Modeling Linguistic Facets of Multimedia Content for Semantic Annotation

Massimo Romanelli,[1] Paul Buitelaar[2] and Michael Sintek[3]

[1] DFKI IUI, Saarbrücken, Germany, firstname.lastname@dfki.de
[2] DFKI LT, Saarbrücken, Germany, firstname.lastname@dfki.de
[3] DFKI KM, Kaiserslautern, Germany, firstname.lastname@dfki.de

Abstract. We provide an integrated ontological framework offering coverage for deep semantic content, including ontological representation of multimedia based on the MPEG-7 standard. We link the deep semantic level with the media-specific semantic level to operationalize multimedia information. Through the link between multimedia representation and the semantics of specific domains we approach the Semantic Gap. The focus of the paper is on the linguistic features of multimedia, the annotation of these features and their analysis.

1 Introduction

An important reason for the so-called 'semantic gap' (the difficulty in assigning high-level semantics to the results of low-level feature analysis) is the lack of alignment between different levels of semantics and levels of analysis for the different modalities. It is therefore important to develop an integrated model that aligns the foundational semantics level with the domain-specific semantics level, the semantics of the different modalities and the semantics of multimedia analysis. Additionally, in order to generalize this for all domains, the alignment of domain-specific and multimedia semantics should be organized on the foundational level.

In this paper we describe such an integrated model (working title 'SmartMediaLing') that we developed in the context of the SmartWeb project[4] on mobile access to the Semantic Web [22]. In SmartWeb we were confronted with a number of different semantic analysis tasks (media annotation and presentation, multi-modal interaction, text analysis, etc.), each of which requiring a different level of representation, realized by a number of separate ontologies. In order to bring these different representation levels together we developed the SmartMediaLing integrated model as a common knowledge space that provides semantic interoperability between the different components of the SmartWeb system.

The SmartMediaLing approach described here uses the DOLCE foundational ontology for this purpose as it already provides patterns for defining so-called 'information objects', on top of which we were able to define the alignment of the

[4] http://www.smartweb-project.org

different semantic levels mentioned above. In particular, we used the DOLCE D&S (Description and Situation) and OIO (Ontology of Information Objects) patterns to align the SmartMedia ontology for defining multimedia objects and the LingInfo ontology for defining linguistic (textual) objects with the DOLCE foundational model.

The paper is organized as follows: in Sec. 2 we describe the different levels of semantic representation that we consider. In Sec. 3 we discuss the constituent ontologies (DOLCE, SmartMedia, LingInfo) that are integrated into SmartMediaLing. In Sec. 4 we discuss the alignment strategy and present the SmartMediaLing ontology in more detail. In Sec. 5 we discuss the relation to other approaches.

2 Narrowing Down the Task: Different Semantic Levels

In order to reach the goal of appropriately represent and processing different information deriving from different analysis perspectives of the same object a complex approach to representation of contents becomes indispensable. Different perspectives on a complex object corresponds to different representation levels specifying features, properties and relationships on the different analysis points of view. In the definition of our task to proper represent semantics of multimedia we evidence basically four different level of representation that has to interact: foundational, domain specific, multimedia, linguistic.

Additional evidence for the definition of different representation levels comes from the semiotic investigation of communication.

Semiotics is "the study of the social production of meaning through signs" [19]. As a Kantian philosopher Peirce, key figure in the early development of semiotics, distinguishes between the "word" and the "sign" [17]. As defined in Peirce semiotic theory, communication takes place between three subjects: a *sign* (also called *representamen*), that denotes an object, an *object* from the world, to which this sign refers and the *interpretant*, the sense made of that sign. Peirce further distinguishes three types of sign depending on the type of relation existing between sign and object: *symbol*, based on a conventional relation (e.g. spoken language, language of gesture), *icon*, based on a similarity relation (e.g. a portrait), and *index* a contextual relation (e.g. smoke indicating the presence of fire). We identify the interpretant as being the concept in an ontological system, the symbol as depicting the linguistic level of representation, the icon as depicting the multimedia level, and the index as depicting the discourse level (see Fig. 1).

The Foundational Level The definition of a complex semantic framework with different levels of representation needs the specification of a conceptual relational common ground offering appropriate instruments for linking together these levels. As soon as complexity increases, the usability reasons suggest applying modularization and distribution of knowledge in an interoperating framework. To accomplish this task successfully a useful

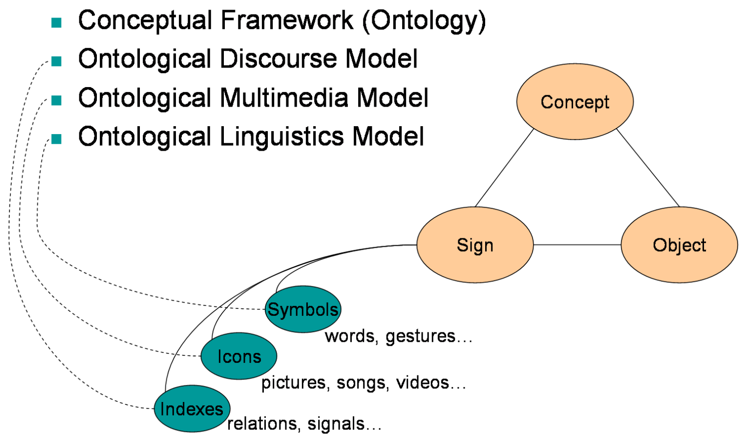

Fig. 1. Peirce semiotic triangle readapted and in context to different ontological levels.

approach is to define a foundational level of representation from which every module of the framework can access basic ontological categories and relations. Foundational ontologies define these top level for the modularization and integration of meaning coming from different analysis sources.

The Domain-Specific Level An ontology is said to be domain specific if it models the semantic of a specific domain. In our framework we define a different ontology for each domain and then align the all ontologies to the foundational one. Working this way we have the possibility to each time expand the world knowledge covered by the ontology by means of just adding new ontology branches without modifying basic relations in the framework.

The Multimedia Level Videos, songs, pictures and so on are information objects with specific properties defining their realization in time (e.g. duration of a video) and space (e.g. number of pixels). On the other hand multimedia objects carry meaning that cannot be identified with the information object itself (e.g. an image depicting the Brazilian football team cannot be identified with the football team itself). We distinguish between a multimedia meta-data representation level, modeling characteristics of multimedia, and a domain specific level where concepts referred from media are completely specified.

The Linguistic Level In order to ensure annotation for multilingual knowledge a rich representation of the linguistic symbols for the object classes that are defined by an ontology is needed. The linguistic level of information correspond to the symbol in Peirce triangle as depicted in Fig. 1. The purpose of such a semantic level is the definition of a grounding to the human cognitive and linguistic domain. Such domain is also important in the context of the interaction with multimedia objects where texts appear also in the perspective of a media object or as part of other media (e.g., the caption of a picture, the subtitles of a video).

The Analysis Level Parallel to the already mentioned levels, that we can define as "static", we regard the "dynamic" dimension of multimedia as being the analysis level. We consider analysis, in both decomposition and annotation cases, as being a process activated by an agent, allayed to a multimedia object (domain) and resulting in his decomposition.

3 The Constituent Modules

Following the approach in [15] in order to define a framework with the features specified in Sec. 2 we have to first select a foundational ontology that matches the described requirements and enables us to reuse existing components. We decided to adopt the DOLCE ontology providing together a well defined formalization for basic relations and a number of modules, among others for the definition of contexts (D&S) and knowledge content (OIO). The second step is the specification of an adequate multimedia domain capable of describing annotation and decomposition of multimedia and a straight forwarded ontology description of linguistic feature.

In this section we shortly present the DOLCE ontology with the two modules *Descriptions & Situations* and *Ontology of Information Objects*. We then describe in two dedicated subsections the *SmartMedia* ontology for the coverage of the surface representation of the multimedia level. Finally we present the *LingInfo* ontology modeling facets of the linguistic domain.

3.1 DOLCE, D&S, OIO

DOLCE belongs to the WonderWeb library of foundational ontologies [12]. It is intended to act as a starting point for comparing and elucidating the relationships and assumptions underlying existing ontologies of the WonderWeb library. DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering) [7] is based on the fundamental distinction between enduring and perduring entities. An endurant is an entity that is wholly present, i.e., whose parts are all present, at any time at which it exists. A perdurant is an entity that enfolds in time, i.e., for any time at which it exists, some of its parts are not present. Meaning that *participation* is the main relation between *Endurants* (i.e., objects or substances) and *Perdurants* (i.e., events or processes): an *Endurant* exists in time by participating in a *Perdurant*. For example, a natural person, which is an *Endurant*, participates in his or her life, which is a *Perdurant*. DOLCE introduces *Qualities* as another category that can be seen as the basic entities we can perceive or measure: shapes, colors, sizes, sounds, smells, as well as weights, lengths or electrical charges. Spatial locations (i.e., a special kind of physical quality) and temporal qualities encode the spatio-temporal attributes of objects or events. Finally, *Abstracts* do not have spatial or temporal qualities and they are not qualities themselves. An example are *Regions* used to encode the measurement of qualities as conventionalized in some metric or conceptual space.

In DOLCE the module *Descriptions&Situations* (D&S) [8] has been defined to standardize a variety of reified contexts and states of affairs.

The DOLCE module **OIO** (**O**ntology of **I**nformation **O**bjects) provides a design pattern that allows us to concisely model the relationship between entities in an information system and the real world. As emphasized in [9] INFORMATION OBJECTS can be seen as NON-PHYSICAL-ENDURANTS participating in computational activities. Information-Objects correspond to the spatio-temporal entities of abstract information formalizing Shannon's communication theory [20].

3.2 SmartMedia

MPEG-7[5] is conceived for describing multimedia content data. MPEG-7 is used to store meta-data about multimedia in order to tag particular events. In the context of the SmartWeb project we defined an MPEG-7 based ontology (Smartmedia) following the approach in [2] and [11] restricting the number of the modeled concepts to those that fit well to the project.

Primarily the concepts mpeg7:MediaFormat for format and the coding parameters, mpeg7:MediaPro for coding schemes like resolution, compression, and mpeg7:SegmentDecomposition for decompositions of the audio, visual, textual segments in space, time, and frequency are imported into Smartmedia in order to offer a well defined background for the specification of meta-data level describing multimedia events like synchronization or decomposition of media.

3.3 Linginfo

Automatic multilingual knowledge markup requires a rich representation of the features of linguistic expressions (such as terms, synonyms and multilingual variants) for ontology classes and properties. Currently, such information is mostly missing or represented in impoverished ways, leaving the semantic information in an ontology without a grounding to the human cognitive and linguistic domain. Linguistic information for terms that express ontology classes and/or properties consists of lexical and context features, such as:

- *language-ID* - ISO-based unique identifier for the language of each term
- *part-of-speech* - representation of the part of speech of the head of the term
- *morpho-syntactic decomposition* - representation of the morphological and syntactic structure (segments, head, modifiers) of a term
- *statistical and/or grammatical context model* - representation of the linguistic context of a term in the form of N-grams, grammar rules or otherwise

[5] 7http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.html

To allow for a direct connection of this linguistic information for terms with corresponding classes and properties in the domain ontology, [4] developed a lexicon model (LingInfo) that enables a linguistically motivated definition of terms for each class or property. The LingInfo model [3] is represented by use of the meta-class `ClassWithLingInfo` (and meta-property `PropertyWithLingInfo`), which allow for the representation of LingInfo instances with each class/property, where each LingInfo instance represents the linguistic features (`feat:lingInfo`) of a term for that particular class.

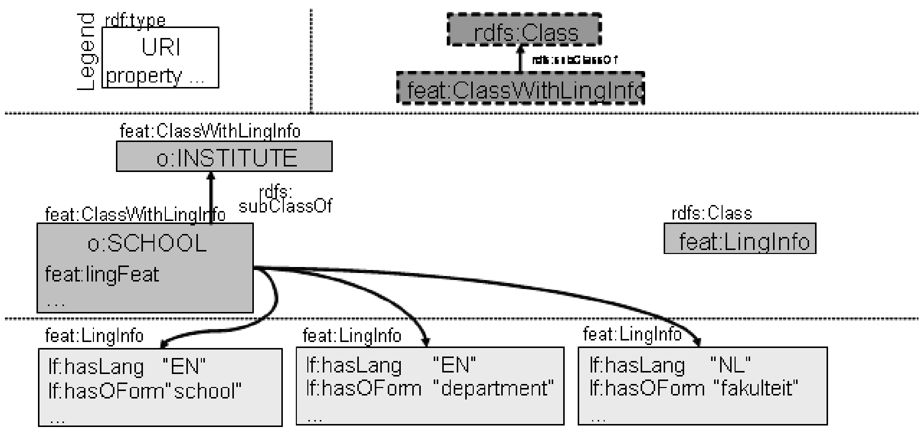

Fig. 2. LingInfo model with example domain ontology classes and LingInfo instances (simplified).

Figure 2 shows an overview of the model with example domain ontology classes and associated LingInfo instances. Figure 3 shows a sample application of the model with a LingInfo instance (and connected 'stem' instances) that represents the decomposition of the Dutch term "fakulteitsgebouw" ("department building"). The example shows a LingInfo instance (`Term-1` with `semantics "SCHOOL"`) that represents the word form "fakulteitsgebouw" (instance WordForm-1), which can be decomposed into "fakulteit" (`Term-2` , "fakulteit" with `semantics "SCHOOL"`) and "gebouw" (`Term-3` with `semantics "BUILDING"`).

4 Bringing It All together

In Sec. 2 and 3 we presented respectively the different levels of representation and how we ontologically cover such levels in order to proper processing multimedia information. In this section we show how we connected these different levels.

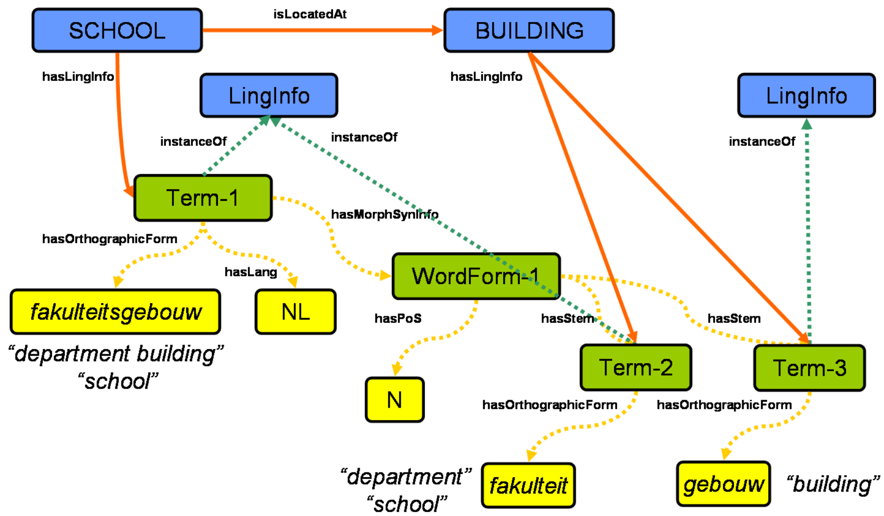

Fig. 3. LingInfo instance (partial) for the morpho-syntactic decomposition of the Dutch term "fakulteitsgebouw" ("department building")

This work were developed in the context of the SmartWeb project where the three ontologies introduced in Sec. 3 were all adopted as part of a comprehensive ontology named SWIntO (SmartWeb Integrated Ontology)[14]. The DOLCE ontology and the modules OIO and D&S were modified to meet the needs of the project and evolved respectively to SmartDOLCE, SmartOIO and SmartD&S. Basic functionality of Dolce remained unaffected. For more details on the use of DOLCE in SWIntO see also [6].

In the context of the SmartWeb project we successfully used this framework for the disambiguation of cross-modal reference expressions and resolution of multi-modal expressions. This work enabled the system the use of multimedia in a multimodal context like in the case of mixed gesture and speech interpretation, where every object that is visible on the screen must have a comprehensive ontological representation in order to be identified and processed [18][21].

4.1 Alignment strategy

The alignment process of a domain ontology to a core ontology is directly dependent from several factors [10]: intended use of the aligned ontology, intended form of the framework (modular, distributed), etc.

In our case we played a particular attention at following parameters:

- A modular reuse of the different component ontologies in other projects.
- Ontology alignment is different from equivalence because any element in the alignment depend on other elements and there will be degree of confidence between aligned elements.

In order to reach these means we decided to align ontologies to DOLCE as follows:

- non destructive: alignment happens without modifications for the core ontology.
- non reusing: properties are completely defined in the domain ontologies and then aligned as sub-properties to properties of DOLCE. No properties of DOLCE are directly reused in the ontology.

4.2 The Integrated Model: 'SmartmediaLing'

In [9] information objects are introduced on the base of an example (Dante's Comedy). To align the smartmedia ontology to the DOLCE we applied this example to the world of multimedia. In Fig. 4 are depicted basic relations and concepts modeled in the OIO framework and adopted for the alignment.

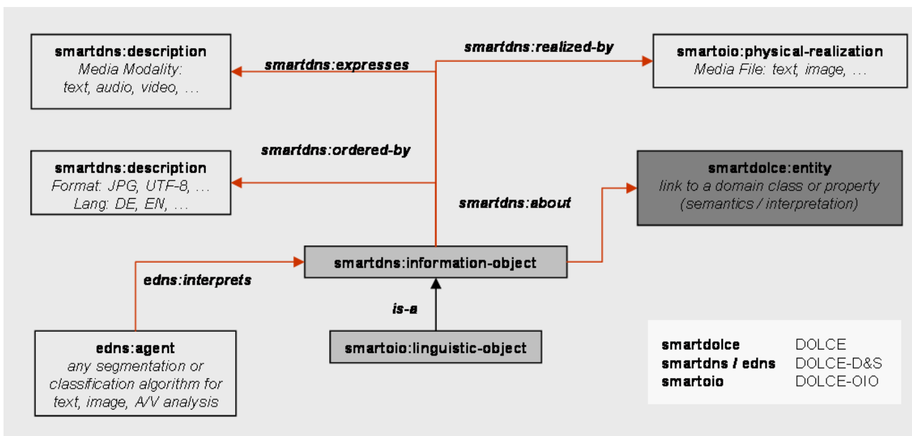

Fig. 4. Basic concepts and patterns of the DOLCE OIO module used in SWIntO for the integration of the foundational, multimedia and linguistic levels.

In the case of the analysis of a picture of the Brazilian football team we identify the picture itself as an *information object* that is *about* an *entity*, a *particular* modeling the Brazilian football team in a DOLCE aligned domain specific ontology. The picture can be decomposed to different *segments*, each segment being about a different player from the same domain specific ontology. A *Segment-Decomposition* is an information object carrying the result of a segmentation process, a *perdurant* applied by some *agent*, some classification or segmentation algorithm that *interprets* the information object *using* visual descriptors. In Fig. 5 we give a graphical representation of such relations.

Exactly the same way we can see e.g., a semantic parser as an agent participating in a parsing process that is identified by an information object of type

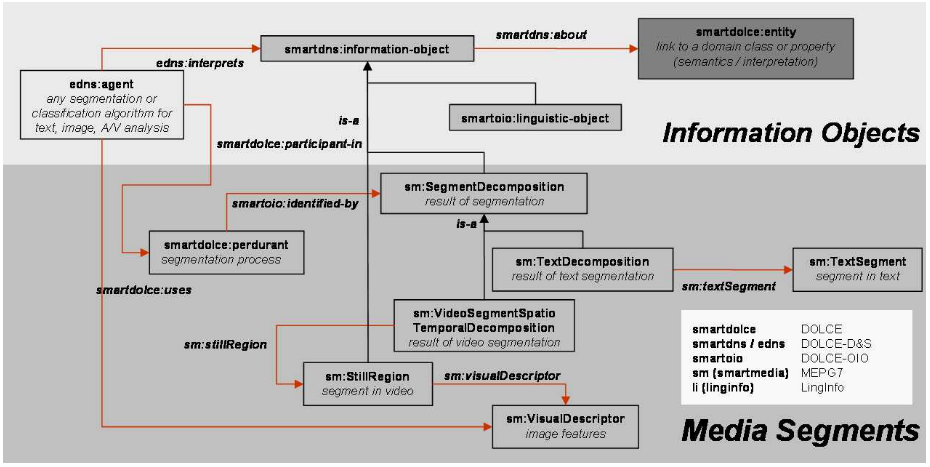

Fig. 5. The Smartmedia ontology integrated in the OIO module.

textDecomposition. The result of such a decomposition are again linguistic entities like sentences, words, morphemes and so on, as represented in the LingInfo ontology (See figure 6). Each linguistic information object is *about* an entity from a domain specific ontology and is itself a sort of *TextSegment* as specified from the Smartmedia ontology.

5 Related Work

A very interesting approach is that followed in [1] where the authors concentrate on the task of creating an ontological framework in the context of annotation of multimedia objects. The approach is based on the DOLCE ontology and makes deep use of the *Descriptions and Situations* DOLCE module defining annotation, description and semantic patterns usable for the realization of an annotation tool. This work offers a well defined specification of interpretation and annotation processes. However deep analysis of relations between surface representation objects and deep semantic objects is not taken into account.

An other approach to semantic annotation for multimedia content similar to the one adopted in this work can be found in [16]. The work is based on the same ontological background and place emphasis mostly on the visual part of the ontology, the context analysis used for the visual analysis and a visual analysis algorithm.

6 Conclusion

In this paper we described an approach to the specification of a semantics for the annotation and use of multimedia objects in a comprehensive ontological framework. We analyzed the characteristics of multimedia objects and evidenced the

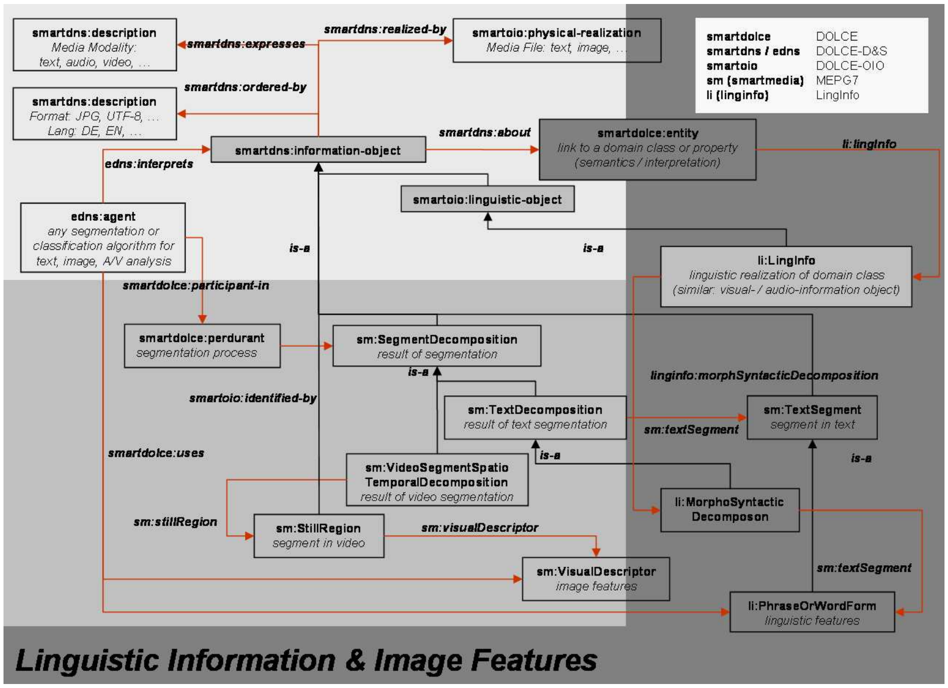

Fig. 6. The integrated framework for the representation of linguistic and multimedia objects as information objects.

necessity of specifying different levels of representation for covering the complexity of the task. On the other hand we established relations between the different level of analysis in order to ensure a proper treatment of processes of analysis and annotation of such multimedia objects. We stressed the necessity of a linguistic representation level in these framework and offered an ontological model of this level. Finally we showed how an alignment of different levels of semantic is possible in the context of a foundational ontology like DOLCE for a successful use in systems like SmartWeb.

Future work is needed to completely specify the processing part of the modeling. The approach in Sec. 5 will be taken into account for this purpose and actual work is concentrating on harmonizing the two approaches.

Acknowledgments

The research presented here is sponsored by the German Ministry of Research and Technology (BMBF) under grant 01IMD01A (SmartWeb). We thank our student assistants and the project partners. The responsibility for this papers lies with the authors.

References

1. Richard Arndt, Steffen Staab, Raphael Troncy and Lynda Hardman: "*Adding Formal Semantics to MPEG-7: Designing a Well-Founded Multimedia Ontology for the Web*". Arbeitsberichte aus dem Fachbereich Informatik, 4/2007, Universit́t Koblenz-Landau, ISSN
2. Benitez, A., Rising, H., Jorgensen, C., Leonardi, R., Bugatti, A., Hasida, K., Mehrotra, R., Tekalp, A., Ekin, A. and Walker, T.: "*Semantics of Multimedia in MPEG-7*". In Proc. of IEEE International Conference on Image Processing (ICIP), 2002.
3. Paul Buitelaar, Thierry Declerck, Anette Frank, Stefania Racioppa, Malte Kiesel, Michael Sintek, Massimo Romanelli, Ralf Engel, Daniel Sonntag, Berenike Loos, Vanessa Micelli, Robert Porzel and Philipp Cimiano: "*LingInfo: A Model for the Integration of Linguistic Information in Ontologies*". In Proceedings of Workshop on Interfacing Ontologies and Lexical Resources for Semantic Web Technologies (OntoLex 2006). Genova, Italy, May 27, 2006.
4. Paul Buitelaar, Michael Sintek and Malte Kiesel: "*A Lexicon Model for Multilingual/Multimedia Ontologies*". In: Proceedings of the 3rd European Semantic Web Conference (ESWC06), Budva, Montenegro, June 2006.
5. Buitelaar, P., Cimiano P., Racioppa S. and Siegel, M.: "*Ontology-based Information Extraction with SOBA*". In Proc. of the 5th Conference on Language Resources and Evaluation (LREC 2006).
6. Cimiano, P., Eberhart, A., Hitzler, P., Oberle, D., Staab, S., Studer, S.: "*The SmartWeb Foundational Ontology*". Technical report, Institute for Applied Informatics and Formal Description Methods (AIFB) University of Karlsruhe, SmartWeb Project, Karlsruhe, Germany, 2004.
7. Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., Schneider, L.: "*Sweetening Ontologies with DOLCE*". In Proc. of the 13th International Conference on Knowledge Engineering and Knowledge Management (EKAW02), volume 2473 of Lecture Notes in Computer Science, Sigünza, Spain, 2002.
8. Aldo Gangemi and Peter Mika:"*Understanding the semantic web through descriptions and situations*". In Databases and Applications of SEmantics (ODBASE 2003), Catania, Italy, November 37, 2003
9. Aldo Gangemi, Stefano Borgo, Carola Catenacci and Jobs Lehmann: "*Deliverable of the EU FP6 project Metokis*", 2005.
10. Hughes, T., C. and Ashpole, B., C.: "*The Semantics of Ontology Alignment*". In Information Interpretation and Integration Conference (I3CON), August 2004.
11. Hunter, J.: "*Adding Multimedia to the Semantic Web - Building an MPEG-7 Ontology*". In Proc. of the International Semantic Web Working Symposium (SWWS), 2001.
12. Claudio Masolo, Stefano Borgo, Aldo Gangemi, Nicola Guarino, Alessandro Oltramari, and Luc Schneider: "*The WonderWeb Library of Foundational Ontologies*". WonderWeb Deliverable D17, August 2002. http://wonderweb.semanticweb.org.
13. Niles, I., Pease, A.: "*Towards a Standard Upper Ontology*". In Proc. of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001), C. Welty and B. Smith, Ogunquit, Maine, 2001.
14. D. Oberle, A. Ankolekar, P. Hitzler, P. Cimiano, C. Schmidt, M. Weiten, B. Loos, R. Porzel, H.-P. Zorn, V. Micelli, M. Sintek, M. Kiesel, B. Mougouie, S. Vembu, S. Baumann, M. Romanelli, P. Buitelaar, R. Engel, D. Sonntag, N. Reithinger, F. Burkhardt, and J. Zhou: "*DOLCE ergo SUMO: On Foundational and Domain Models in SWIntO (SmartWeb Integrated Ontology)*". In Journal of Web Semantics: Science, Services and Agents on the World Wide Web (2007).

15. Daniel Oberle, Steffen Lamparter, StephanGrimm, Denny Vrandecic, Steffen Staab and Aldo Gangemi: "*Towards Ontologies for Formalizing Modularization and Communication*". In Large Software Systems. Applied Ontology, Volume 1, Number 2/2006, pp. 163-202, IOS Press.
16. G. Th. Papadopoulos, V. Mezaris, I. Kompatsiaris and M. G. Strintzis: "*Ontology-Driven Semantic Video Analysis Using Visual Information Objects*". In Proceedings of the Second International Conference on Semantic and Digital Media Technology (SAMT), p. 37-38, Genova, Italy, December 5-7, 2007
17. Peirce, C.S., Collected Papers of Charles Sanders Peirce, vols. 1-6, Charles Hartshorne and Paul Weiss (eds.), vols. 7-8, Arthur W. Burks (ed.), Harvard University Press, Cambridge, MA, 1931-1935, 1958.
18. Massimo Romanelli, Daniel Sonntag and Norbert Reithinger. "*Connecting Foundational Ontologies with MPEG-7 Ontologies for Multimodal QA*". In Proceedings of the 1st International Conference on Semantic and Digital Media Technology (SAMT), p. 37-38, Athens, Greece, December 6-8, 2006
19. Ron Scollon, Suzie Wong Scollon. "*Discourses in Place - Language in the Material World*". Routeledge, London, 2003.
20. Claude Shannon, "A Mathematical Theory of Communication", Bell System Technical Journal, vol. 27, pp. 379-423, 623-656, July, October, 1948.
21. Sonntag, D., Romanelli, M.: "*A Multimodal Result Ontology for Integrated Semantic Web Dialogue Applications*". In Proc. of the 5th Conference on Language Resources and Evaluation (LREC 2006).
22. Wahlster, W.: SmartWeb: Mobile applications of the semantic web. In: P. Dadam and M. Reichert, editors, *GI Jahrestagung 2004*, Springer, 2004.