https://www.dfki.uni-kl.de/~maus/dok/MausSchwarzHaas+11.pdf (cit. on pp. 38, 54, 60, 65).

[244] H. Maus, S. Schwarz, C. Jilek and B. Eldesouky. *ForgetIT Personal Preservation Pilot*. Web documentation. 2015. URL: https://pimo.opendfki.de/wp9-pilot/ (cit. on pp. 4, 25, 27, 91, 138, 184, 185, 225–227).

[245] H. Maus, S. Schwarz, C. Jilek and F. Gallo. *ForgetIT Deliverable D9.4: Personal Preservation Pilot II: Concise Preserving Mobile Information Assistant*. Deliverable. ForgetIT Consortium, Dec. 2015. URL: https://www.dfki.de/fileadmin/user_upload/import/8360_D9.4.pdf (cit. on pp. 4, 31, 125, 135, 138, 139, 195, 226).

[246] H. Maus, S. Schwarz, C. Jilek, M. Wolters, S. Rhodes, A. Ceroni and G. Gür. *ForgetIT Deliverable D9.5: Personal Preservation Report*. Deliverable. ForgetIT Consortium, Jan. 2016. URL: https://www.forgetit-project.eu/fileadmin/fm-dam/deliverables/ForgetIT_WP9_D9.5New.pdf (cit. on pp. 4, 125, 140, 226).

[247] V. Mayer-Schönberger. *Delete: The Virtue of Forgetting in the Digital Age*. Princeton University Press, 2009. ISBN: 978-0-691-13861-9 (cit. on pp. 18, 57).

[248] A. Melnikov and B. Leiba. *Internet Message Access Protocol (IMAP) - Version 4rev2*. RFC 9051. RFC Editor, Aug. 2021. DOI: 10.17487/RFC9051. URL: https://www.rfc-editor.org/rfc/rfc9051.html (cit. on pp. xxx, 81, 153, 261).

[249] P. N. Mendes, M. Jakob, A. García-Silva and C. Bizer. "DBpedia spotlight: shedding light on the web of documents." In: *Proceedings of the 7th International Conference on Semantic Systems (I-Semantics)*. ACM. 2011, pp. 1–8. DOI: 10.1145/2063518.2063519. URL: https://www.dbpedia-spotlight.org/docs/spotlight.pdf. (Implementation available as *DBpedia Spotlight* [Sw74].) (Cit. on p. 269).

[250] M. Mesnier, E. Thereska, G. R. Ganger, D. Ellard and M. Seltzer. "File classification in self-* storage systems." In: *Proceedings of the International Conference on Autonomic Computing, New York, NY, USA, May 17-18, 2004*. IEEE, 2004, pp. 44–51. DOI: 10.1109/ICAC.2004.1301346. URL: https://www.pdl.cmu.edu/PDL-FTP/ABLE/CMU-PDL-04-101.pdf (cit. on p. 63).

[251] V. Mezaris, C. Niederée and R. H. Logie, eds. *Personal Multimedia Preservation: Remembering or Forgetting Images and Video*. Springer Series on Cultural Computing. Springer Cham, 2018. ISBN: 978-3-319-73464-4. DOI: 10.1007/978-3-319-73465-1 (cit. on pp. 57, 193, 194).

[252] Microsoft Corporation. *Description of how Word creates temporary files*. Microsoft Support (online). Microsoft Corporation, Aug. 2020. URL: https://support.microsoft.com/en-us/topic/description-of-how-word-creates-temporary-files-66b112fb-d2c0-8f40-a0be-70a367cc4c85 (cit. on pp. 73, 154).

[253] Microsoft Corporation. *Windows Accessibility API overview*. Microsoft Docs. Microsoft Corporation, 2020. URL: https://docs.microsoft.com/en-us/windows/win32/winauto/windows-automation-api-portal (cit. on p. 93).

[254] Microsoft Corporation. *Server Message Block (SMB) Protocol Versions 2 and 3*. Microsoft Docs. Windows Protocols. Open Specifications (Technical report) MS-SMB2 v20210625. June 2021. URL: https://docs.microsoft.com/openspecs/windows_protocols/ms-smb2 (cit. on pp. xxxi, 25, 153, 196).

[255] R. B. Miller. "Response Time in Man-Computer Conversational Transactions." In: *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I, AFIPS '68 (Fall, part I), San Francisco, California*. ACM, 1968, pp. 267–277. DOI: 10.1145/1476589.1476628. URL: https://dl.acm.org/doi/pdf/10.1145/1476589.1476628 (cit. on p. 97).

[256] T. Milo. "Getting Rid of Data." In: *Journal of Data and Information Quality* 12.1 (Nov. 2019). DOI: 10.1145/3326920. Related keynote slides available online: https://vldb2019.github.io/files/VLDB19-keynote-2-slides.pdf (cit. on p. 125).

[257] E. Minack, L. Sauermann, G. Grimnes, C. Fluit and J. Broekstra. *The Sesame LuceneSail: RDF queries with full-text search*. Tech. rep. NEPOMUK Consortium, Feb. 2008. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.125.9864&rep=rep1&type=pdf (cit. on pp. 53, 77).

[258] S. Mori, H. Nishida and H. Yamada. *Optical Character Recognition*. Wiley, 1999. ISBN: 047308196 (cit. on p. 93).

[259] D. Morris, M. Ringel Morris and G. Venolia. “SearchBar: A Search-Centric Web History for Task Resumption and Information Re-Finding.” In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI’08), Florence, Italy*. ACM, 2008, pp. 1207–1216. DOI: 10.1145/1357054.1357242. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.131.7400&rep=rep1&type=pdf (cit. on p. 169).

[260] Mozilla Project. *XUL*. Mozilla Wiki. 2004. URL: https://wiki.mozilla.org/XUL:Home_Page (cit. on pp. 91, 92).

[261] Mozilla Project. *WebExtensions*. Mozilla Wiki. 2015. URL: https://wiki.mozilla.org/WebExtensions (cit. on pp. 92, 94, 153, 157, 196).

[262] Mozilla Project. *Firefox Data Stores Documentation: places.sqlite*. Online. 2017. URL: https://github.com/mozilla/firefox-data-store-docs#placessqlite (cit. on p. 81).

[263] S. Muggleton and L. De Raedt. “Inductive Logic Programming: Theory and methods.” In: *The Journal of Logic Programming* 19–20 (1994). Special Issue: Ten Years of Logic Programming, pp. 629–679. DOI: 10.1016/0743-1066(94)90035-3. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.27.4632&rep=rep1&type=pdf (cit. on p. 58).

[264] Mymory Project Team. *Mymory – Situated Documents in Personal Information Spaces*. Project website. Originally online at http://www.dfki.uni-kl.de/mymory/ (not available anymore). Archived copy of 2016-11-14. 2006. URL: https://web.archive.org/web/20161114142751/http://www.dfki.uni-kl.de/mymory. Also see https://www.dfki.de/web/forschung/projekte-publikationen/projekte-uebersicht/projekt/mymory (cit. on p. 52).

[265] K. Needham. *The Future of Developing Firefox Add-ons*. Mozilla Add-ons Community Blog. Aug. 2015. URL: https://blog.mozilla.org/addons/2015/08/21/the-future-of-developing-firefox-add-ons/ (cit. on p. 91).

[266] B. J. Nelson. “Remote Procedure Call.” PhD thesis. Carnegie Mellon University, 1981 (cit. on pp. xxxi, 25).

[267] NEPOMUK Consortium. *NEPOMUK – The Social Semantic Desktop – FP6-027705*. Project website. 2006. URL: https://nepomuk.semanticdesktop.org/. Also see https://cordis.europa.eu/project/id/027705 (cit. on pp. 18, 23, 24, 49, 50, 52, 53, 64, 233).

[268] Nepomuk-KDE Development Team. *Nepomuk: The Semantic Desktop in KDE*. Project website, originally online at https://nepomuk.kde.org/ (not available anymore). Archived copy of 2013-04-01. 2012. URL: https://web.archive.org/web/20130401085641/https://nepomuk.kde.org/ (cit. on pp. 5, 22, 53, 63).

[269] G. Nguyen, Š. Dlugolinský, M. Laclavík, M. Šeleng and V. Tran. “Next Improvement Towards Linear Named Entity Recognition Using Character Gazetteers.” In: *Advanced Computational Methods for Knowledge Engineering*. Springer, 2014, pp. 255–265. DOI: 10.1007/978-3-319-06569-4_19. URL: http://ikt.ui.sav.sk/archive/vega15/AEC11_Springer_Giang_Gazetteer.pdf. (Implementation available as *IKT Gazetteers* [Sw93].) (Cit. on pp. 100, 110–113, 270).

[270] C. Niederée, N. Kanhabua, F. Gallo and R. H. Logie. "Forgetful Digital Memory: Towards Brain-Inspired Long-Term Data and Information Management." In: *SIGMOD Record* 44.2 (2015), pp. 41–46. DOI: 10.1145/2814710.2814718. URL: https://sigmodrecord.org/publications/sigmodRecord/1506/09_vision_Niederee.pdf (cit. on pp. 33, 35).

[271] J. Nielsen. *Usability Engineering*. Morgan Kaufmann, 1993. ISBN: 0125184050 (cit. on p. 97).

[272] K. North, A. Brandner and T. Steininger. "Die Wissenstreppe: Information – Wissen – Kompetenz [The knowledge ladder: information – knowledge – expertise]." In: *Wissensmanagement für Qualitätsmanager: Erfüllung der Anforderungen nach ISO 9001:2015 [Knowledge Management for Quality Managers: Meeting the Requirements of ISO 9001:2015]*. Springer Fachmedien Wiesbaden, 2016, pp. 5–8. ISBN: 978-3-658-11250-9. DOI: 10.1007/978-3-658-11250-9_2 (cit. on pp. 13, 14).

[273] P. Norvig. *English Letter Frequency Counts: Mayzner Revisited or ETAOIN SRHLDCU*. Online. 2013. URL: http://norvig.com/mayzner.html (cit. on p. 100).

[274] R. Novik. "Searching Forgetful Information Systems." Bachelor's thesis. TU Kaiserslautern (in cooperation with DFKI), 2019 (cit. on pp. 81, 95, 201–204, 212, 227, 243).

[275] T. O'Connor, T. Leithead, I. Hickson, S. Faulkner, S. Pfeiffer, E. D. Navara and R. Berjon. *HTML5*. W3C Recommendation. W3C, Oct. 2014. URL: https://www.w3.org/TR/2014/REC-html5-20141028/ (cit. on pp. xxx, 71, 95, 148, 178).

[276] Oracle Corporation. *Date (Class)*. Java Platform, Standard Edition 8 API Specification (online documentation). 2022. URL: https://docs.oracle.com/javase/8/docs/api/java/util/Date.html (cit. on p. 129).

[277] R. Parmar, A. Leiponen and L. D. W. Thomas. "Building an organizational digital twin." In: *Business Horizons* 63.6 (2020), pp. 725–736. DOI: 10.1016/j.bushor.2020.08.001. URL: https://www.researchgate.net/profile/Llewellyn-Thomas/publication/336408969_Building_an_organizational_digital_twin/links/5d9f811592851c6b4bcb6385/Building-an-organizational-digital-twin.pdf (cit. on p. xxx).

[278] G. Probst, S. Raub and K. Romhardt. *Wissen managen: Wie Unternehmen ihre wertvollste Ressource optimal nutzen [Managing knowledge: How enterprises optimally utilize their most valuable resource]*. Gabler, 1997. DOI: 10.1007/978-3-8349-8597-2 (cit. on p. 45).

[279] E. Prud'hommeaux and A. Seaborne. *SPARQL Query Language for RDF*. W3C Recommendation. W3C, Jan. 2008. URL: https://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/ (cit. on pp. xxxi, 19, 75).

[280] L. De Raedt, ed. *Logical and Relational Learning*. Cognitive Technologies. Springer Berlin, Heidelberg, 2008. DOI: 10.1007/978-3-540-68856-3 (cit. on p. 58).

[281] A. S. Rath. "User Interaction Context – Studying and Enhancing Automatic User Task Detection on the Computer Desktop via an Ontology-based User Interaction Context Model." PhD thesis. Graz University of Technology, 2010 (cit. on pp. 56, 91, 116).

[282] A. S. Rath, D. Devaurs and S. N. Lindstaedt. "UICO: An Ontology-Based User Interaction Context Model for Automatic Task Detection on the Computer Desktop." In: *Proceedings of the 1st Workshop on Context, Information and Ontologies (CIAO '09), Heraklion, Greece*. ACM, 2009. DOI: 10.1145/1552262.1552270. URL: https://www.know-center.at/filemaker/pdf_publikationen/a8-rath.pdf (cit. on pp. 56, 91, 116).

[283] A. S. Rath, D. Devaurs and S. N. Lindstaedt. "Studying the Factors Influencing Automatic User Task Detection on the Computer Desktop." In: *Sustaining TEL: From Innovation to Learning and Practice, 5th European Conference on Technology Enhanced Learning, EC-TEL 2010, Barcelona, Spain, September 28 - October 1, 2010, Proceedings*. Ed. by M. Wolpers, P. A. Kirschner, M. Scheffel, S. Lindstaedt and V. Dimitrova. Springer Berlin Heidelberg, 2010, pp. 292–307. DOI: 10.1007/978-3-642-16020-2_20. URL: https://hal.archives-ouvertes.fr/file/index/docid/872201/filename/rath-10-ec-tel.pdf (cit. on pp. 56, 91, 116).

[284] A. S. Rath, N. Weber, M. Kröll, M. Granitzer, O. Dietzel and S. N. Lindstaedt. "Context-Aware Knowledge Services." In: *Workshop on Personal Information Management (PIM2008) at the 26th Computer Human Interaction Conference (CHI2008), Florence Italy, April 5-10, 2008*. 2008. URL: https://www.know-center.at/filemaker/pdf_publikationen/Context-aware_knowledge_services.pdf (cit. on pp. 56, 116).

[285] D. Reinsel, J. Gantz and J. Rydning. *The Digitization of the World – From Edge to Core*. Tech. rep. US44413318. IDC White Paper. International Data Corporation (IDC), Nov. 2018. URL: https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf (cit. on p. 15).

[286] R. Al-Rfou and S. Skiena. "SpeedRead: A Fast Named Entity Recognition Pipeline." In: *Proceedings 24th International Conference on Computational Linguistics (COLING 2012)* (2012), pp. 51–66. arXiv: 1301.2857. URL: https://aclanthology.org/C12-1004.pdf (cit. on p. 100).

[287] J. Richter, M. Völkel and H. Haller. "DeepaMehta – A Semantic Desktop." In: *Proceedings of the ISWC 2005 Workshop on The Semantic Desktop - Next Generation Information Management & Collaboration Infrastructure. Galway, Ireland, November 6, 2005*. Ed. by S. Decker, J. Park, D. Quan and L. Sauermann. Vol. 175. CEUR Workshop Proceedings. CEUR-WS.org, 2005. URL: http://ceur-ws.org/Vol-175/30_dm_poster.pdf (cit. on p. 53).

[288] M. Ringel, E. Cutrell, S. Dumais and E. Horvitz. "Milestones in time: The value of landmarks in retrieving information from personal stores." In: *Proceedings of Interact*. Vol. 2003. 2003, pp. 184–191. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.436.8285&rep=rep1&type=pdf (cit. on pp. 55, 180).

[289] E. F. Risko and S. J. Gilbert. "Cognitive Offloading." In: *Trends in Cognitive Sciences* 20.9 (2016), pp. 676–688. DOI: 10.1016/j.tics.2016.07.002. URL: http://samgilbert.net/pubs/Risko2016TiCS.pdf (cit. on p. 161).

[290] U. Riss, H. Maus, S. Javaid and C. Jilek. "Digital Twins of an Organization for Enterprise Modeling." In: *The Practice of Enterprise Modeling – 13th IFIP WG 8.1 Working Conference on the Practice of Enterprise Modelling (PoEM 2020), Riga, Latvia, November 25-27, 2020, Proceedings*. Ed. by J. Grabis and D. Bork. Lecture Notes in Business Information Processing (LNBIP). Springer, 2020, pp. 25–40. DOI: 10.1007/978-3-030-63479-7_3. URL: https://www.dfki.uni-kl.de/~maus/dok/RissMausJavaid+2020.pdf (cit. on pp. 211, 217, 227).

[291] T. Rizzo. *WinFS 101: Introducing the New Windows File System*. Microsoft Docs. Microsoft Corporation, Mar. 2004. URL: http://msdn.microsoft.com/de-de/library/aa480687.aspx (cit. on p. 55).

[292] E. Roche and Y. Schabes. *Finite-State Language Processing*. MIT Press, 1997. ISBN: 0-262-18182-7 (cit. on pp. xxx, 101, 102).

[293] Y. Runge, C. Frings and T. Tempel. "Hiding to hedge against information overload: Diverging distraction effects due to task-irrelevant information from misleading associations." In: *AIS Transactions on Enterprise Systems* 4.1 (May 2019). DOI: 10.30844/aistes.v4i1.12. URL: https://www.aes-journal.com/index.php/ais-tes/article/view/12/12 (cit. on p. 147).

[294] Y. Runge, C. Frings and T. Tempel. "Specifying the mechanisms behind benefits of saving-enhanced memory." In: *Psychological Research* 85.4 (June 2021), pp. 1633–1644. DOI: 10.1007/s00426-020-01341-0 (cit. on p. 161).

[295] G. Salton, A. Wong and C.-S. Yang. "A vector space model for automatic indexing." In: *Communications of the ACM* 18.11 (1975), pp. 613–620. DOI: 10.1145/361219.361220. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.107.7453&rep=rep1&type=pdf (cit. on p. 118).

[296] R. Sanderson, P. Ciccarese and B. Young. *Web Annotation Data Model*. W3C Recommendation. W3C, Feb. 2017. URL: https://www.w3.org/TR/annotation-model/ (cit. on p. 157).

[297] M. Sappelli. "Knowledge Work in Context – User Centered Knowledge Worker Support." PhD thesis. Radboud University, Nijmegen, 2016. URL: https://repository.ubn.ru.nl/bitstream/handle/2066/151710/151710.pdf?sequence=1&isAllowed=y (cit. on pp. 56, 116).

[298] M. Sappelli, G. Pasi, S. Verberne, M. de Boer and W. Kraaij. "Assessing e-mail intent and tasks in e-mail messages." In: *Information Sciences* 358–359 (2016), pp. 1–17. DOI: 10.1016/j.ins.2016.03.002 (cit. on pp. 56, 116).

[299] M. Sappelli, S. Verberne, S. Koldijk and W. Kraaij. "Collecting a Dataset of Information Behaviour in Context." In: *Proceedings of the 4th Workshop on Context-Awareness in Retrieval and Recommendation (CARR '14), Amsterdam, The Netherlands*. ACM, 2014, pp. 26–29. DOI: 10.1145/2601301.2601306. URL: https://repository.ubn.ru.nl/bitstream/handle/2066/127981/127981.pdf (cit. on pp. 56, 91, 142).

[300] M. Sappelli, S. Verberne and W. Kraaij. “Combining textual and non-textual features for e-mail importance estimation.” In: *Proceedings of the 25th Benelux Conference on Articial Intelligence (BNAIC 2013), Delft, Nov 7-8, 2013*. Ed. by K. Hindriks. Nov. 2013, pp. 168–174. URL: https://repository.ubn.ru.nl/bitstream/handle/2066/122950/122950.pdf (cit. on pp. 56, 125).

[301] M. Sappelli, S. Verberne and W. Kraaij. “Adapting the Interactive Activation Model for Context Recognition and Identification.” In: *ACM Transactions on Interactive Intelligent Systems* 6.3 (Sept. 2016). DOI: 10.1145/2873067. URL: https://msappelli.github.io/assets/pdf/2016_TiiSmanuscript-Adapting%20the%20interactive%20activation%20model%20for%20context%20recognition%20and%20identification.pdf (cit. on pp. 56, 116).

[302] M. Sappelli, S. Verberne and W. Kraaij. “Evaluation of context-aware recommendation systems for information re-finding.” In: *Journal of the Association for Information Science and Technology* 68.4 (2017), pp. 895–910. DOI: 10.1002/asi.23717. URL: https://msappelli.github.io/assets/pdf/2016_JASISTmanuscript_Evaluation%20of%20context-aware%20recommendation%20systems%20for%20information%20re-finding.pdf (cit. on p. 56).

[303] L. Sauermann. “The Gnowsis Semantic Desktop approach to Personal Information Management.” PhD thesis. TU Kaiserslautern, 2009. URL: https://www.dfki.de/fileadmin/user_upload/import/4261_Sauermann2009phd.pdf (cit. on pp. 21, 39, 53, 54, 63, 64).

[304] L. Sauermann, A. Bernardi and A. Dengel. “Overview and Outlook on the Semantic Desktop.” In: *Proceedings of the ISWC 2005 Workshop on The Semantic Desktop - Next Generation Information Management & Collaboration Infrastructure. Galway, Ireland, November 6, 2005*. Ed. by S. Decker, J. Park, D. Quan and L. Sauermann. Vol. 175. CEUR Workshop Proceedings. CEUR-WS.org, 2005, pp. 74–91. URL: http://ceur-ws.org/Vol-175/18_sauermann_overviewsemdesk_final.pdf (cit. on pp. xxix, xxxi, 4–6, 17, 18, 20, 21, 25, 51, 52, 91).

[305] L. Sauermann, A. Dengel, L. van Elst, A. Lauer, H. Maus and S. Schwarz. “Personalization in the EPOS Project.” In: *Proceedings of the International Workshop on Semantic Web Personalization, Budva, Montenegro, June 12, 2006*. Ed. by M. Bouzid and N. Henze. 2006, pp. 42–52. URL: https://www.dfki.uni-kl.de/~elst/papers/SemanticWebPersonalization@ESWC2006.pdf (cit. on pp. 22, 23, 52, 71, 148, 233).

[306] L. Sauermann, L. van Elst and A. Dengel. “PIMO – a Framework for Representing Personal Information Models.” In: *Proceedings of I-Media '07 and I-Semantics '07*. Know-Center, Austria, 2007, pp. 270–277. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.162.843&rep=rep1&type=pdf (cit. on pp. xxxi, 6, 22, 23, 52, 53, 56, 71, 148, 233).

[307] L. Sauermann, L. van Elst and K. Möller. *Personal Information Model (PIMO) Ontology V1.3*. Online. 2013. URL: https://www.semanticdesktop.org/ontologies/2007/11/01/pimo/. Also see http://www.semanticdesktop.org/ontologies/2007/11/01/pimo/v1.1/pimo_v1.1.pdf (cit. on pp. 23, 71, 148, 233, 271).

[308] L. Sauermann and D. Heim. "Evaluating Long-Term Use of the Gnowsis Semantic Desktop for PIM." In: *The Semantic Web – ISWC 2008*. 2008, pp. 467–482. DOI: 10.1007/978-3-540-88564-1_30. URL: https://link.springer.com/content/pdf/10.1007/978-3-540-88564-1_30.pdf (cit. on pp. 21, 39, 45, 53, 63, 64).

[309] K. Sauerwald, L. Reuter, C. Jilek, C. Beierle, J. O. Berndt, H. Dames, A. Dengel, D. Howey, G. Kern-Isberner, T. Kraemer, H. Maus, D. Nandini, C. Niederée, T. H. Nguyen, N. Potyka, M. Ragni, U. Schmid, S. Staab and I. J. Timm. "How to Intentionally Forget in AI Systems: A Multidisciplinary Approach to Belief Dynamics." Journal manuscript in preparation (preliminary title and author list). 2022 (cit. on pp. 58, 141).

[310] A. Savary and J. Piskorski. "Lexicons and grammars for named entity annotation in the National corpus of Polish." In: *18th International Conference Intelligent Information Systems, Siedlce, Poland*. 2010, pp. 141–154. URL: http://www.piskorski.waw.pl/papers/iis10-15.pdf (cit. on pp. 100, 102, 108).

[311] S. Scerri, J. Attard, I. Rivera, M. Valla and S. Handschuh. "DCON: Interoperable Context Representation for Pervasive Environments." In: *Proceedings of the AAAI Workshops, Technical Report WS-12-05*. 2012, pp. 90–97 (cit. on pp. 57, 91, 117).

[312] B. Schandl. "SemDAV: A File Exchange Protocol for the Semantic Desktop." In: *Proceedings of the Semantic Desktop and Social Semantic Collaboration Workshop at ISWC 2006, Athens, GA, USA*. Ed. by S. Decker, J. Park, L. Sauermann, S. Auer and S. Handschuh. Vol. 202. CEUR Workshop Proceedings. CEUR-WS.org, 2006. URL: http://ceur-ws.org/Vol-202/SEMDESK2006_0009.pdf (cit. on p. 61).

[313] J.-G. Schettler-Köhler. "SemanticFileExplorer: Semantisches Annotieren und Filtern im Windows Datei-Explorer [SemanticFileExplorer: Semantic Annotation and Filtering in the Windows File Explorer]." Bachelor's thesis. TU Kaiserslautern (in cooperation with DFKI), 2012 (cit. on pp. 26, 28).

[314] J.-G. Schettler-Köhler. "PimoCloud: a cloud-based, versioning document storage as a service for the PIMO." Master's thesis. TU Kaiserslautern (in cooperation with DFKI), 2014 (cit. on pp. 28, 29, 183).

[315] D. R. Schilling. *Knowledge Doubling Every 12 Months, Soon to be Every 12 Hours*. Weblog. Industry Tap – Tap Into News, Apr. 2013. URL: https://www.industrytap.com/knowledge-doubling-every-12-months-soon-to-be-every-12-hours/3950 (cit. on p. 14).

[316] A. Schmidt. "A Layered Model for User Context Management with Controlled Aging and Imperfection Handling." In: *Modeling and Retrieval of Context, Second International Workshop, MRC 2005, Edinburgh, UK, July 31-August 1, 2005, Revised Selected Papers*. Ed. by T. R. Roth-Berghofer, S. Schulz and D. B. Leake. Springer Berlin Heidelberg, 2006, pp. 86–100. DOI: 10.1007/11740674_6. URL: https://www.researchgate.net/profile/Andreas-Schmidt-11/publication/220743719_A_Layered_Model_for_User_Context_Management_with_Controlled_Aging_and_Imperfection_Handling/links/0912f50c7aa1d84d5e000000/A-Layered-Model-for-User-Context-Management-with-Controlled-Aging-and-Imperfection-Handling.pdf (cit. on p. 59).

[317] B. Schmidt. "Information Work Support Based on Activity Data." PhD thesis. TU Darmstadt, 2013. URL: https://tuprints.ulb.tu-darmstadt.de/id/eprint/3652 (cit. on pp. 54, 91).

[318] B. Schmidt and E. Godehardt. "Interaction Data Management." In: *Knowlege-Based and Intelligent Information and Engineering Systems*. Ed. by A König, A. Dengel, K. Hinkelmann, K. Kise, R. J. Howlett and L. C. Jain. Springer Berlin Heidelberg, 2011, pp. 402–409. DOI: 10.1007/978-3-642-23863-5_41 (cit. on pp. 54, 117).

[319] B. Schmidt, J. Kastl, T. Stoitsev and M. Mühlhäuser. "Hierarchical Task Instance Mining in Interaction Histories." In: *Proceedings of the 29th ACM international conference on Design of communication (SIGDOC '11), Pisa, Italy*. ACM, 2011, pp. 99–106. DOI: 10.1145/2038476.2038495. URL: https://fileserver.tk.informatik.tu-darmstadt.de/Publications/2011/p99-schmidt.pdf (cit. on pp. 54, 117, 118).

[320] B. Schmidt, T. Stoitsev and M. Mühlhäuser. "Task Models for Intention-Aware Systems." In: *Journal of Universal Computer Science (JUCS)*. Vol. 17. 10. Verlag der Technischen Universität Graz, 2011, pp. 1511–1526. DOI: 10.3217/jucs-017-10-1511. URL: https://fileserver.tk.informatik.tu-darmstadt.de/Publications/2011/jucs.pdf (cit. on pp. 54, 91).

[321] A. Schrijver. "On the History of the Shortest Path Problem." In: *Documenta Mathematica – Journal der Deutschen Mathematiker-Vereinigung* Extra Volume Optimization Stories, 21st International Symposium on Mathematical Programming (ISMP), Berlin, August 19–24, 2012 (2012). Ed. by M. Grötschel, pp. 155–167. URL: https://www.math.uni-bielefeld.de/documenta/vol-ismp/32_schrijver-alexander-sp.pdf (cit. on p. 120).

[322] M. Schröder. "Efficient High-Level Semantic Enrichment of Undocumented Enterprise Data." In: *The Semantic Web: ESWC 2019 Satellite Events*. Ed. by P. Hitzler, S. Kirrane, O. Hartig, V. de Boer, M.-E. Vidal, M. Maleshkova, S. Schlobach, K. Hammar, N. Lasierra, S. Stadtmüller, K. Hose and R. Verborgh. Springer International Publishing, 2019, pp. 220–230. URL: https://www.dfki.uni-kl.de/~mschroeder/data/ESWC2019-PhDSymposium_paper_4.pdf (cit. on p. 69).

[323] M. Schröder. *Building Knowledge Graphs from Messy Enterprise Data*. PhD thesis under review. TU Kaiserslautern, 2022 (cit. on pp. 9, 69, 89).

[324] M. Schröder, C. Jilek and A. Dengel. "Deep Linking Desktop Resources." In: *The Semantic Web: ESWC 2018 Satellite Events – ESWC 2018 Satellite Events, Heraklion, Crete, Greece, June 3-7, 2018, Revised Selected Papers*. Ed. by A. Gangemi, A. L. Gentile, A. G. Nuzzolese, S. Rudolph, M. Maleshkova, H. Paulheim, J. Z. Pan and M. Alam. Springer, 2018, pp. 202–207. DOI: 10.1007/978-3-319-98192-5_38. arXiv: 1805.03491. URL: https://arxiv.org/pdf/1805.03491.pdf (cit. on pp. 69, 75, 89).

[325] M. Schröder, C. Jilek and A. Dengel. "Interactive Concept Mining on Personal Data – Bootstrapping Semantic Services." In: *Computing Research Repository (CoRR)* (2019). arXiv: 1903.05872. URL: https://arxiv.org/pdf/1903.05872.pdf (cit. on pp. 69, 83, 84, 89, 225).

[326] M. Schröder, C. Jilek and A. Dengel. "A Linked Data Application Framework to Enable Rapid Prototyping." In: *Computing Research Repository (CoRR)* (2021). arXiv: 2104.13605. URL: https://arxiv.org/pdf/2104.13605.pdf (cit. on pp. 69, 90).

[327] M. Schröder, C. Jilek and A. Dengel. "Dataset Generation Patterns for Evaluating Knowledge Graph Construction." In: *The Semantic Web: ESWC 2021 Satellite Events – ESWC 2021 Satellite Events, Virtual Event, June 6-10, 2021, Revised Selected Papers*. Ed. by R. Verborgh, A. Dimou, A. Hogan, C. d'Amato, I. Tiddi, A. Bröring, S. Maier, F. Ongenae, R. Tommasini and M. Alam. Springer, 2021, pp. 27–32. DOI: 10.1007/978-3-030-80418-3_5. arXiv: 2104.13576. URL: https://arxiv.org/pdf/2104.13576.pdf (cit. on pp. 69, 90, 145).

[328] M. Schröder, C. Jilek and A. Dengel. "Mapping Spreadsheets to RDF: Supporting Excel in RML." In: *Proceedings of the 2nd International Workshop on Knowledge Graph Construction (KGCW 2021) at 18th Extended Semantic Web Conference (ESWC 2021), Virtual Event, June 6-10, 2021*. Ed. by D. Chaves-Fraga, A. Dimou, P. Heyvaert, F. Priyatna and J. Sequeda. Vol. 2873. CEUR Workshop Proceedings. CEUR-WS.org, 2021. arXiv: 2104.13600. URL: http://ceur-ws.org/Vol-2873/paper3.pdf (cit. on pp. 69, 89).

[329] M. Schröder, C. Jilek and A. Dengel. "Spread2RML: Constructing Knowledge Graphs by Predicting RML Mappings on Messy Spreadsheets." In: *Proceedings of the 11th International Conference on Knowledge Capture (K-CAP 2021), Virtual Event, USA, December 2-3, 2021*. ACM, 2021, pp. 145–152. DOI: 10.1145/3460210.3493544. arXiv: 2110.12829. URL: https://arxiv.org/pdf/2110.12829.pdf (cit. on pp. 69, 89, 320).

[330] M. Schröder, C. Jilek and A. Dengel. "A Human-in-the-Loop Approach for Personal Knowledge Graph Construction from File Names." In: *Proceedings of the 3rd International Workshop on Knowledge Graph Construction (KGCW 2022) at 19th Extended Semantic Web Conference (ESWC 2022), Hersonissos, Crete, Greece, May 29 - June 2, 2022*. Ed. by D. Chaves-Fraga, A. Dimou, P. Heyvaert, F. Priyatna and J. Sequeda. Vol. 3141. CEUR Workshop Proceedings. CEUR-WS.org, 2022. URL: http://ceur-ws.org/Vol-3141/paper2.pdf (cit. on pp. 69, 89).

[331] M. Schröder, C. Jilek, J. Hees and A. Dengel. "Towards Semantically Enhanced Data Understanding." In: *Computing Research Repository (CoRR)* (2018). arXiv: 1806.04952. URL: https://arxiv.org/pdf/1806.04952.pdf (cit. on pp. 69, 90).

[332] M. Schröder, C. Jilek, J. Hees, S. Hertling and A. Dengel. "RDF Spreadsheet Editor: Get (G)rid of Your RDF Data Entry Problems." In: *Proceedings of the ISWC 2017 Posters & Demonstrations and Industry Tracks co-located with 16th International Semantic Web Conference (ISWC 2017), Vienna, Austria, October 23-25, 2017*. Ed. by N. Nikitina, D. Song, A. Fokoue and P. Haase. Vol. 1963. CEUR Workshop Proceedings. CEUR-WS.org, 2017. URL: http://ceur-ws.org/Vol-1963/paper635.pdf (cit. on pp. 69, 89).

[333] M. Schröder, C. Jilek, J. Hees, S. Hertling and A. Dengel. "An Easy & Collaborative RDF Data Entry Method using the Spreadsheet Metaphor." In: *Computing Research Repository (CoRR)* (2018). arXiv: 1804.04175. URL: https://arxiv.org/pdf/1804.04175.pdf (cit. on pp. 69, 89).

[334] M. Schröder, C. Jilek, M. Schulze and A. Dengel. "Interactively Constructing Knowledge Graphs from Messy User-Generated Spreadsheets." In: *Computing Research Repository (CoRR)* (2021). arXiv: 2103.03537. URL: https://arxiv.org/pdf/2103.03537.pdf (cit. on pp. 69, 89).

[335] M. Schröder, C. Jilek, M. Schulze and A. Dengel. "The Person Index Challenge: Extraction of Persons from Messy, Short Texts." In: *Proceedings of the 13th International Conference on Agents and Artificial Intelligence (ICAART 2021), Online Streaming, February 4-6, 2021*. Ed. by A. P. Rocha, L. Steels and H. J. van den Herik. Vol. 2. SciTePress, 2021, pp. 531–537. DOI: 10.5220/0010188405310537. arXiv: 2011.07990. URL: https://www.scitepress.org/Papers/2021/101884/101884.pdf (cit. on pp. 69, 90, 145, 320).

[336] M. Schröder, M. Schulze, C. Jilek and A. Dengel. "Bridging the Technology Gap Between Industry and Semantic Web: Generating Databases and Server Code From RDF." In: *Proceedings of the 13th International Conference on Agents and Artificial Intelligence (ICAART 2021), Online Streaming, February 4-6, 2021*. Ed. by A. P. Rocha, L. Steels and J. van den Herik. Vol. 2. SciTePress, 2021, pp. 507–514. DOI: 10.5220/0010186005070514. arXiv: 2011.07957. URL: https://www.scitepress.org/Papers/2021/101860/101860.pdf (cit. on pp. 69, 90).

[337] J. Schuhmacher. "Challenges of designing intelligent support systems: Experimental study with a digital assistant for task resumption after interruptions." Master's thesis. TU Berlin (in cooperation with University of Trier and DFKI), 2020 (cit. on pp. 147, 166).

[338] M. Schulze, M. Pelzer, M. Schröder, C. Jilek, H. Maus and A. Dengel. "Towards Knowledge Graph Based Services in Accounting Use Cases." In: *Proceedings of Poster and Demo Track and Workshop Track of the 18th International Conference on Semantic Systems co-located with 18th International Conference on Semantic Systems (SEMANTiCS 2022)*. Ed. by U. Şimşek, D. Chaves-Fraga, T. Pellegrini and S. Vahdat. Vol. 3235. CEUR Workshop Proceedings. CEUR-WS.org, 2022. URL: https://ceur-ws.org/Vol-3235/paper19.pdf (cit. on p. 220).

[339] M. Schulze, M. Schröder, C. Jilek, T. Albers, H. Maus and A. Dengel. "P2P-O: A Purchase-To-Pay Ontology for Enabling Semantic Invoices." In: *The Semantic Web – 18th International Conference, ESWC 2021, Virtual Event, June 6-10, 2021, Proceedings*. Ed. by R. Verborgh, K. Hose, H. Paulheim, P.-A. Champin, M. Maleshkova, O. Corcho, P. Ristoski and M. Alam. Springer, 2021, pp. 647–663. DOI: 10.1007/978-3-030-77385-4_39. URL: https://openreview.net/pdf?id=50N86aYZ3n (cit. on p. 220).

[340] M. Schulze, M. Schröder, C. Jilek and A. Dengel. "ptpDG: A Purchase-To-Pay Dataset Generator for Evaluating Knowledge-Graph-Based Services." In: *Proceedings of the ISWC 2021 Posters & Demonstrations and Industry Tracks: From Novel Ideas to Industrial Practice co-located with 20th International Semantic Web Conference (ISWC 2021), Virtual Conference, October 24-28, 2021*. Ed. by O. Seneviratne, C. Pesquita, J. Sequeda and L. Etcheverry. Vol. 2980. CEUR Workshop Proceedings. CEUR-WS.org, 2021. URL: http://ceur-ws.org/Vol-2980/paper386.pdf (cit. on pp. 145, 220).

[341] S. Schwarz. “A Context Model for Personal Knowledge Management.” In: *Proceedings of 2nd International Workshop on Modeling and Retrieval of Context (MRC ’05) in conjuction with IJCAI-05*. Ed. by T. R. Roth-Berghofer, S. Schulz and D. B. Leake. Vol. 146. CEUR Workshop Proceedings. CEUR-WS.org, 2005. URL: http://ceur-ws.org/Vol-146/paper4.pdf (cit. on pp. 52, 69, 70, 91).

[342] S. Schwarz. “A Context Model for Personal Knowledge Management Applications.” In: *Modeling and Retrieval of Context – Second International Workshop, MRC 2005, Edinburgh, UK, July 31 - August 1, 2005, Revised Selected Papers*. Ed. by T. R. Roth-Berghofer, S. Schulz and D. B. Leake. Lecture Notes in Artificial Intelligence. Springer Berlin Heidelberg, 2006, pp. 18–33. DOI: 10.1007/11740674_2 (cit. on pp. 52, 69, 70, 91).

[343] S. Schwarz. “Context-Awareness and Context-Sensitive Interfaces for Knowledge Work Support.” PhD thesis. TU Kaiserslautern, 2010. ISBN: 978-3-86853-388-0 (cit. on pp. 38, 39, 54, 60, 61, 65, 69, 116, 137, 192).

[344] A. Seaborne and S. Harris. *SPARQL 1.1 Query Language*. W3C Recommendation. W3C, Mar. 2013. URL: https://www.w3.org/TR/2013/REC-sparql11-query-20130321/ (cit. on pp. xxxi, 19, 75).

[345] A. Seeliger, B. Schmidt, I. Schweizer and M. Mühlhäuser. “What Belongs Together Comes Together: Activity-Centric Document Clustering for Information Work.” In: *Proceedings of the 21st International Conference on Intelligent User Interfaces (IUI ’16), Sonoma, California, USA*. ACM, 2016, pp. 60–70. DOI: 10.1145/2856767.2856777. URL: https://fileserver.tk.informatik.tu-darmstadt.de/Publications/2016/TUD-CS-2016-0010.pdf (cit. on pp. 54, 117).

[346] SensAI Project Team. *SensAI – Self-organizing Personal Knowledge Assistants in Evolving Corporate Memories*. Project website. 2020. URL: https://comem.ai/sensai/ (cit. on pp. 54, 319).

[347] J. Shen. “Activity Recognition in Desktop Environments.” PhD thesis. Oregon State University, 2009. URL: https://ir.library.oregonstate.edu/dspace/bitstream/1957/11179/1/ActivityRecognition.pdf (cit. on pp. 56, 116).

[348] J. Shen, W. Geyer, M. Muller, C. Dugan, B. Brownholtz and D. R. Millen. “Automatically Finding and Recommending Resources to Support Knowledge Workers’ Activities.” In: *Proceedings of the 13th International Conference on Intelligent User Interfaces (IUI ’08), Gran Canaria, Spain*. ACM, 2008, pp. 207–216. DOI: 10.1145/1378773.1378801. URL: https://www.researchgate.net/profile/David-Millen/publication/221608121_Automatically_finding_and_recommending_resources_to_support_knowledge_workers'_activities/links/00b7d51a6e2193f94f000000/Automatically-finding-and-recommending-resources-to-support-knowledge-workers-activities.pdf (cit. on p. 56).

[349] J. Shen, L. Li and T. G. Dietterich. “Real-Time Detection of Task Switches of Desktop Users.” In: *Proceedings of the 20th International Joint Conference on Artificial Intelligence*. Vol. 7. Morgan Kaufmann, 2007, pp. 2868–2873. URL: https://www.aaai.org/Papers/IJCAI/2007/IJCAI07-460.pdf (cit. on pp. 56, 116).

[350] J. Shen, L. Li, T. G. Dietterich and J. L. Herlocker. "A Hybrid Learning System for Recognizing User Tasks from Desktop Activities and Email Messages." In: *Proceedings of the 11th International Conference on Intelligent User Interfaces (IUI '06), Sydney, Australia*. ACM, 2006, pp. 86–92. DOI: 10.1145/1111449.1111473. URL: https://web.engr.oregonstate.edu/~tgd/publications/IUI06-TaskPredictor.pdf (cit. on pp. 56, 116).

[351] Y. Shen, B. Guo, Y. Shen, X. Duan, X. Dong and H. Zhang. "A pricing model for Big Personal Data." In: *Tsinghua Science and Technology* 21.5 (Oct. 2016), pp. 482–490. DOI: 10.1109/TST.2016.7590317. URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7590317 (cit. on p. 81).

[352] D. Shin. "The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI." In: *International Journal of Human-Computer Studies* 146 (2021). DOI: 10.1016/j.ijhcs.2020.102551. URL: https://www.academia.edu/download/66687074/IJHCS_2020_Explainability.pdf (cit. on p. 208).

[353] B. Shneiderman. "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations." In: *Proceedings of IEEE Symposium on Visual Languages 1996 (VL '96)*. IEEE. 1996, pp. 336–343. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.445.8909&rep=rep1&type=pdf (cit. on p. 3).

[354] M. Siebers, K. Göbel, C. Niessen and U. Schmid. "Requirements for a companion system to support identifying irrelevancy." In: *2017 International Conference on Companion Technology (ICCT)*. July 2017, pp. 1–2. DOI: 10.1109/COMPANION.2017.8287076. URL: https://cogsys.uni-bamberg.de/publications/requirements_companion_ICCT2017.pdf (cit. on p. 58).

[355] M. Siebers and U. Schmid. "Please delete that! Why should I?" In: *KI – Künstliche Intelligenz, German Journal of Artificial Intelligence - Organ des Fachbereichs "Künstliche Intelligenz" der Gesellschaft für Informatik e.V.* 33.1 (Mar. 2019), pp. 35–44. DOI: 10.1007/s13218-018-0565-5. URL: https://fis.uni-bamberg.de/bitstream/uniba/46644/1/fisba46644.pdf (cit. on p. 58).

[356] R. Sinclair. *Microsoft Active Accessibility: Architecture*. Microsoft Docs. Microsoft Corporation, May 2000. URL: https://msdn.microsoft.com/en-us/library/ms971310.aspx (cit. on p. 93).

[357] R. Sollfrank. "Die "Superkraft" des Vergessens digital managen [Digitally Managing the "Super-human Power" of Forgetting]." In: *Die Wirtschaftszeitung* (Jan. 2021). URL: https://www.m-medienfabrik.de/epaper/karriere-2020/#8 (cit. on pp. 214, 321).

[358] L. Sommer. *Künstliche Intelligenz am Arbeitsplatz [Artificial Intelligence at the Workplace]*. enviaM-Gruppe Blog (Online). envia Mitteldeutsche Energie AG, May 2021. URL: https://blog.enviam.de/kuenstliche-intelligenz-am-arbeitsplatz/ (cit. on pp. 26, 319).

[359] SPP 1921 Project Coordination Team (Chair of Business Psychology, Ruhr-Universität Bochum). *'Intentional Forgetting in Organizations' as a DFG priority program (SPP 1921)*. Project website. 2016. URL: http://www.spp1921.de/ (cit. on pp. 33, 58).

[360] SRI International's Artificial Intelligence Center. *CALO: Cognitive Assistant that Learns and Organizes*. Project website. Originally online at `http://www.ai.sri.com/project/CALO` (not available anymore). Archived copy of 2022-03-08. 2003. URL: https://web.archive.org/web/20220308041153/http://www.ai.sri.com/project/CALO. Also see https://www.sri.com/hoi/artificial-intelligence-calo/ (cit. on p. 55).

[361] Stack Exchange. *Stack Overflow Developer Survey 2021: Programming, scripting, and markup languages*. Online. May 2021. URL: https://insights.stackoverflow.com/survey/2021#section-most-popular-technologies-programming-scripting-and-markup-languages (cit. on p. 95).

[362] T. Stajner and D. Mladenic. "Modeling Knowledge Worker Activity." In: *Proceedings of the First Workshop on Applications of Pattern Analysis, Sep 1-3, 2010*. Ed. by T. Diethe, N. Cristianini and J. Shawe-Taylor. Vol. 11. Proceedings of Machine Learning Research. Cumberland Lodge, Windsor, UK: PMLR, 2010, pp. 127–133. URL: http://proceedings.mlr.press/v11/stajner10a/stajner10a.pdf (cit. on pp. 56, 116).

[363] T. Stajner, D. Mladenic and M. Grobelnik. "Exploring Contexts and Actions in Knowledge Processes." In: *Proceedings of the Second Workshop on Context, Information and Ontologies (CIAO-2010), Lisbon, Portugal, October 11, 2010*. Ed. by V. Ermolayev, J. M. Gomez-Perez, P. Haase and P. Warren. Vol. 626. CEUR Workshop Proceedings. CEUR-WS.org, 2010. URL: http://ceur-ws.org/Vol-626/regular4.pdf (cit. on pp. 56, 116).

[364] T. Stajner, D. Mladenic and M. Grobelnik. "Information Resource Recommendation in Knowledge Processes." In: *The Semantic Web: ESWC 2011 Workshops*. Ed. by R. García-Castro, D. Fensel and G. Antoniou. Springer Berlin Heidelberg, 2012, pp. 186–193. URL: https://link.springer.com/content/pdf/10.1007/978-3-642-25953-1_15.pdf (cit. on p. 56).

[365] B. C. Storm and S. M. Stone. "Saving-enhanced memory: The benefits of saving on the learning and remembering of new information." In: *Psychological Science* 26.2 (2015), pp. 182–188. DOI: 10.1177/0956797614559285. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.940.6563&rep=rep1&type=pdf (cit. on pp. 35, 161).

[366] Student. "The Probable Error of a Mean." In: *Biometrika* 6.1 (Mar. 1908), pp. 1–25. DOI: 10.1093/biomet/6.1.1. URL: https://bayes.wustl.edu/Manual/Student.pdf (cit. on pp. 164, 171).

[367] R. Studer, V. R. Benjamins and D. Fensel. "Knowledge engineering: Principles and methods." In: *Data & Knowledge Engineering* 25.1 (1998), pp. 161–197. DOI: 10.1016/S0169-023X(97)00056-6. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.97.3940&rep=rep1&type=pdf (cit. on p. 20).

[368] S. Stumpf, X. Bao, A. Dragunov, T. G. Dietterich, J. Herlocker, K. Johnsrude, L. Li and J. Shen. "Predicting User Tasks: I Know What You're Doing!" In: *20th National Conference on Artificial Intelligence (AAAI-05), workshop on Human Comprehensible Machine Learning*. 2005. URL: https://www.aaai.org/Papers/Workshops/2005/WS-05-04/WS05-04-003.pdf (cit. on pp. 56, 116).

[369] SuGraBo Project Team. *SuGraBo: Suchmaschine für die grafische Benutzeroberfläche [SuGraBo: Search Engine for Graphical User Interfaces]*. Project website. 2016. URL: https://comem.ai/sugrabo/ (cit. on pp. 93, 319).

[370] supSpaces Consortium. *supSpaces – Semantische Support-Wissensräume [supSpaces – Semantic Support (Knowledge) Spaces]*. Project website. Originally online at http://www.supspaces.de/ (not available anymore). Archived copy of 2019-01-09. 2015. URL: https://web.archive.org/web/20190109230722/http://www.supspaces.de/ (cit. on pp. 54, 80, 84, 151, 186, 213–215, 319).

[371] SWELL Consortium. *SWELL: Smart Reasoning Systems for Well-being at Work and at Home*. Project website. Originally online at http://swell-project.net/ (not available anymore). Archived copy of 2021-06-12. 2011. URL: https://web.archive.org/web/20210612234945/http://swell-project.net/. Also see https://www.commit-nl.nl/projects/swell-smart-reasoning-systems-for-well-being-at-work-and-at-home (cit. on pp. 56, 142).

[372] T. Tempel and C. Frings. "Directed forgetting benefits motor sequence encoding." In: *Memory & Cognition* 44.3 (2016), pp. 413–419. DOI: 10.3758/s13421-015-0565-8. URL: https://link.springer.com/content/pdf/10.3758/s13421-015-0565-8.pdf (cit. on p. 161).

[373] T. Tempel, C. Niederée, C. Jilek, A. Ceroni, H. Maus, Y. Runge and C. Frings. "Temporarily Unavailable: Memory Inhibition in Cognitive and Computer Science." In: *Interacting with Computers* 31.3 (May 2019), pp. 231–249. DOI: 10.1093/iwc/iwz013. arXiv: 1912.00760. URL: https://arxiv.org/pdf/1912.00760.pdf (cit. on pp. 4, 33, 35, 211, 218, 219, 225, 227).

[374] I. J. Timm, S. Staab, M. Siebers, C. Schon, U. Schmid, K. Sauerwald, L. Reuter, M. Ragni, C. Niederée, H. Maus, G. Kern-Isberner, C. Jilek, P. Friemann, T. Eiter, A. Dengel, H. Dames, T. Bock, J. O. Berndt and C. Beierle. "Intentional Forgetting in Artificial Intelligence Systems: Perspectives and Challenges." In: *KI 2018: Advances in Artificial Intelligence – 41st German Conference on AI, Berlin, Germany, September 24–28, 2018, Proceedings*. Springer, 2018, pp. 357–365. DOI: 10.1007/978-3-030-00111-7_30. URL: https://fis.uni-bamberg.de/bitstream/uniba/46646/1/fisba46646.pdf (cit. on pp. 33, 58).

[375] E. F. Tjong Kim Sang. "Introduction to the CoNLL-2002 Shared Task: Language-Independent Named Entity Recognition." In: *COLING-02: The 6th Conference on Natural Language Learning 2002 (CoNLL-2002)*. 2002. URL: https://aclanthology.org/W02-2024.pdf (cit. on pp. 109, 110).

[376] T. A. Tran, S. Schwarz, C. Niederée, H. Maus and N. Kanhabua. "The Forgotten Needle in My Collections: Task-Aware Ranking of Documents in Semantic Information Space." In: *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval (CHIIR '16), Carrboro, North Carolina, USA*. ACM, 2016, pp. 13–22. DOI: 10.1145/2854946.2854971. URL: https://homes.cs.aau.dk/~nattiya/papers/CHIIR2016-task-aware-ranking.pdf (cit. on pp. 126, 225).

[377] P. Trkman and K. McCormack. "Estimating the Benefits and Risks of Implementing E-Procurement." In: *IEEE Transactions on Engineering Management* 57.2 (2010), pp. 338–349. DOI: 10.1109/TEM.2009.2033046. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.669.3035&rep=rep1&type=pdf (cit. on p. 220).

[378] J. W. Tukey. *Exploratory Data Analysis*. Addison-Wesley, 1977. ISBN: 9780201076165 (cit. on p. 262).

[379] L. Turczyk, M. Groepl, N. Liebau and R. Steinmetz. "A method for file valuation in information lifecycle management." In: *AMCIS 2007 Proceedings* (2007), p. 38. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.891.8088&rep=rep1&type=pdf (cit. on p. 125).

[380] L. A. Turczyk. "Information Lifecycle Management – Eine Methode zur Wertzuweisung von Dateien [A Method for Valuation of Files]." PhD thesis. TU Darmstadt, 2009 (cit. on p. 125).

[381] L. A. Turczyk, O. Heckmann, R. Berbner and R. Steinmetz. "A Formal Approach to Information Lifecycle Management." In: *Proceedings of 17th Annual IRMA International Conference, Washington DC*. 2006. URL: https://www.kom.tu-darmstadt.de/papers/THBSt06_675.pdf (cit. on p. 125).

[382] L. A. Turczyk, O. Heckmann and R. Steinmetz. "File Valuation in Information Lifecycle Management." In: *Managing Worldwide Operations and Communications with Information Technology, 2007 IRMA International Conference*. Ed. by M. Khosrow-Pour. 2007, pp. 347–351. URL: https://www.kom.tu-darmstadt.de/papers/THSt07_649.pdf (cit. on p. 125).

[383] A.-S. Ulfert-Blank, J. Knabe and C. Jilek. "Forgetting-enabled AI systems: Exploring the role of transparency perceptions and trust." Short paper accepted at 24th International Conference on Human-Computer Interaction (HCII 2022) but withdrawn due to unforeseen scheduling conflicts. Extended journal manuscript in preparation. 2022 (cit. on pp. 141, 206, 207, 227).

[384] Unicode Consortium. *The Unicode Standard Version 14.0 – Core Specification, Chapter 21.1 Braille*. Online. Sept. 2021. URL: https://www.unicode.org/versions/Unicode14.0.0/ch21.pdf (cit. on p. 156).

[385] Unicode Consortium. *FAQ – UTF and BOM*. Online. 2022. URL: https://www.unicode.org/faq/utf_bom.html#BOM (cit. on p. 154).

[386] A. van Drunen, E. L. van den Broek and T. Heffelaar. "uLog: Towards attention aware user-system interactions measurement." In: *Proceedings of Measuring Behavior 2008, 6th International Conference on Methods and Techniques in Behavioral Research, MECC Maastricht, Maastricht, Netherlands, Aug 26-29, 2008*. Ed. by A. Spink, M. Ballintijn, N. Bogers, F. Grieco, L. Loijens, L. Noldus, G. Smit and P. Zimmerman. Noldus Information Technology, Aug. 2008, pp. 109–110. URL: https://research.utwente.nl/files/208429083/Drunen2008ulog.pdf. (Implementation available as *uLog* [Sw94].) (Cit. on p. 270).

[387] K. Voit, K. Andrews and W. Slany. "Why personal information management (PIM) technologies are not widespread." In: *PIM Workshop, ASIS&T*. 2009, pp. 60–64. URL: http://pimworkshop.org/2009/papers/voit-pim2009.pdf (cit. on pp. 18, 37).

[388] Wacom Co., Ltd. *Wacom and DFKI realize enhanced semantic technologies with digital ink, combining Wacom's WILL 3.0 and DFKI's AI expertise*. Online. Wacom Co., Ltd., Nov. 2018. URL: https://www.wacom.com/en-us/about-wacom/news-and-events/2018/1335 (cit. on pp. 26, 212, 319).

[389] Wacom Co., Ltd. *Wacom – Interactive pen displays, pen tablets and stylus products*. Corporate website. 2022. URL: https://www.wacom.com/ (cit. on pp. 26, 212).

[390] Q. Wang and H. Chang. "Multitasking Bar: Prototype and Evaluation of Introducing the Task Concept into a Browser." In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'10), Atlanta, Georgia, USA*. ACM, 2010, pp. 103–112. DOI: 10.1145/1753326.1753343. URL: https://www.sysu-hcp.net/userfiles/files/2021/02/28/fe10b4e611a8cdec.pdf (cit. on p. 169).

[391] P. Warren. "Personal Information Management: The Case for an Evolutionary Approach." In: *Interacting with Computers* 26.3 (July 2013), pp. 208–237. DOI: 10.1093/iwc/iwt034. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.975.4643&rep=rep1&type=pdf (cit. on pp. 16–18, 37, 44).

[392] P. Warren, I. Thurlow, J. Davies, I. Dolinsek, C. Ruiz and J. M. Gomez-Perez. "Context as a Tool for Organizing and Sharing Knowledge." In: *Proceedings of the Second Workshop on Context, Information and Ontologies (CIAO-2010), Lisbon, Portugal, October 11, 2010*. Ed. by V. Ermolayev, J. M. Gomez-Perez, P. Haase and P. Warren. Vol. 626. CEUR Workshop Proceedings. CEUR-WS.org, 2010. URL: http://ceur-ws.org/Vol-626/regular1.pdf (cit. on pp. 56, 61, 62).

[393] T. Warren. *The story of Cortana, Microsoft's Siri killer*. Online. TheVerge.com, Apr. 2014. URL: https://www.theverge.com/2014/4/2/5570866/cortana-windows-phone-8-1-digital-assistant (cit. on p. 55).

[394] T. Warren. *Microsoft shuts down Cortana on iOS and Android*. Online. TheVerge.com, Apr. 2021. URL: https://www.theverge.com/2021/4/1/22361687/microsoft-cortana-shut-down-ios-android-mobile-app (cit. on p. 55).

[395] R. Watson, S. Dekeyser and N. Albadri. "Exploring the Design Space of Metadata-Focused File Management Systems." In: *Proceedings of the Australasian Computer Science Week Multiconference (ACSW '17), Geelong, Australia*. ACM, 2017. DOI: 10.1145/3014812.3014833. URL: https://core.ac.uk/download/pdf/211500287.pdf (cit. on p. 62).

[396] K. Weitz, D. Schiller, R. Schlagowski, T. Huber and E. André. ""Let me explain!": exploring the potential of virtual agents in explainable AI interaction design." In: *Journal on Multimodal User Interfaces* 15.2 (June 2021), pp. 87–98. DOI: 10.1007/s12193-020-00332-0. URL: https://link.springer.com/content/pdf/10.1007/s12193-020-00332-0.pdf (cit. on p. 208).

[397] F. Wijnhoven, C. Amrit and P. Dietz. "Value-Based File Retention: File Attributes As File Value and Information Waste Indicators." In: *Journal of Data and Information Quality* 4.4 (2014), 15:1–15:17. DOI: 10.1145/2567656. URL: https://www.researchgate.net/profile/Chintan-Amrit/publication/274483776_Value-Based_File_Retention/links/5851183c08aecb6bd8d21d2e/Value-Based-File-Retention.pdf (cit. on p. 125).

[398] Wikimedia Foundation. *Wikipedia: The Free Encyclopedia*. Online. 2001. URL: https://www.wikipedia.org/ (cit. on pp. 98, 99, 108, 162).

[399] Wikimedia Foundation. *Wiktionary: Free dictionary*. Online. 2002. URL: https://www.wiktionary.org/ (cit. on p. 101).

[400] Wikimedia Foundation. *List of Wikipedias / Table*. Wikimedia Meta-Wiki. State of 2016-11-01. Nov. 2016. URL: https://meta.wikimedia.org/w/index.php?title=List_of_Wikipedias/Table&oldid=16025729. (Includes statistics corresponding to a *German Wikipedia dump* [Sw122] downloaded at the time.) (Cit. on pp. 225, 272).

[401] M. K. Wolters, E. Niven, M. Runardotter, F. Gallo, H. Maus and R. H. Logie. "Personal Photo Preservation for the Smartphone Generation." In: *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI-15)*. ACM, Apr. 2015, pp. 1549–1554. DOI: 10.1145/2702613.2732793. URL: https://www.dfki.uni-kl.de/~maus/dok/WoltersNivenRunardotter+2015.pdf (cit. on pp. 135, 137).

[402] H. Xiao and I. F. Cruz. "A Multi-Ontology Approach for Personal Information Management." In: *Proceedings of the ISWC 2005 Workshop on The Semantic Desktop - Next Generation Information Management & Collaboration Infrastructure. Galway, Ireland, November 6, 2005*. Ed. by S. Decker, J. Park, D. Quan and L. Sauermann. Vol. 175. CEUR Workshop Proceedings. CEUR-WS.org, 2005, pp. 1–15. URL: http://ceur-ws.org/Vol-175/32_xiaocruz_multiontology_final.pdf (cit. on p. 57).

[403] R. R. Yager and L. Liu. *Classic Works of the Dempster-Shafer Theory of Belief Functions*. Springer Berlin, Heidelberg, 2008. ISBN: 978-3-540-25381-5. DOI: 10.1007/978-3-540-44792-4 (cit. on pp. 137, 192).

[404] YAML Language Development Team. *YAML Ain't Markup Language (YAML) version 1.2, Revision 1.2.2*. Online. Oct. 2021. URL: https://yaml.org/spec/1.2.2/ (cit. on pp. xxxi, 146).

[405] T. Zesch, C. Müller and I. Gurevych. "Extracting Lexical Semantic Knowledge from Wikipedia and Wiktionary." In: *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*. Ed. by N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis and D. Tapias. European Language Resources Association (ELRA), May 2008, pp. 1646–1652. ISBN: 2-9517408-4-0. URL: http://lrec-conf.org/proceedings/lrec2008/pdf/420_paper.pdf. (Implementation available as *Java-based Wiktionary Library (JWKTL)* [Sw115].) (Cit. on pp. 11, 272).

[406] Z. Zhang, J. Iria, C. Brewster and F. Ciravegna. "A Comparative Evaluation of Term Recognition Algorithms." In: *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08), Marrakech, Morocco*. European Language Resources Association (ELRA), May 2008. URL: http://www.lrec-conf.org/proceedings/lrec2008/pdf/538_paper.pdf. (Implementation available as *JATE 1.11* [Sw125].) (Cit. on p. 272).

[407] X. Zhu, C. Niederée, T. Tran, A. Ceroni, K. D. Naini, N. K. Tran, H. Maus and C. Jilek. *ForgetIT Deliverable D3.4: Strategies and Components for Managed Forgetting*. Deliverable. ForgetIT Consortium, Mar. 2016. URL: https://www.forgetit-project.eu/fileadmin/fm-dam/deliverables/ForgetIT_WP3_D3.4.pdf (cit. on pp. 4, 125, 191, 227).

Academic Curriculum Vitae: Christian Jilek

E-mail:	christian.jilek@dfki.de / c_jilek@cs.uni-kl.de
Website:	https://www.dfki.uni-kl.de/~jilek/
ORCiD:	https://orcid.org/0000-0002-5926-1673
Google Scholar:	https://scholar.google.com/citations?user=F4J63d4AAAAJ

Experience

German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Smart Data & Knowledge Services Research Dept. (Prof. Dr. Andreas Dengel)

since 2015 Researcher and PhD student

- Project leader and contributions to research and development:
  - SensAI [346] (BMBF, 2020–2023),
  - Managed Forgetting Phase II [234] (DFG, 2020–2023), and
  - Managed Forgetting [234] (DFG, 2016–2020).
- Contributions to research and development:
  - SuGraBo [369] (BMBF, 2016–2017),
  - supSpaces [370] (BMBF, 2015–2017),
  - ForgetIT [120] (EU, 2013–2016), and
  - CoMem [87] (ongoing meta project, 2011–).
- Contributions to research, development and consulting in 11 other research and industry projects, e.g.
  - Semantic Technologies with Digital Ink [388] (Wacom Co., Ltd., 2018–),
  - CoMem @ enviaM [358] (Envia Mitteldeutsche Energie AG, 2017–)
- Supporting the teaching activities of the Artificial Intelligence Group, Computer Science Dept., TU Kaiserslautern (Prof. Dr. Andreas Dengel):
  - Guest lecturer (3 lectures) and (co-)supervision of 12 bachelor/master theses, 8 seminar papers, 15 student projects and 2 pupil projects.
  - Developer of *Preferant: The Preference Ant*, an online tool for preference votings (e.g. used for seminar/project topic assignments)

2014 Diploma thesis and student research assistant

Chair of Business Inf. Systems and Operations Research (BISOR), Business Studies and Economics Dept., TU Kaiserslautern (Prof. Dr. Oliver Wendt)

2012 – 2014 Student research assistant

John Deere GmbH & Co. KG, John Deere Werk Zweibrücken, Advanced Engineering Department

2010 – 2014 Working student

Education

TU Kaiserslautern (TUK)

2003 – 2014 Studies of *Computer Science* and *Economics*

- Major subjects: *Intelligent Systems*, *Software Engineering* and *Business Information Systems / Operations Research*
- Selected works:
  - *PIMO Diary: Diary Generation from Personal Information Models* [178], Diploma thesis in cooperation with the German Research Center for Artificial Intelligence (DFKI)
  - *Usability Evaluation of the Communication and Information System (KIS) of TU Kaiserslautern using Usability Tests*, Student project in cooperation with the Research Institute of Technology and Work (ITA)
  - *Design and Implementation of a Publication Workflow Management System*, Student project in cooperation with the Fraunhofer Institute for Experimental Software Engineering (FhG IESE)
  - *Model-driven Software Development*, Seminar paper
  - *A Memetic Algorithm for the Vehicle Routing Problem with Soft Time Windows*, Student research paper
  - *A Hybrid Tabu Search/Variable Neighborhood Search Solution Method for the Electric Vehicle Routing Problem with Time Windows and Mixed Fleet and Driving Speed and Load Considerations*, Student research paper
- Prediplomas in *Computer Science* and *Industrial Engineering/Computer Science* (changed course of studies)
- Diploma in *Industrial Engineering/Computer Science* (best final grade of this course of studies in 2014)

Paper Awards and Nominations

2021 M. Schröder, C. Jilek and A. Dengel [329]
received Best Paper Award
at 11th International Conference on Knowledge Capture (K-CAP 2021)

M. Schröder, C. Jilek, M. Schulze and A. Dengel [335]
nominated for Best Poster Award
at 13th International Conference on Agents and Artificial Intelligence (ICAART 2021)

2019 C. Jilek, M. Schröder, R. Novik, S. Schwarz, H. Maus and A. Dengel [186]
received Best Research Paper Award
at 2nd Conference on Language, Data and Knowledge (LDK 2019)

2017 S. Hertling, M. Schröder, C. Jilek and A. Dengel [164]
received Best Paper Award (Area: Artificial Intelligence)
at 9th International Conference on Agents and Artificial Intelligence (ICAART 2017)

2016 S. Hertling, M. Schröder, C. Jilek and A. Dengel [163]
won Top-k Shortest Paths in Large Typed RDF Graphs Challenge
at 13th Extended Semantic Web Conference (ESWC 2016)

Selected Talks and Presentations

2023 *Teaching Computers to Forget*,
Newspaper Interview by J. Luttenberger, Die Rheinpfalz, online [233]

2021 *SensAI: Personal and Organizational Knowledge Graphs for Context-specific Assistance*,
Invited Talk at Network AI Research Germany, online

Digitally Managing the "Super-human Power" of Forgetting,
Newspaper Interview by R. Sollfrank, Die Wirtschaftszeitung, online [357]

2020 *Managed Forgetting: Towards a Personal Computer that Tidies Up Itself*,
Invited Talk at Sensorik X. HR-Expertenforum, Regensburg [181]

Psychology Meets AI Using the Example of Forgetful & Self-(re)organizing Knowledge Assistants,
Invited Talk at Institute of Psychology, Goethe University Frankfurt am Main

2019 *Self-organizing Context Spaces to Support Information Management and Knowledge Work*,
Invited Talk at Department of Computer Science, University of Trier

Inflection-tolerant Ontology-based Named Entity Recognition for Real-time Applications,
Invited Talk at 13th DBpedia Community Meeting, Leipzig [180]

Self-organizing Context Spaces to Support Information Management and Knowledge Work,
Talk at Department of Computer Science, TU Kaiserslautern

2017 *Programmed Forgetting: If Machines Learn and Help to Forget*,
Invited Talk at Department of Psychology, University of Trier

2016 *supSpaces: Semantic Support Spaces for Agile Knowledge Management in 3rd-level IT Support*,
Poster Presentation at KMU-innovativ: IKT Symposium 2016, Hannover [179]

2015 *ForgetIT: Concise Preservation by Combining Managed Forgetting and Contextualized Remembering*,
Demo Presentation at CeBIT 2015, Hannover [239]

Peer Reviewing

2022 *ACM CHI Conference on Human Factors in Computing Systems 2022 (CHI 2022) (Late-Breaking Work)*, New Orleans, Louisiana, USA

2018 *20th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI 2018)*, Barcelona, Spain

Advances in Human-Computer Interaction (AHCI), Journal (sub-reviewer)

2017 *4th Workshop on Ubiquitous Technologies for Augmenting the Human Mind – Sharing Experiences (WAHM 2017 @ UbiComp 2017)*, Maui, Hawaii, USA

2016 *3rd Workshop on Ubiquitous Technologies for Augmenting the Human Mind (WAHM 2016 @ UbiComp 2016)*, Heidelberg, Germany (sub-reviewer)