A Multimodal Result Ontology for Integrated Semantic Web Dialogue Applications

Daniel Sonntag*, **Massimo Romanelli***

*DFKI GmbH – German Research Center for Artificial Intelligence
Stuhlsatzenhausweg 3,
d-66123 Saarbrücken, Germany
{sonntag, romanelli}@dfki.de

Abstract

General purpose ontologies and domain ontologies make up the infrastructure of the Semantic Web, which allow for accurate data representations with relations, and data inferences. In our approach to multimodal dialogue systems providing question answering functionality (SMARTWEB), the ontological infrastructure is essential. We aim at an integrated approach in which all knowledge-aware system modules are based on interoperating ontologies in a common data model. The discourse ontology is meant to provide the necessary dialogue- and HCI concepts. We present the ontological syntactic structure of multimodal question answering results as part of this discourse ontology which extends the W3C EMMA annotation framework and uses MPEG-7 annotations. In addition, we describe an extension to ontological result structures where automatic and context-based sorting mechanisms can be naturally incorporated.

1. Introduction

Integrating ontologies means to allow for transformations between generated ontologies and the already existing ones as Semantic Web artefacts available at the start of a Semantic Web project.

General ontologies such as SUMO (Niles and Pease, 2001) and DOLCE (Gangemi et al., 2002) exist to be used and exploited as general purpose conceptual terminologies. The use of different ontologies in a unique application assumes the presence of a upper model ontology defining base-concepts and relations to be inherited from the subordinate domain ontologies, providing suitability for fast reasoning and consistent representation.

For SMARTWEB[1] SUMO and DOLCE have been merged. The SMARTWEB foundational ontology (Cimiano et al., 2004) brings together the conceptual clarity of DOLCE, due to the Ontoclean(Guarino and Welty, 2004) methodology used to model it, and the accessibility of SUMO.

Connecting and merging different data- and knowledge sources representing knowledge of special domains like *football* or *chemistry* is a difficult task, even if such upper level ontology models exist. Although difficult to establish, an integrated data approach is the key for functionalities such as query decomposition, ontological inference, and at least trust and transparency in the results obtained. Therefore, all knowledge-intensive sub-modules of a Semantic Web application should themselves be represented by an ontological infrastructure, e.g. RDF-based data exchange[2] (Randal, 1998) which we used finally.

With the help of our discourse ontology (DISCONTO) the pertinent concepts and relations for content-based fusion of input modalities, semantic-based interpretation of user input (query template filling) for ontology-based similarity search, and multimodal output filtering, aggregation, and generation can be modelled concurrently.

What follows in section 2. is a brief introduction of the SMARTWEB ontological framework, focusing on a specific branch of the discourse ontology, the concepts for representing multimedia results syntactically, with content-based semantic annotations. In the question answering (QA) scenario as metaphorical basis of the user-system interaction, the presentation of multimodal results is very important, which is reflected in the detailed ontologicial description for this data types. Nonetheless the integration of dialogue- and QA-based concepts into the upper model ontology is a difficult task, since the conceptual clarity of the upper model must not become blurred. In addition, specific domain- and task-oriented ontologies like ontologies for dialogue and discourse processing must not present a narrow-sighted view on the relevant concepts and relations. In the remainder of this text, we present our approach to integrate question answering and multimodal dialogue concepts into the ontological frame, focusing on concept for result presentation. Group (list) results and sorting criteria for QA results are presented in detail.

2. Ontological Framework

SMARTWEB joins together multimodal and mobile user interface technology with question answering technology.

The dialogue functionality of Smartweb can basically be summarised as a question answer game; the user can ask constituent interrogative questions and refer to objects previously mentioned in the dialogue. Most prominent are single instance questions of semantic classes such as persons, locations, and dates. The application we describe is a dialogue system, where the user can ask questions about a specific domain (football) as well as open domain questions. In addition, the input and output is multimodal, which on the one hand increases naturalness in query formulation, but on

[1] SMARTWEB (Wahlster, 2004; Reithinger et al., 2005) aims at developing a context-aware, mobile and multimodal interface to the Semantic Web. In the main scenario, the user carries a smartphone and is able to pose multimodal open-domain questions using speech, pen, and gesture among other input modalities. The user input is transmitted via UMTS or WLAN to a backend server, where the multimodal recogniser, the Semantic Web access subsystems and the dialogue manager for result generation resist.

[2] http://www.w3.org/TR/rdf-primer/

the other hand restricts the usability of a textual question answering system.
Example questions for single instance constituents are the following:

- Who was world champion in 1990?
- Show me the goal of Alexandersson.
- Show me the world cup mascot 2006.
- Where is Brandenburg Gate located?
- What is the capital of Germany?
- When does the next football match take place?

Example questions for enumeration of constituents are the following:

- Show me the World Cup mascot.
- When was Italy world champion?
- Who was world champion for more than three times?
- And between 1970 and 1980?
- Which games had more than 100000 spectators?
- Which games took place in Stuttgart?
- Who scored in the final match Germany vs. Brasil?

To reach the goal to correctly answer the questions we combine different kinds of domain ontologies into an integrated and modular knowledge base. For this purpose we defined an upper model ontology based on SUMO and DOLCE and integrated each domain ontology in it. This way we could guarantee interoperability and modelling consistency between the different ontologies.
A domain-ontology is the primary knowledge server for answering question. This offers great opportunity for more specific questions in a dialogical interaction, on the other hand the difficulties in speech recognition, language understanding and information coverage uncover the demand for open-domain QA functionality (Neumann and Sacaleanu, 2005) obvious, which does not necessarily require complex syntactic/semantic question parsing, and may use answer redundancy as a surrogate for semantic understanding of questions (Lin, 2002). However, for particular question types, i.e. enumeration questions, ontology-based systems outperform standard open-domain QA system for the following two reasons: (1) Redundancy information [3] often does not help very much for list questions, (2) ontological representation of multimodal list result allow for deeper structuring, filtering, and sorting.
We present a multimodal ontological result structure where automatic and context-based sorting mechanism can be naturally incorporated.

[3] The more frequently an answer appears in a document corpus, the easier it is to find and hence expected to be correct. Confer e.g. (Breck et al., 2001)

2.1. The Upper Model Ontology

Over the DISCONTO we link concepts of the domain specific ontologies to a media representation ontology SMARTMEDIA that provides concepts for multimodal presentation.
The SMARTWEB upper model ontology basically proposes the upper-level concepts of DOLCE with the distinction between `endurant`, `perdurant`, `abstract` and `qualities`, and the adoption of several concepts from SUMO adding rich taxonomies which feature specific concepts such as `Hotel` or `Organization` which we use as semantic answer types.
The upper model (SmartSUMO) was modelled in three steps:

- Define relevant concepts from the DOLCE ontology (SmartDOLCE).
- Define relevant concepts from the SUMO ontology.
- Align SUMO concepts in the DOLCE superstructure (SmartSUMO).

The SmartDOLCE ontology includes a relevant module (sub-ontology) of DOLCE called *Descriptions & Situations* (D&S) (Gangemi and Mika, 2003) to standardise a variety of reified contexts and states of affairs. We use this module for representing contexts for sorting criteria on group results (section 2.6.).

2.2. The Discourse Ontology

The DISCONTO discourse ontology is meant to provide concepts for dialogical interaction with the user (HCI) and the interaction with the Semantic Web access sub-systems, which are based on the ontological infrastructure. We identified the following top-level constructs of the DISCONTO ontology: (1) dialogue acts, (2) dialogue memory, dialogue model, (3) lexical rules for syntactic/semantic mapping, and (4) concepts for multimodal results to be presented to the user which incorporate concepts for question answering as a sub-category.
User interaction is modelled in a generic way in the discourse ontology. As opposed to other contributes like in (Pfleger et al., 2003) where a three layer model (linguistic, discourse and dialog layer) is used, or (Niekrasz and Purver, 2005), a semiotic approach to discourse, in our approach we concentrate on discourse interactions within a question answering scenario. The DISCONTO (Smartweb Discourse Ontology) divides dialogue acts in `Queries` and `Results` concepts. Specific discourse phenomena like ellipsis or anaphora are resolved in a dedicated multimodal fusion and discourse processing module (Pfleger, 2005).

2.2.1. Concepts for Multimedia Results

The discourse ontology branch describing the multimodal interaction is based on the W3C EMMA standard[4]. EMMA was primarily designed to code interpretations of multimodal input devices such as the output of a speech recogniser. We extended the EMMA language to meet the requirements of representing special multimodal answer

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

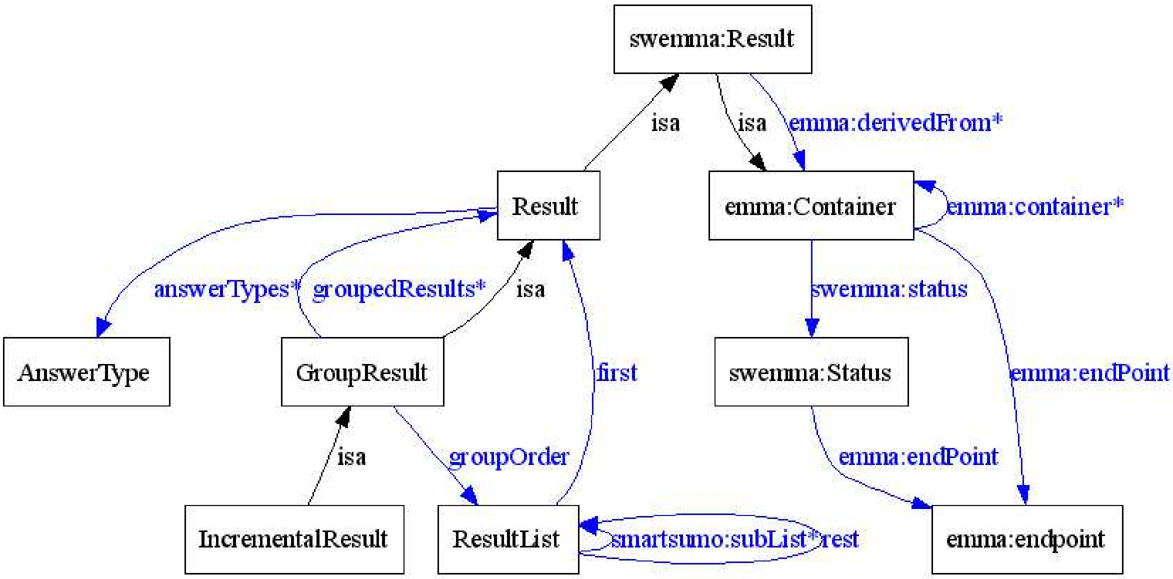

Figure 1: Graphical representation of the *swemma:Result* class as EMMA extension. *GroupResult* groups several results in a linked list structure. *IncrementalResult* provides an incremental result structure, according to different time constraints during the Semantic Web lookup.

types in the context of the ontological infrastructure. What we present here, is the ontological representation of multimodal result objects which serve as the answers obtained from the Semantic Web information retrieval step.

We provide a new namespace *swemma* for our extensions to the *emma:container* and *emma:interpretation* classes. One of these extensions is the *swemma:Result* class, extended by *disconto:GroupResult*, and *disconto:IncrementalResult*. Figure 1 shows the ontological structure. [5] We focus on the *answerTypes* slot of the *Result* base class.

Answer types are multimodal, hierarchical dialogue concepts of the results to be presented to the user. To ensure full coverage of all forms of multiple answers, and to be able to factor in deviations and special requirements, a typology is being developed. More general answer types should be overwritten by more specific ones, if possible. (See e.g. (Eduard Hovy and Ravichandran, 2001) for similar approaches for a uni-modal textual context. This taxonomy consists of mainly *semantic* answer types[6]. The result structure as a whole combines *semantic* answer types with a *syntactic* result structure by the GroupResult concept. We exemplify the modelling of results by discussing an actual result instance.

2.3. An MPEG-7 Ontology

We represent output results with a MPEG-7-based media annotation ontology named SMARTMEDIA.

MPEG-7[7] is conceived for describing multimedia content data. Primarily the concepts *mpeg7:MediaFormat* for file format and the coding parameters, *mpeg7:MediaProfile* 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 following the approach in (Hunter, 2001) restricting the number of the modelled concepts to those that fit well to the project.

Concept instances (ContentAnswer) of the SMARTMEDIA ontology are linked to the result instances by the *discourse:answerTypes* slot in the representation of results as shown in figure 2.

The important thing to notice is that all answer types refer to media types as multimodal SMARTMEDIA concepts imported into our upper layer ontology with namespaces *smartmedia* and *mpeg7*. In this way we bridge the discourse ontology with the extended SUMO ontology to arrive at a common data model.

2.4. SmartWeb Integrated Ontology (SWINTO)

The concepts of the DISCONTO have been aligned to the SMARTWEB integrated ontology (SWINTO). The EMMA derived concepts inherit from *smartsumo:Software*. Dialog acts have been mapped as intentional processes to *smartsumo:Communication*. The *discourse:Query* concept relates to the domain specific concepts over a *discourse:content* relation defining a partially instantiated instance used as search pattern for the semantic knowledge base (ontobroker)(Decker et al., 1999), and the *discourse:focus* relation expresses which part of the content represents the knowledge base instance or atomic value the user asks for.

The SMARTMEDIA ontology is connected to the SmartSUMO by a high level semantic concept called *smart-*

[5] The inherited classes of *swemma:Result* have the disconto namespace, because this guaranties access to list concepts of the upper level ontology, due to technical reasons.

[6] Semantic answer types are semantic named entity classes such as Location and Person in the multimodal context, e.g. music in a disco location or portraits of persons. They can be both open classes (e.g. Person) or closed classes (e.g. Nationality).

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

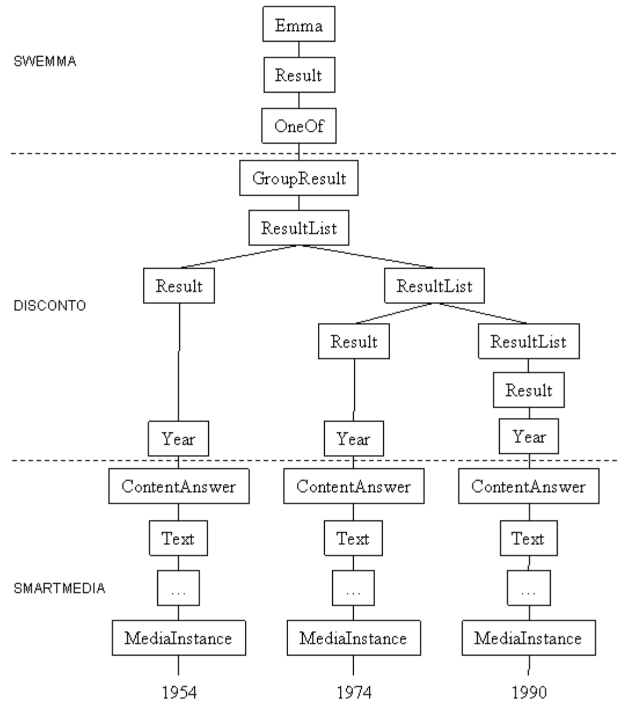

Figure 2: Graphical representation of a *swemma:Result* instance for the question *"When did Germany win the soccer world championship?"*. *GroupResult* groups several results in a linked list structure. The three ontological levels are: the extended EMMA (SWEMMA), the discourse ontology (DISCONTO), and the MPEG7 ontology (SMARTMEDIA).

media:ContentAnnotation. This concept directly relates to the top level concept *smartdolce:entity* and to the higher concept in the smartmedia taxonomy *smartmedia:ContentOrSegment*. Over this concept a medium instance (e.g. video, picture, text) is associated with domain specific concepts.

2.5. Defining and Declaring Sort Criteria for Group Results

In order to properly sort results for presentation purpose, we realised a context sensitive mechanism for determining parameters of an arbitrary sorting function. We observe by inspecting different list results, that correct sorting of results for presentations heavily depends on user expectations, similar to answer types. Those expectation can be partly deduced from context information. For example, if a user asks for "nearby cities where a football match takes place", he expects the cities to be ordered with respect to the distance from the place he is situated.
We adopted the approach of ontological patterns for modelling schematic knowledge of the pragmatics of spatial navigation. as described in (Loos and Porzel, 2005).
In our approach we use ontological pattern to generate a description for an ordering function in a situative context. The description of `SortResults` would sequences a `SortGroupResults` use roles - such as *discourse:SortCriterion* played by *smartsumo:City* and a parameter such as *smartsumo:Nation*, which follow the constraints in D&S, for which roles are played by endurants, parameterized by regions.

2.6. Declaring a Sort Function

We presented a multimodal ontological result structure where automatic and context-based sorting mechanism can be naturally incorporated. In this section we focus on the sorting aspect.
Expressive power of logic-based ontology languages can be measures using a so-called *bisimulation* (Blackburn et al., 2001). The expressive power of the RDF-based ontological representation is quite low, on the other hand allows for decidable and fast inferences in the knowledge base (ONTOBROKER)(Decker et al., 1999) as well as for similar representations among different components, such as a dialogue system and the backend server (Reithinger et al., 2005)(Reithinger and Sonntag, 2005). On interesting research question is how to boost expressiveness of ontological languages by adding additional operators and relations in addition to subsumptions. Unfortunately, description languages do not offer the necessary expressiveness for arbitrary (semantic) sorting criteria we would like to use. Driven by pragmatic issues, and algorithmic concerns, we do not attempt to extend the expressive power of the ontology for Group Results. Instead, we use the JENA-Java framework (McBride, 2001) to implement arbitrary sorting functions on instance lists. Since the JENA-Java framework uses the same terminological box (T-Box) and assertion box (A-Box) as the reasoning mechanism of the ontology, a type-save but rather independent bridge to object-oriented programming can be established. The sorting functions can thus be implemented in an arbitrary fashion, and can be executed on deeper structured ontological group result instances. Hence al relevant phenomena and contextual factors that have to be taken into account for sorting QA results for presentation can be taken into account without the need for a direct representation within the ontological framework.

3. Conclusion

In project SMARTWEB, the ontology serves as a basis for defining the semantics and the content of information exchanged between various system modules so that modules only receive messages whose content is congruent to the terminological and structural distinctions of the ontology. This serves as communication interfaces between NLP components. The natural extension is an ontological framework towards QA result design.
Our multimodal result structure extends the W3C EMMA annotation framework (for syntactic) and uses MPEG7 annotations (for semantic) ontological QA result representations.
We made the domain assumptions more explicit and operational by linking the Sorting Criteria ontology concepts towards a JENA-Java framework to write and execute arbitrary (semantic) sorting functions.

4. Acknowledgements

We would like to thank our partners in SmartWeb. This research was funded by the German Federal Ministry for

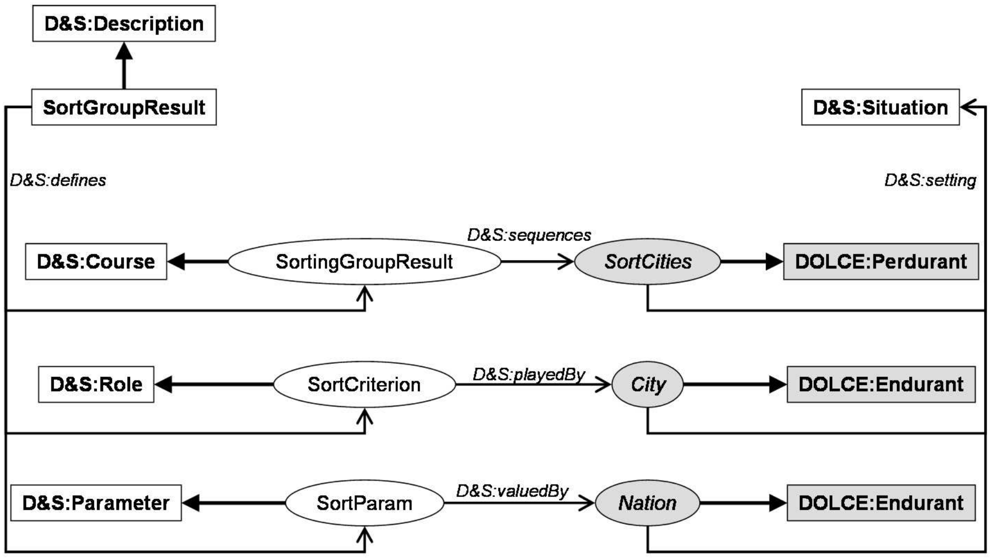

Figure 3: D&S Representation of SortResult criterion.

Education and Research under grant number 01IMD01A. The responsibility for this article lies with the authors.

5. References

Patrick Blackburn, Maarten de Rijke, and Yde Venema. 2001. *Modal Logic*. Cambridge University Press.

E. Breck, M. Light, G. Mann, E. Rilo, B. Brown, P. Anand, M. Rooth, and M. Thelen. 2001. Looking under the hood: Tools for diagnosing your question answering engine.

Philipp Cimiano, Andreas Eberhart, Pascal Hitzler, Daniel Oberle, Steffen Staab, , and Rudi Studer. 2004. The smartweb foundational ontology. Technical report, Institute for Applied Informatics and Formal Description Methods (AIFB) University of Karlsruhe, Karlsruhe, Germany. SmartWeb Project.

Stefan Decker, Michael Erdmann, Dieter Fensel, and Rudi Studer. 1999. Ontobroker: Ontology based access to distributed and semi-structured unformation. In *DS-8*, pages 351–369.

Ulf Hermjakob Chin-Yew Lin Eduard Hovy, Laurie Gerber and Deepak Ravichandran. 2001. Towards semantic-based answer pinpointing.

Aldo Gangemi and Peter Mika. 2003. Understanding the semantic web through descriptions and situations. In *Databases and Applications of SEmantics (ODBASE 2003)*, Catania, Italy, November 3–7.

Aldo Gangemi, Nicola Guarino, Claudio Masolo, Alessandro Oltramari, and Luc Schcneider. 2002. Sweetening Ontologies with DOLCE. In *In 13th International Conference on Knowledge Engineering and Knowledge Management (EKAW02)*, volume 2473 of Lecture Notes in Computer Science, page 166 ff, Sigünza, Spain, Oct. 1–4.

Nicola Guarino and Christopher A. Welty. 2004. An overview of ontoclean. In Steffen Staab and Rudi Studer, editors, *Handbook on Ontologies*, International Handbooks on Information Systems, pages 151–172. Springer.

Jane Hunter. 2001. Adding Multimedia to the Semantic Web - Building an MPEG-7 Ontol ogy. In *Proceedings of the International Semantic Web Working Symposium (SWWS)*.

Jimmy Lin. 2002. The web as a resource for question answering: Perspective and challenges. In *Proceedings of the third International Conference on Language Resources and Evaluation (LREC 2002)*, volume Canary Islands, Spain.

Berenike Loos and Robert Porzel. 2005. Towards Ontology-based Pragmatic Analysis. In Claire Gardent and Bertrand Gaiffe, editors, *Proceedings of 9th Workshop on the Semantics and Pragmatics of Dialogue (Dialor05)*, pages 163–166, Nancy, France, June 9-11.

Brian McBride. 2001. Jena: Implementing the rdf model and syntax specification. In *Proceedings of the 2nd International Workshop on the Semantic Web.*, Hongkong, May 1.

Günter Neumann and Bogdan Sacaleanu. 2005. Multiple language question answering track. In *Working Notes for the CLEF 2005 Workshop*, Vienna, Austria, 21-23 September.

John Niekrasz and Matthew Purver. 2005. A multimodal discourse ontology for meeting understanding. In *MLMI*, pages 162–173.

Ian Niles and Adam Pease. 2001. Towards a Standard Upper Ontology. In Chris Welty and Barry Smith, editors, *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001)*, Ogunquit, Maine, October 17–19.

Norbert Pfleger, Jan Alexandersson, and Tilman Becker. 2003. A robust and generic discourse model for multimodal dialogue. In *Workshop Notes of the IJCAI-03 Workshop on "Knowledge and Reasoning in Practical Dialogue Systems"*, Acapulco, Mexico, August.

Norbert Pfleger. 2005. Fade - an integrated approach to multimodal fusion and discourse processing. In *Proceedings of the Dotoral Spotlight at ICMI 2005*, Trento, Italy.

Damian Mac Randal. 1998. Resource description framework (rdf). Technical report, W3C Office, Rutherford Appleton Laboratory, UK, April.

Norbert Reithinger and Daniel Sonntag. 2005. An integration framework for a mobile multimodal dialogue system accessing the semantic web. In *Proc. of Interspeech'05*, Lisbon, Portugal.

Norbert Reithinger, Simon Bergweiler, Ralf Engel, Gerd Herzog, Norbert Pfeger, Massimo Romanelli, and Daniel Sonntag. 2005. A Look Under the Hood Design and Development of the First SmartWeb System Demonstrator. In *Proceedings of 7th International Conference on Multimodal Interfaces (ICMI 2005)*, Trento, Italy, October 04-06.

Wolfgang Wahlster. 2004. SmartWeb: Mobile Applications of the Semantic Web. In Peter Dadam and Manfred Reichert, editors, *GI Jahrestagung 2004*, pages 26–27. Springer.