Connecting Foundational Ontologies with MPEG-7 Ontologies for Multimodal QA

Massimo Romanelli, Daniel Sonntag, and Norbert Reithinger

***Abstract*— In the SMARTWEB project [1] we aim at developing a context-aware, mobile, and multimodal interface to the Semantic Web. In order to reach this goal we provide a integrated ontological framework offering coverage for deep semantic content, including ontological representation of multimedia based on the MPEG-7 standard[1]. A discourse ontology covers concepts for multimodal interaction by means of an extension of the W3C standard EMMA[2]. For realizing multimodal/multimedia dialog applications, we link the deep semantic level with the media-specific semantic level to operationalize multimedia information in the system. Through the link between multimedia representation and the semantics of specific domains we approach the Semantic Gap.**

***Index Terms*— Multimedia systems, Knowledge representation, Multimodal ontologies, ISO standards.**

I. INTRODUCTION

WORKING with multimodal, multimedia dialog applications with question answering (QA) functionality assumes the presence of a knowledge model that ensures appropriate representation of the different levels of descriptions. Ontologies provide instruments for the realization of a well modeled knowledge base with specific concepts for different domains. For related work, see e.g. [2]–[5].

Within the scope of the SMARTWEB project[3] we realized a multi-domain ontology where a media ontology based on MPEG-7 supports meta-data descriptions for multimedia audio-visual content; a discourse ontology based on the W3C standard EMMA covers multimodal annotation. In our approach we assign conceptual ontological labels according to the ontological framework (figure 1) to either complete multimedia documents, or entities identified therein. We employ an abstract foundational ontology as a means to facilitate domain ontology integration (combined integrity, modeling consistency, and interoperability between the domain ontologies). The ontological infrastructure of SMARTWEB, the SWINTO (**S**mart**W**eb **Int**egrated **O**ntology), is based on a upper model ontology realized by merging well chosen concepts from two established foundational ontologies, DOLCE [7] and SUMO [8], into the SMARTWEB foundational ontology SMARTSUMO [9]. The domain-specific knowledge like sport-event, navigation, or webcam is defined in dedicated ontologies modeled as sub-ontologies of SMARTSUMO. Semantic integration takes place for heterogeneous information sources: extraction results from semi-structured data such as tabular structures which are stored in an ontological knowledge base [10], and hand-annotated multimedia instances such as images of football teams. In addition, Semantic Web Services deliver MPEG-7 annotated city maps with points of interest.

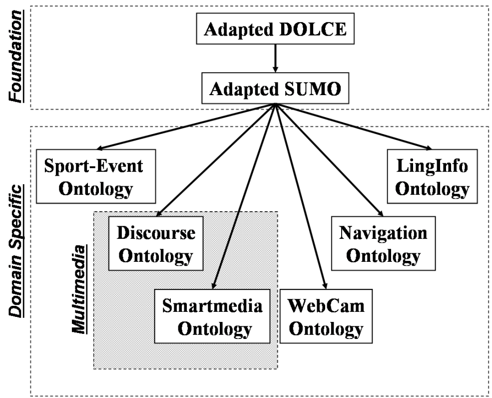

Fig. 1. SMARTWEB's Ontological Framework for Multimodal QA

II. THE DISCONTO AND SMARTMEDIA ONTOLOGIES

The SWINTO supplements QA specific knowledge in a discourse ontology (DISCONTO) and represents multimodal information in a media ontology (SMARTMEDIA). The DISCONTO provides concepts for dialogical interaction with the user and with the Semantic Web sub-system. The DISCONTO models multimodal dialog management using SWEMMA, the SMARTWEB extention of EMMA, dialog acts, lexical rules for syntactic-semantic mapping, HCI concepts (pattern language for interaction design), and semantic question/answer types. Concepts for QA functionality are realized with the discourse:Query concept specifying emma:interpretation. It models the user query to the system in form of a partially filled ontology instance. The discourse:Result concept references information the user is asking for [11]. In order to efficiently search and browse multimedia content SWINTO specifies a multimedia sub-ontology called SMARTMEDIA. SMARTMEDIA is an ontology for semantic annotations based

This research was funded by the German Federal Ministry for Education and Research under grant number 01IMD01A.

M. Romanelli, D. Sonntag and N. Reithinger are with DFKI GmbH – German Research Center for Artificial Intelligence, Stuhlsatzenhausweg 3, d-66123 Saarbrücken, Germany {romanell,sonntag,bert}@dfki.de

[1] http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm

[2] http://www.w3.org/TR/emma/

[3] SMARTWEB aims to realize a mobile and multimodal interface to Semantic Web Services and ontological knowledge bases [6]. The project moves through three scenarios: handheld, car, and motorbike. In the handheld scenario the user is able to pose multimodal closed- and open-domain questions using speech and gesture. The system reacts with a concise and short answer and the possibility to browse pictures, videos, or additional text information found on the Web or in Semantic Web sources (http://www.smartweb-project.de/).

on MPEG-7. It offers an extensive set of audio-visual descriptions for the semantics of multimedia [12]. Basically, the SMARTMEDIA ontology uses the MPEG-7 multimedia content description and multimedia content management (see [13] for details on description schemes in MPEG-7) and enriches it to account for the integration with domain-specific ontologies. A relevant contribution of MPEG-7 in SMARTMEDIA is the representation of multimedia decomposition in space, time, and frequency as in the case of the general mpeg7:Segment-Decomposition concept. In addition we use file format and coding parameters (mpeg7:MediaFormat, mpeg7:MediaProfile, etc.).

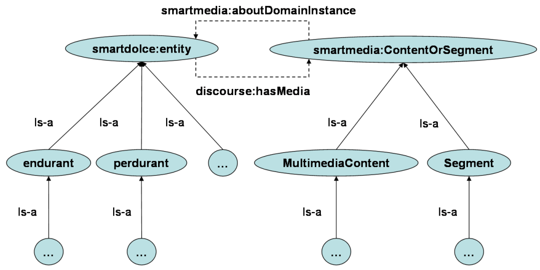

Fig. 2. The SWINTO - DISCONTO - SMARTMEDIA Connection

III. CLOSING THE SEMANTIC GAP

In order to close the Semantic Gap deriving from the different levels of media representations, namely the surface level referring to the properties of realized media as in the SMARTMEDIA, and the deep semantic representation of these objects, the smartmedia:aboutDomainInstance property with range smartdolce:entity has been added to the top level class smartmedia:ContentOrSegment (see fig. 2). In this way the link to the upper model ontology is inherited to all segments of a media instance decomposition, so that we can guarantee for deep semantic representation of the SMARTMEDIA instances referencing the specific media object, or the segments making up the decomposition. Through the discourse:hasMedia property with range smartmedia:ContentOrSegment located in the smartdolce:entity top level class and inherited to each concept in the ontology, we realize a pointer back to the SMARTMEDIA ontology.

This type of representation is useful for pointing gesture interpretation and co-reference resolution. A map which is obtained from the Web Services to be displayed on the screen shows selectable objects (e.g. restaurants, hotels), and the map is represented in terms of an mpeg7:StillRegion instance, decomposed into different mpeg7:StillRegion instances for each object segment of the image. The MPEG-7 instances are linked to a domain-specific instance, i.e., the deep semantic description of the picture (in this case the smartsumo:Map) or the segment of picture (e. g., navigation:ChineseRestaurant). In this way the user can refer to the restaurant by touching on the displayed map. Hence a multimodal fusion component can directly process the referred navigation:ChineseRestaurant instance performing linguistic co-reference resolution: *What's the phone number here?*

IV. CONCLUSION

We presented the connection of our foundational ontology with an MPEG-7 ontology for multimodal QA in the context of the SMARTWEB project. The foundational ontology is based on two upper model ontologies and offers coverage for deep semantic ontologies in different domains. To capture multimedia low-level semantics we adopted an MPEG-7 based ontology that we connected to our domain-specific concepts by means of relations in the top level classes of the SWINTO and SMARTMEDIA. This work enables 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 on the discourse level.

ACKNOWLEDGMENT

We would like to thank our partners in SmartWeb. The responsibility for this article lies with the authors.

REFERENCES

[1] Wahlster, W.: Smartweb: Mobile applications of the semantic web. In: P. Dadam and M. Reichert, editors, *GI Jahrestagung 2004*, Springer, 2004.
[2] Reyle, U., Saric, J.: Ontology Driven Information Extraction, Proceedings of the 19th Twente Workshop on Language Technology, 2001.
[3] Lopez, V., Motta, E.: Ontology-driven Question Answering in AquaLog In Proceedings of 9th International Conference on Applications of Natural Language to Information Systems (NLDB), 2004.
[4] Niekrasz, J. and Purver, M.: A multimodal discourse ontology for meeting understanding. In Bourlard, H. and Bengio, S., editors, Proceedings of MLMI'05. LNCS.
[5] Nirenburg, S., Raskin, V.: Ontological Semantics, MIT Press, 2004.
[6] Reithinger, N., Bergweiler, S.,Engel, R., Herzog, G., Pfleger, N., Romanelli, M., and Sonntag, D.: A Look Under the Hood - Design and Development of the First SmartWeb System Demonstrator. In: Proc. ICMI 2005, Trento, 2005.
[7] Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., Schcneider, 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] 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.
[9] 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.
[10] Buitelaar, P., Cimiano P., Racioppa S., Siegel, M.: Ontology-based Information Extraction with SOBA In Proc. of the 5th Conference on Language Resources and Evaluation (LREC 2006).
[11] 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).
[12] Benitez, A., Rising, H., Jorgensen, C., Leonardi, R., Bugatti, A., Hasida, K., Mehrotra, R., Tekalp, A., Ekin, A., Walker, T.: Semantics of Multimedia in MPEG-7. In Proc. of IEEE International Conference on Image Processing (ICIP), 2002.
[13] Hunter, J.: Adding Multimedia to the Semantic Web - Building an MPEG-7 Ontology. In Proc. of the Internatioyesnal Semantic Web Working Symposium (SWWS), 2001.