Explanation of Similarities in Process-Oriented Case-Based Reasoning by Visualization*

Alexander Schultheis[1], Maximilian Hoffmann[1,2], Lukas Malburg[1,2], and Ralph Bergmann[1,2]

[1] German Research Center for Artificial Intelligence (DFKI) Branch University of Trier, Behringstraße 21, 54296 Trier, Germany
[2] Artificial Intelligence and Intelligent Information Systems, University of Trier, 54296 Trier, Germany, http://www.wi2.uni-trier.de
{alexander.schultheis,maximilian.hoffmann,lukas.malburg, ralph.bergmann}@dfki.de
{hoffmannm,malburgl,bergmann}@uni-trier.de

Abstract Modeling similarity measures in Case-Based Reasoning is a knowledge-intensive, demanding, and error-prone task even for domain experts. Visualizations offer support for users, but are currently only available for certain subdomains and case representations. Currently, there are only visualizations that can be used for local attributes or specific case representations. However, there is no possibility to visualize similarities between complete processes accordingly so far, although complex domains may be present. Therefore, an extension of existing approaches or the design of new suitable concepts for this application domain is necessary. The contribution of this work is to enable a more profound understanding of similarity for knowledge engineers who create a similarity model and support them in this task by using visualization methods in *Process-Oriented Case-Based Reasoning* (POCBR). For this purpose, we present related approaches and evaluate them against derived requirements for visualizations in POCBR. On this basis, suitable visualizations are further developed as well as new approaches designed. Three such visualizations are created: (1) a graph mapping approach, (2) a merge graph, and (3) a visualization based on heatmaps. An evaluation of these approaches has been performed based on the requirements in which the domain experts determine the graph-mapping visualization as best-suited for engineering of similarity models.

Keywords: Visualization · Explanation · Similarity · Process-Oriented Case-Based Reasoning · Explainable Case-Based Reasoning

1 Introduction

In *Case-Based Reasoning* (CBR) [1], cases are retrieved by approximating the a-posteriori utility of a case with mostly knowledge-intensive a-priori similarity

* The final authenticated publication is available online at https://doi.org/10.1007/978-3-031-40177-0_4

measures [5, 37]. Developing such similarity measures can be performed manually by a domain expert or learned automatically, e. g., with machine learning methods [18, 29]. Whereas automatically generated similarity measures limit the knowledge acquisition and modeling effort, the entire required knowledge cannot always be learned and, thus, integrated into the similarity measure. For this reason, similarity measures are mainly created manually by a knowledge engineer, based on information from domain experts. However, this knowledge-intensive acquisition process is a demanding and error-prone task, since the knowledge from the domain expert must be acquired, transferred, and, finally, encoded by the knowledge engineer during similarity measure development. This especially holds for complex and deeply nested measures based on the local-global principle. For instance, a global similarity measure usually aggregates individual local similarities between several attributes, which, in turn, use further similarity measures for different data types [5], so that this nested structure can be cluttered for a knowledge engineer. In this context, a profound understanding of the similarity and its aggregation results in a better explainability that increases trust and transparency of the CBR system [12]. Although similarity-based retrieval is one of the most important phases in CBR, and, thus, similarity measures should be well-defined, the development process is only weakly supported by current research (e. g., [3]). Therefore, the knowledge engineer should be supported in manual modeling of complex and knowledge-intensive similarity measures. An understanding of the similarity is necessary for this, including both the value of the similarity itself and its calculation.

One approach for making similarities more accessible and providing confidence in CBR applications is *visualization* [3]. Visualizations map discrete data to a visual representation and support users in problem-solving [26]. In the context of *Explainable Case-Based Reasoning* (XCBR), they are successfully used in CBR domains to investigate similarities between simple case representations [39]. Of particular interest is to make complex similarities, e. g., based on nested similarity measures, accessible using visualizations. Another application area is *Process-Oriented Case-Based Reasoning* (POCBR) [31] in which CBR methods are applied to procedural experiential knowledge. Cases in POCBR are expressed as complex graph representations, and similarity assessment is mainly performed by aggregating several local similarities on a global level. Especially these complex case representations are currently not supported by any visualization method. For this purpose, this paper presents how visualizations can be used in POCBR and, in general, for more complex case representations that go beyond attribute-value-based cases. To determine the eligibility of a visualization for this purpose, we survey fundamental requirements from CBR developers. We present three visualization methods for complex cases in POCBR based on these requirements, namely a *Graph Mapping*, a *Merge Graph*, and a *Heatmap* approach. We compare them regarding the requirements in an experimental evaluation to validate their suitability. The key question for CBR developers in this evaluation is how well which requirement is being fulfilled by the corresponding visualization method and, whether, the approaches provide a better explainability of similarities for knowledge engineers.

The paper is structured as follows: First, an overview of the necessary foundations of POCBR is provided, and related work is discussed (see Sect. .2). Then requirements for visualizations in POCBR are established by interviews with experts and from literature (see Sect. 3). Based on these requirements, applicable approaches from the literature are examined and adapted for the use in POCBR (see Sect. 4). The suitability of the approaches is determined based on a user evaluation (see Sect. 5). Finally, the paper is concluded and areas for future work are discussed (see Sect. 6).

2 Foundations and Related Work

In this section, the semantic workflow graph representation as well as the similarity assessment between such workflow graphs is introduced (see Sect. 2.1). Afterward, the research topic of this work is differentiated to other works, presenting approaches that are thematically relevant or not applicable w. r. t. this question (see Sect. 2.2 and Sect. 2.3).

2.1 Semantic Workflow Representation and Similarity Assessment

The proposed visualizations address process-oriented cases. In the research area of POCBR, various types of graphs exist [20] as well as multiple applicable similarity measures for these [9, 34]. To be able to perform the requirements analysis and design the visualizations, a concrete form of case representation must be addressed. So, this paper uses semantic workflow graphs represented as *NEST* graphs [7] for this purpose and aims at similarity measures based on graph matching.

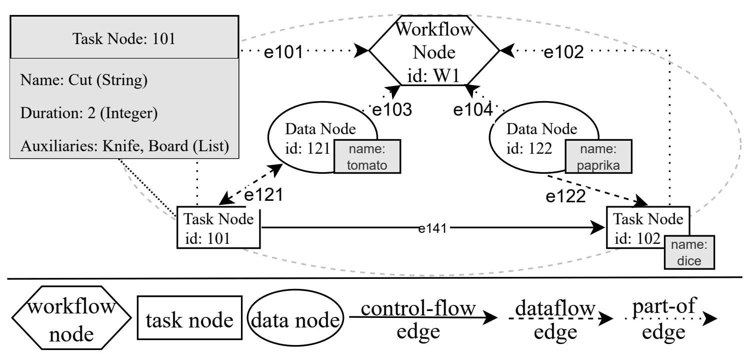

Fig. 1. An Exemplary Cooking Recipe represented as *NEST* Graph.

A NEST graph is a semantically labeled directed graph [7] represented as a quadruple $W = (N, E, S, T)$, whereby N is a set of nodes and $E \subseteq N \times N$

a set of edges. $T : N \cup E \to \Omega$ assigns a type from Ω to each node and each edge. Furthermore, $S : N \cup E \to \Sigma$ allocates a semantic description from a semantic metadata language Σ to each node and edge and annotates the graph structure with semantic knowledge. Each NEST graph is part of a case base, which is denoted as $W \in CB$. In Fig. 1, an exemplary NEST graph is shown that represents a simple cooking recipe with different node and edge types and an exemplary semantic description.

For the similarity computation between two such NEST graphs, suitable similarity measures for this case representation are required. Bergmann and Gil [7] have developed a similarity measure that computes this similarity based on the local-global principle [5, 10]. Here, the global similarity is composed of the local similarities computed between the individual graph elements, based on corresponding similarity measures [7]. The local similarity is always 0.0 if the types of the graph elements to be compared are not identical. If the types are equal, a similarity calculation is initiated, which calculates a similarity value based on the semantic descriptions. When computing the global similarity, several possible mappings of the query and case graphs to each other exist. To determine the suitability of a case, the highest possible global similarity value must be calculated, which is based on the mapping that leads to the highest local similarities. This mapping is found by conducting an A* search in the solution space, for which various heuristics are used. An optimized form of this search procedure is presented by Zeyen and Bergmann [44]. The described similarity assessment based on the local-global principle and the large search space of possible mappings of graph items make it difficult for a knowledge engineer to understand the composition of a similarity value on a conceptual level.

2.2 Distinction from Related Explanation Approaches

This contribution uses visualizations as a method to explain similarities in CBR applications. Thereby, the proposed approach differs from thematically related work in the area of XCBR. In some publications of this research field, CBR is utilized as a method to explain other artificial intelligence approaches, such as Kenny and Keane [21], Nugent and Cunningham [33], Gates et al. [15] or Recio-García et al. [36]. In other XCBR work, visualizations are used to explain similarity measures, as in Batyrshin et al. [4]. This work differs in that it is not the measure itself that is to be visualized, but the individual similarity value. Furthermore, visualizations for similarity distributions are used, for example, in case bases, as in Rostami et al. [38] or Namee and Delany [32]. Other works deal with visualizing retrieval results, such as Lamy et al. [23] or Paola and Bach [27]. These approaches visualize the similarity distribution in a case base or from a query to the best retrieval results, but they do not offer an explanation for the respective similarities, which are considered in this work.

2.3 Approaches for Visualizing Similarities

The following methods visualize similarities in CBR or between graphs and are, thus, identified as thematically relevant approaches. Bach and Mork [3] examine visualizations of similarities in CBR. The focus of their work is the visualization of similarity measures during the retrieval phase. In the underlying similarity computation, the local-global principle is applied, to which this approach is restricted. Massie et al. [28] address explanations in CBR by visualization. They aim to extend the explanation provided by the CBR system in the form of a single case with a similarity value to make the knowledge more accessible to users. For this purpose, they visualize the similarities of cases representing lists of numerical values in coordinate diagrams. The merge graph designed by Andrews et al. [2] stems from the field of Business Process Visualization. A merged graph is computed from two input graphs and the corresponding similarities between the graph nodes, allowing a visual comparison of the two graphs. For the merging procedure, it is required that the two graphs are pretty similar to have useful overlaps in the merged graph. The authors also mention that this approach is not infinitely scalable and that the computation of the graph layout results in a high computational complexity. Ivanov et al. [19] deal with an automatic comparison of business processes in the field of Business Process Management. They present a tool that can be used to detect and visualize discrepancies in two processes to check them for conformance. It can also help to find duplicates or to classify a new process. When mapping two processes onto each other, the similarity can be computed as a label match, a structural similarity or a graph edit distance. Also related is the general approach of heatmaps [42], which is a data visualization technique that uses colors and their intensity to visually indicate clusters in a two-dimensional table. These are used in CBR to show how the similarity values of one case relate to all other cases in the case base [30]. The use of such heatmaps can also be adapted to other domains. The approaches presented in this section are evaluated in Sect. 4.1 based on the elicited requirements.

3 Requirements for Visualization

To review the related approaches presented in Sect. 2.3 regarding their suitability to be used in POCBR, a list of requirements is needed. These requirements can also be used to derive necessary extensions for the existing approaches or to provide a guideline for the creation of a new visualization technique. Initially, established requirements are identified by a literature research and adapted for the application area of this contribution. Due to the small number of publications in this area, we derived further requirements from researchers, who are involved in this area, through a focus group interview [35]. We conducted this method of elicitation with a group of five people. First, they were confronted with the requirements that are collected based on the literature. Then, two visualization approaches from the literature according to Bach and Mork [3] and Andrews et al. [2] (see Sec. 2.3) were presented, whereupon the researchers had to justify why they found the approach suitable or not suitable in detail. In the interview, the participants discussed their answers, but overall agreed on the desired requirements.

Based on the literature review and the focus group interview, the following seven requirements are derived:

Req. 1 Explanation for Resulting Similarity: The visualization should comprehensibly justify the calculated similarity value and its composition. This is inspired by the goal of justification according to Sørmo et al. [41] as well as by the requirements for XCBR systems of Hall et al. [16].
Req. 2 Representation of Process Data: The visualization must support a case representation of processes, in the context of this contribution, as semantic graphs [7]. In addition to representing processes, the visualization approach must be able to show the similarities between them. So, the similarity measure used at the global level for process-oriented cases must be included in the visualization.
Req. 3 Representation of Similarities Between Graph Components: The visualization should include the local similarities by explaining their values and the measures used for their calculation. This is analogous to the requirement of Hall et al. [16] that explanations in XAI systems should be provided on the local level. If no local information contributes to the similarity computation, then no similarities need to be visualized there.
Req. 4 Visualization From the Query's Perspective: The explanation of the visualization should be from the query's perspective. For example, if a case contains an element that is not considered in the global similarity calculation, it is irrelevant knowledge in this context. This should be visually marked accordingly.
Req. 5 Representation of All Features Involved in the Similarity and of their Influence on the Similarity: The visualization should represent all features involved in the composition of the similarity. This is inspired by the prioritization of decision information by Hall et al. [16], according to which appropriately relevant features and feature relationships are selected for representation in the explanation. In the context of POCBR, this refers to the influence of these elements on the overall similarity, which can also be explored interactively. This requirement is specified in four sub-requirements, which were identified in the focus group interview.

Req. 5a Marking All Similarities With a Desired Value: The visualization should mark all similarities with a defined value, so these local similarities can be considered in detail. For example, the origin of similarities that values are very high like 1.0 or very low like 0.0 can be investigated.

Req. 5b Flexible Changes of Similarity Measures: To be able to estimate the influence of used similarity measures on the overall similarity value, the visualization should provide an interactive element that allows a flexible adjustment of these measures. This allows changes to the similarity model to be tracked. These adjustments should be possible on both, the global and the local level.

Req. 5c Filter Functions for Graph Elements: The visualization shall provide a function to hide selected types of graph items. This should make it possible to better assess the influence of certain graph item types on the overall similarity.

Req. 5d Pre-definition of Mappings: In the visualization, it should be possible to manually specify mappings based on which a new similarity calculation is performed. This is a functional requirement where the developer can manually check the influence of mapping pairs. It should only be possible to set valid mappings according to the definition of the semantic graphs [7].

Req. 6 Pairwise Visualization of Similarities: The similarity to be visualized always consists of a pair of query and case to ensure its traceability. For example, this would not occur, if the similarity of a query to an entire case base is visualized. The elements of the tuple are restricted to the query and case only.

Req. 7 Relation to Other Cases: The visualization should classify the similarity within the case base as well as in comparison with specific other cases. This requirement comes entirely from the focus group interview and is specified in two sub-requirements.

Req. 7a Classification Within Similarities of the Case Base: The similarity value should be ranked within the similarity distribution of the case base and visually represented. This allows to detect if all similarity values are close together or widely distributed, thus, identifying necessary adjustments to the similarity model for higher discriminatory power.

Req. 7b Comparison With Other Cases: The visualization shall include multiple pairwise visualizations to provide comparison to other similarity calculations. For this purpose, it should be possible to compare several visualizations in parallel. This should be used to identify possible adjustments in the similarity model that would not be noticeable when inspected individually.

4 Techniques for Similarity Visualization

Based on the presented requirements, the identified publications from the related work (see Sect. 2.3) are examined for their suitability, and new visualization forms are considered (see Sect. 4.1). The graph mapping approach is then presented (see Sect. 4.2), followed by the merge graph approach (see Sect. 4.3) and the heatmap approach (see Sect. 4.4)[1].

4.1 Suitable Visualization Approaches

Based on the presented requirements, the approaches introduced in Sect. 2.3 are reviewed for their suitability for use as visualizations for POCBR. The approaches of Bach and Mork [3] and Massie et al. [28] are identified as not suitable because they could not fulfill Req. 2 of representing the process data. Both approaches are constructed too restrictively for the respective form of case representation, and it is not possible to convert the cases into a suitable form of representation. For the visualization of local similarities demanded in Req. 3,

[1] A detailed description of all visualization approaches, including various mock-ups, is available at https://git.opendfki.de/easy/explanation-of-similarities-in-pocbr-by-visualization/-/blob/main/Detailed_Description_of_Visualization_Approaches.pdf

these approaches are suitable under the condition that an appropriate data structure is available there. The approach of Ivanov et al. [19] does not meet Req. 3 of representing the similarities between graph components, as it only supports similarity measures such as a graph edit distance that does not consider local similarities in its pure form. Since in this contribution only similarities based on a mapping of graphs to each other are considered, this requirement is not fulfilled. The merge graph approach according to Andrews et al. [2] and an extended heatmap approach, which represents the graph elements in the tables, can generally fulfill all requirements. Some requirements need further extensions to be fulfilled, like Req. 5b, Req. 5c and Req. 7. Thus, based on the requirements' analysis, the merge graph and heatmap approaches are identified as suitable for visualization in POCBR.

Due to the lack of suitable related work, attempts are made to design further visualization approaches from scratch. Thereby, one visualization is designed fundamentally new: the graph mapping approach. There are no comparable approaches to this visualization form in the literature. Some existing ones follow a similar basic idea, for example in psychology by Xuu [43] or for the visualization of similarity graphs by Rostami et al. [38]. But their adaptation for the application field in POCBR would be so extensive that it would also become a new approach.

These three visualization approaches are considered according to the requirements and appropriate adjustments are made to satisfy them. All three methods are designed according to Shneiderman's Visual Information-Seeking Mantra [40] which is a common method in the field of information visualization [11]. It specifies that a visualization should first give an overview of all the available information [40]. Then, it should be possible to zoom and filter, so that ultimately details are provided on demand. The visualizations are illustrated using various static mock-ups to present their design and indicate the functionalities. A simple comparison between two short recipes is used as an example.

4.2 Approach of Graph Mapping

A mock-up for the graph mapping approach can be found in Fig. 2, where the display of multiple visualization instances described in Req. 7b is omitted for a clearer overview. This is given by a configurable number of instances that are displayed below each other. In the graph mapping visualization, the similarities are represented as directed edges starting from the query. These contain the information about the local similarities and their composition. The thickness of the edges provides additional information about the similarity, which is a common method in information visualization techniques [14]. When displaying edges, a distinction is made according to whether they are mappings between nodes or edges. The mapping edges between nodes are solid, the edges between edges are dashed.

The visualization has a header that contains a menu with general information about the case and the similarity calculation, where the global similarity measure can also be adjusted, according to Req. 5b. In addition, there is the possibility to

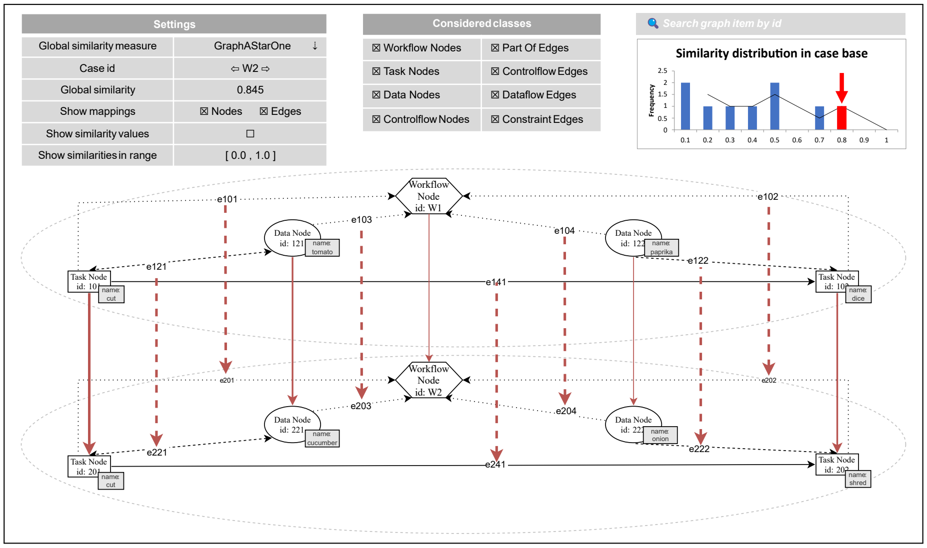

Fig. 2. Overview of the Graph Mapping Approach.

hide nodes or edges, to display similarity values on the edges, as well as to display only similarities whose values are within a given interval according to Req. 5a. A second menu contains the filter functions for the different graph element types according to Req. 5c. Next to it, a histogram illustrates the similarity distribution in the case base according to Req. 7a. Above this, a search function is provided that can be used to find elements from the query and case, for example based on their ID or a textual attribute expression.

In the graph mapping approach, the query graph as well as the case graph are each aligned according to the longest path algorithm [13]. The mapping edges always run between the elements mapped onto each other and are represented with sufficient spacing. The example in Fig. 2 is the optimal case, where all edges are parallel, which is not true for most examples. If 50% or more edges run in the opposite direction, the case graph is mirrored to provide a better overview. If the case graph contains unmapped elements, these are displayed transparent and in a light shade of gray. They are irrelevant for the similarity calculation and, thus, Req. 4 is also considered. As defined in Req. 5d, mappings can be set by drawing a line from an element of the query to an element of the case, whereby only valid mappings are possible. Afterward, a new similarity calculation is started with this mandatory mapping, whereupon all similarity values in the visualization change. After such a change, the function for annotating the similarity values, which are displayed at the edges, is also selected there. Selecting the edge opens a menu for viewing the local similarity values. There, the locally used similarity measure can be viewed and replaced by a compatible measure according to Req. 5b. Besides, information about the elements from the query and case is included, as well as the local similarity values. By clicking a button, a visualization for the local similarity can be opened, where it can be traced and adjusted according to Req. 3.

4.3 Approach of Merge Graph

The basic idea of the merge graph is to merge two graphs to be compared into one graph, thereby showing the similarities between them. For the application in POCBR, similarities between graphs are displayed independent of the height of their global similarity value. Moreover, mapping between edges must also be visualized and not only between nodes.

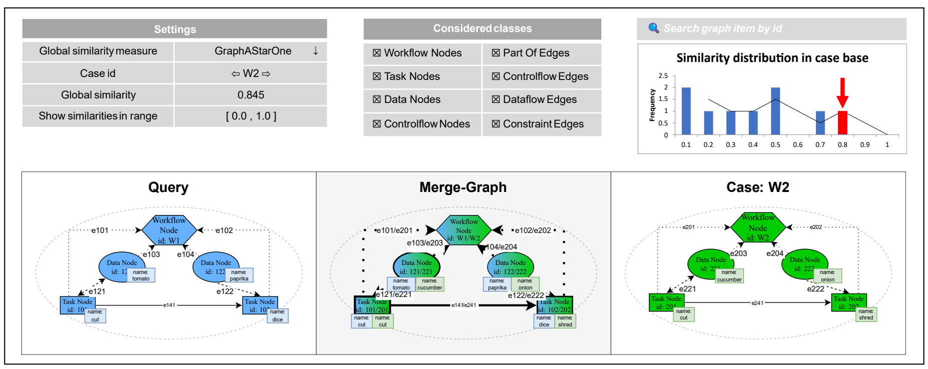

Fig. 3. Overview of the Merge Graph Approach.

Fig. 3 shows a mock-up of the adapted merge graph, where Req. 7b is integrated as in the graph mapping and omitted here. The header is identical to that of the graph mapping approach, except that the entries for displaying specific mappings and annotating the similarity values are omitted. The query and case graphs are again based on the longest common path algorithm [13], and both are color-coded. The merge graph always adopts the layout of the query graph to ensure comparability with other similarity calculations, according to Req. 4 and Req. 7b. The nodes in the merge graph contain a gradient of both colors if they are mapping nodes. Unmapped elements of the query are taken over unchanged, while unmapped elements of the case are weakly transparent. The mapped nodes can always be mapped to the nodes of the query in the merge graph. This is not always the case for edges. If a mapped edge runs between two nodes to which their corresponding nodes have also been mapped, then this mapping edge is displayed in the merge graph. However, if this is not the case, the edge is hidden from the case and can only be identified by looking at the local similarities. The representation of similarities is done analogously to the graph mapping approach, using different thicknesses of the nodes' frames as well as of the edges. Selecting an element of the merge graph opens the menu for displaying local similarities, which is analogous to graph mapping. To set mappings as required by Req. 5d, a connection is drawn from an element in the query to an element in the case graph. With this preset, the other mappings are recalculated and visualized, whereby the old mapping cannot be marked.

4.4 Approach of Heatmaps

Instead of using heatmaps to visualize similarity distributions in the case base as in the literature [30, 42], heatmaps are used in this contribution to compare elements of two different cases. Each entry in the heatmap thereby symbolizes a mapping of an element of the query to an element of the case. An analogous idea exists in the original merge graph approach to represent metadata [2]. In addition to the color representation, where a bright red represents high similarity and a dark red low similarity, the numerical similarity values are also included in the tables. The heatmap requires the calculation of the individual similarity values if they have not already been determined and cached as part of the original similarity calculation.

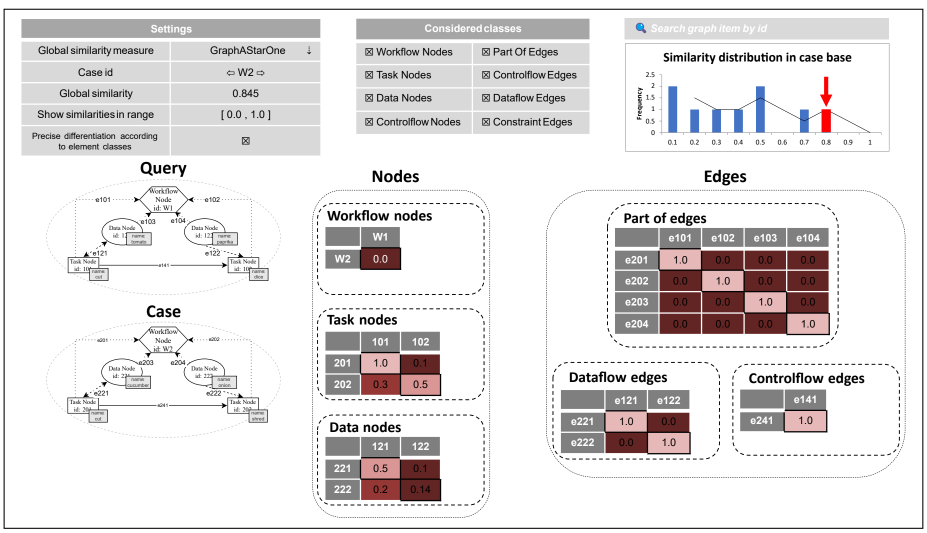

Fig. 4. Overview of the Heatmap Approach.

Fig. 4 shows a mock-up of such an extended merge graph, where Req. 7b is considered as in the other two approaches and omitted here. Basically, two heatmaps are shown, one for the nodes and one for the edges. They contain several non-selectable fields, since no mapping is allowed at these points. These can be split as shown in the mock-up into several individual tables for the different graph element types, omitting the invalid fields by a function provided in the header. This header is similar to the one in the graph mapping approach, except that the entries for displaying specific mappings and annotating the similarity values are not contained. Selecting a field opens the menu for displaying local similarities, which is analogous to graph mapping and merge graph. The fields that represent the current mappings are outlined and thus made recognizable. In addition to the heatmaps, the query and the case graphs are shown, again using the longest path algorithm [13]. Both are displayed independently and serve to clarify the graph elements in the heatmaps. When a field is selected, the elements in the graphs are marked accordingly. Mappings according to Req. 5d can be set by selecting other fields. The originally selected mappings are marked by dashed borders.

5 Experimental Evaluation

The evaluation examines the hypothesis that the designed visualizations of similarities enable domain experts and knowledge engineers to acquire a more profound understanding of the similarity model and make targeted adjustments to it. First, the setup of the evaluation is presented (see Sect. 5.1). It is followed by a presentation of the results and their discussion (see Sect. 5.2).

5.1 Experimental Setup

The evaluation of the developed approaches was performed based on the presented requirements (see Sect. 3). A total of ten researchers in POCBR were recruited, who have been involved in this area for several years, so they are familiar with the subject, and have built similarity models themselves in the role of knowledge engineers. They were presented with the requirements and the three developed approaches, which were shown by mock-ups. With this background, they rated the three visualization approaches using five-point, ordinal, and interval-scaled Likert scales [25] and school grades. The participants were asked to select which approach best met the respective requirements compared with the other approaches, and to select their overall favorite approach.

5.2 Experimental Results and Discussion

In the individual evaluation for Req. 1 of the explanation of similarity, all three approaches are rated positively, with graph mapping and heatmaps both scoring best. To compare the individual evaluations, the Likert items are evaluated in the interval $[0, 4]$ and the mean value per requirement is calculated. This mean value based on each requirement is shown in Fig. 5. Overall, the average value indicates the graph mapping approach as the most suitable, with an absolute score of 3.12. The other two approaches are not rated significantly lower in total, with heatmap second scoring 2.76 and merge graph third scoring 2.75. The grading of the approaches also suggests this order when looking at the mean, with merge graph and heatmap again having a minimal difference. Looking at the median, confirms graph mapping in first place, with the other two approaches tied.

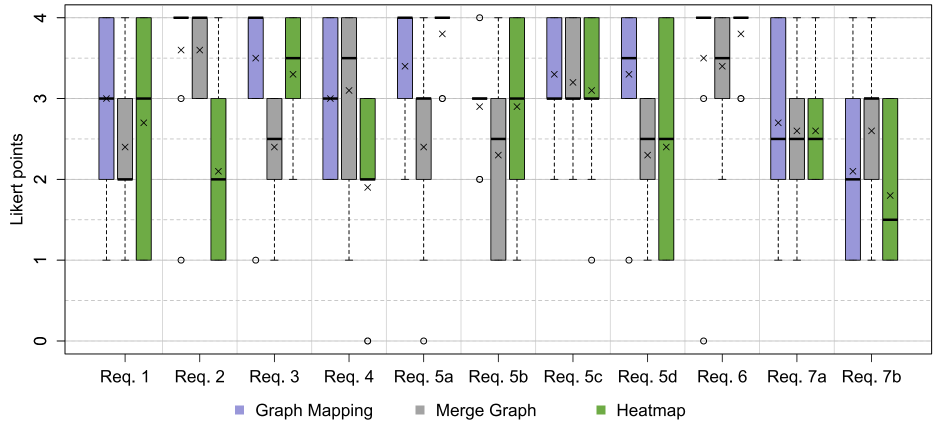

Fig. 5. Scored Likert Values for the Visualization Approaches Visualized as Boxplots.

In the selection of the favored approach for each requirement, graph mapping clearly dominates. Overall, the graph mapping approach is rated best for five requirements and tied with others for one requirement. In the other five requirements, participants could not decide on a favorite. When asked for the overall best approach, seven experts rated the graph mapping approach best. In this question, the margin is the clearest, with merge graph second and heatmap third.

The working hypothesis of this contribution is to achieve better explainability of similarities in POCBR for knowledge engineers by using visualization techniques. Based on the experts' assessment of requirements fulfillment, the evaluation suggests that this goal is achieved for all approaches created. Both, based on the Req. 1 of explainability and generally recognizable from the requirement evaluation, the approach of graph mapping stands out as the best approach to increase explainability. Based on the evaluation results, the graph mapping approach has been prototypically implemented in the ProCAKE framework [8] to be reused by the research community[2].

6 Conclusion and Future Work

In this paper, we present novel visualization approaches for complex case representations, enabling a better explainability of similarity for knowledge engineers in POCBR. For this purpose, first, we conduct a literature study and, second, we build a catalog of requirements for visualization approaches . Based on these requirements, the merge graph approach [2] and a heatmap approach are further investigated as they promise good support for developers. However, the approaches need to be tailored and extended for the application in POCBR. In

[2] The implementation is available at https://git.opendfki.de/easy/explanation-of-similarities-in-pocbr-by-visualization

addition, a new visualization approach based on graph mapping is introduced. Thus, three visualizations are provided to increase the explanation in POCBR. To verify their successful conceptualization, the evaluation is conducted, explicitly addressing explanation. The explanation capabilities of all approaches are rated high, with graph mapping receiving the best scores and being the preferred approach of the experts. Overall, the evaluation suggests a contribution to explanation for all approaches based on the requirements. Thus, the evaluation indicates that the proposed visualizations are suitable for supporting CBR developers in their task and to increase the explanation of similarities in POCBR.

In future work, further case representations should be investigated for using visualizations to support CBR developers in their work. As recent literature shows [3, 28], only rather simple case representations are currently investigated and many case representations have not even been considered, such as textual cases. Visualizations can also be created for other case representations, such as the simple ones, which have not been covered in the literature so far, or more complex structures. On a global level, an extension to represent dependencies between processes [22] or an adaptation for usage with other similarity measures are conceivable. Another research possibility is the application to other domains, such as argument graphs [6], in which a similarity visualization can be used to explain argumentation processes [24]. To support the knowledge engineer in the pre-definitions of mappings as stated in the requirements , a procedure must also be designed to search for a configuration of the similarity model that actually achieves the specified mapping during the similarity calculation. Currently, this is only possible by a manual configuration. For this purpose, methods from the field of intelligent search [17] can be applied, for example. In this work, we focus on supporting CBR system developers for their work. In future work, we want to examine what changes need to be performed on the visualization approaches to support end users. In this context, other aspects of the similarity calculation are probably more important, e. g., the differences between multiple query-case pairs and their similarities.

Acknowledgments. This work is funded by the Federal Ministry for Economic Affairs and Climate Action under grant No. 01MD22002C *EASY*.

References

1. Aamodt, A., Plaza, E.: Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. AI Commun. **7**(1), 39–59 (1994)
2. Andrews, K., Wohlfahrt, M., Wurzinger, G.: Visual Graph Comparison. In: 13th IV. pp. 62–67. IEEE (2009)
3. Bach, K., Mork, P.J.: On the Explanation of Similarity for Developing and Deploying CBR Systems. In: 33rd FLAIRS. pp. 413–416. AAAI Press (2020)
4. Batyrshin, I.Z., Kubysheva, N., Solovyev, V., Villa-Vargas, L.A.: Visualization of Similarity Measures for Binary Data and 2x2 Tables. CyS **20**(3), 345–353 (2016)
5. Bergmann, R.: Experience Management: Foundations, Development Methodology, and Internet-Based Applications, LNCS, vol. 2432. Springer (2003)

6. Bergmann, R., Biertz, M., Dumani, L., Lenz, M., Ludwig, A., Neumann, P.J., Ollinger, S., Sahitaj, P., Schenkel, R., Witry, A.: The ReCAP Project. Datenbank-Spektrum **20**(2), 93–98 (2020)
7. Bergmann, R., Gil, Y.: Similarity Assessment and Efficient Retrieval of Semantic Workflows. Inf. Syst. **40**, 115–127 (2014)
8. Bergmann, R., Grumbach, L., Malburg, L., Zeyen, C.: ProCAKE: A Process-Oriented Case-Based Reasoning Framework. In: 27th ICCBR Workshop Proc. (2019)
9. Bunke, H., Messmer, B.T.: Similarity Measures for Structured Representations. In: Topics in Case-Based Reasoning: 1st European Workshop. pp. 106–118. Springer (1994)
10. Burkhard, H., Richter, M.M.: On the Notion of Similarity in Case Based Reasoning and Fuzzy Theory. In: Soft Computing in CBR, pp. 29–45. Springer (2001)
11. Card, S.K., Mackinlay, J.D., Shneiderman, B.: Readings in Information Visualization: Using Vision to Think. Academic Press (1999)
12. Das, A., Rad, P.: Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey. CRR **2006.11371** (2020)
13. Eades, P., Xuemin, L.: How to Draw a Directed Graph. In: IEEE Workshop on Visual Languages. pp. 13–14. IEEE Lett. Comput. Soc. (1989)
14. Epskamp, S., Cramer, A.O., Waldorp, L.J., Schmittmann, V.D., Borsboom, D.: qgraph: Network Visualizations of Relationships in Psychometric Data. J. Stat. Softw. **48**, 1–18 (2012)
15. Gates, L., Kisby, C., Leake, D.: CBR Confidence as a Basis for Confidence in Black Box Systems. In: 27th ICCBR. LNCS, vol. 11680, pp. 95–109. Springer (2019)
16. Hall, M., Harborne, D., Tomsett, R., Galetic, V., Quintana-Amate, S., Nottle, A., Preece, A.: A Systematic Method to Understand Requirements for Explainable AI (XAI) Systems. In: 28th IJCAI Workshop Proc. vol. 11 (2019)
17. Hart, P.E., Nilsson, N.J., Raphael, B.: A Formal Basis for the Heuristic Determination of Minimum Cost Paths. IEEE Trans. Syst. Sci. and Cybern. **4**(2), 100–107 (1968)
18. Hoffmann, M., Bergmann, R.: Informed Machine Learning for Improved Similarity Assessment in Process-Oriented Case-Based Reasoning. CoRR **abs/2106.15931** (2021)
19. Ivanov, S., Kalenkova, A.A., van der Aalst, W.M.P.: BPMNDiffViz: A tool for BPMN Models Comparison. In: 13th BPM. CEUR Workshop Proc., vol. 1418, pp. 35–39. CEUR-WS.org (2015)
20. Kendall-Morwick, J., Leake, D.: A Study of Two-Phase Retrieval for Process-Oriented Case-Based Reasoning. In: Succ. CBR Appl.-2, pp. 7–27. Springer (2014)
21. Kenny, E.M., Keane, M.T.: Twin-Systems to Explain Artificial Neural Networks Using Case-Based Reasoning: Comparative Tests of Feature-Weighting Methods in ANN-CBR Twins for XAI. In: 28th IJCAI. pp. 2708–2715 (2019)
22. Kumar, R., Schultheis, A., Malburg, L., Hoffmann, M., Bergmann, R.: Considering Inter-Case Dependencies During Similarity-Based Retrieval in Process-Oriented Case-Based Reasoning. In: 35th FLAIRS. FloridaOJ (2022)
23. Lamy, J.B., Sekar, B., Guezennec, G., Bouaud, J., Séroussi, B.: Explainable Artificial Intelligence for Breast Cancer: A Visual Case-Based Reasoning Approach. Artif. Intell. Med. **94**, 42–53 (2019)
24. Lenz, M., Sahitaj, P., Kallenberg, S., Coors, C., Dumani, L., Schenkel, R., Bergmann, R.: Towards an Argument Mining Pipeline Transforming Texts to Argument Graphs. In: 8th COMMA. FAIA, vol. 326, pp. 263–270. IOS Press (2020)

25. Likert, R.: A Technique for the Measurement of Attitudes. Arch. Psychol. (1932)
26. Manovich, L.: What is Visualization. paj: The Journal of the Initiative for Digital Humanities, Media, and Culture **2**(1) (2010)
27. Marín-Veites, P., Bach, K.: Explaining CBR Systems Through Retrieval and Similarity Measure Visualizations: A Case Study. In: 30th ICCBR. LNCS, vol. 13405, pp. 111–124. Springer (2022)
28. Massie, S., Craw, S., Wiratunga, N.: Visualisation of Case-Base Reasoning for Explanation. In: 7th ECCBR Proc. pp. 135–144 (2004)
29. Mathisen, B.M., Aamodt, A., Bach, K., Langseth, H.: Learning Similarity Measures from Data. Prog. Artif. Intell. **9**(2), 129–143 (2020)
30. McArdle, G., Wilson, D.C.: Visualising Case-Base Usage. In: 5th ICCBR Workshop Proc. pp. 105–114 (2003)
31. Minor, M., Montani, S., Recio-García, J.A.: Process-Oriented Case-Based Reasoning. Inf. Syst. **40**, 103–105 (2014)
32. Namee, B.M., Delany, S.J.: CBTV: Visualising Case Bases for Similarity Measure Design and Selection. In: 18th ICCBR. LNCS, vol. 6176, pp. 213–227. Springer (2010)
33. Nugent, C., Cunningham, P.: A Case-Based Explanation System for Black-Box Systems. Artif. Intell. Rev. **24**(2), 163–178 (2005)
34. Ontañón, S.: An Overview of Distance and Similarity Functions for Structured Data. Artif. Intell. Rev. **53**(7), 5309–5351 (2020)
35. Rabiee, F.: Focus-Group Interview and Data Analysis. PNS **63**(4), 655–660 (2004)
36. Recio-García, J.A., Parejas-Llanovarced, H., Orozco-del-Castillo, M.G., Brito-Borges, E.E.: A Case-Based Approach for the Selection of Explanation Algorithms in Image Classification. In: 29th ICCBR. LNCS, vol. 12877, pp. 186–200. Springer (2021)
37. Richter, M.M.: Knowledge Containers. In: Readings in CBR. MKP (2003)
38. Rostami, M.A., Saeedi, A., Peukert, E., Rahm, E.: Interactive Visualization of Large Similarity Graphs and Entity Resolution Clusters. In: 21th EDBT. pp. 690–693. OpenProceedings.org (2018)
39. Schoenborn, J.M., Weber, R.O., Aha, D.W., Cassens, J., Althoff, K.D.: Explainable Case-Based Reasoning: A Survey. In: AAAI-21 Workshop Proceedings (2021)
40. Shneiderman, B.: The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In: Proc. of IEEE Symposium on Visual Languages, pp. 336–343. IEEE Computer Society (1996)
41. Sørmo, F., Cassens, J., Aamodt, A.: Explanation in Case-Based Reasoning-Perspectives and Goals. Artif. Intell. Rev. **24**(2), 109–143 (2005)
42. Wilkinson, L., Friendly, M.: The History of the Cluster Heat Map. Am. Stat. **63**(2), 179–184 (2009)
43. Xuu, A.B.: Structure Mapping in the Comparison Process. AJP **113**(4), 501–538 (2000)
44. Zeyen, C., Bergmann, R.: A*-Based Similarity Assessment of Semantic Graphs. In: 28th ICCBR. LNCS, vol. 12311, pp. 17–32. Springer (2020)