Towards Combining Finite State, Ontologies, and Data Driven Approaches to Dialogue Management for Multimodal Question Answering

Daniel Sonntag

DFKI GmbH – German Research Center for Artificial Intelligence
Stuhlsatzenhausweg 3
D-66123 Saarbrücken, Germany
sonntag@dfki.de

Abstract

Information-providing dialogue systems typically use one of the following dialogue strategies: finite-state based, frame-based, or agent-based. Recent extensions are concerned with hybrid models, where mixed automaton and information state approaches are combined. We report on our multimodal mobile Semantic Web access system SMARTWEB which uses a more rigid dialogue management strategy to ensure operability and robustness of the system while allowing for flexible dialogical interaction in a question answering scenario. In addition, a strategy to incorporate information state approaches into the running system to be extended towards a machine learning scenario for very particular, e.g. domain or language specific, dialogue management decisions is proposed.

Korak h kombiniranju končnih avtomatov, ontologij in pristopov na podlagi podatkov za vodenje dialoga pri multimodalnem odgovarjanju na vprašanja

Sistemi dialoga za dajanje informacij običajno uporabljajo eno od strategij, značilnih za dialog: takšno, ki temelji bodisi na končnih avtomatih, na okvirih ali agentih. Najnovejše širitve se ukvarjajo s hibridnimi modeli, pri katerih se kombinirata pristopa avtomatov in informacijskega stanja. V prispevku predstavljamo svoj multimodalni mobilni sistem SMARTWEB za dostopanja do mreže Semantic Web, ki uporablja bolj togo strategijo vodenja dialoga in s tem zagotavlja operabilnost in robustnost sistema, hkrati pa dovoljujemo prilagodljivo dialoško interakcijo v scenariju odgovarjanja na vprašanja. Poleg tega predlagamo strategijo za vključevanje pristopov informacijskega stanja v delujoč sistem, ki bi se razširil v scenarij strojnega učenja za zelo specifične odločitve vodenja dialoga, npr. za določeno področje ali jezik.

1. Introduction

Dialogue systems often use finite-state-automata (FSA) based dialogue management strategies where the dialogue flow is represented by a path through a finite-state machine. More flexible strategies are frame-based (frame slots are filled dynamically), or agent-based (the interaction is free as far as possible according to some dialogue objectives, e.g. user objectives) (Chu et al., 2005; McTear, 2002). Recent extensions are concerned with hybrid models, whereby automaton and information state (IS) approaches are combined (e.g. (Horacek and Wolska, 2005)). In context of the SMARTWEB system (Wahlster, 2004; Reithinger et al., 2005), we extend hybrid dialogue models for mobile multimodal interaction and explore, how information states can be extracted from dialogue processing data, in particular from ontology structures. The goal is to integrate data-driven approaches to dialogue management.

In our approach, Semantic Web (Fensel et al., 2003) structures form the representation basis of dialogue processing data which allows for extracting machine learning features for dialogue adaptations in a specific application scenario: SMARTWEB aims to develop a context-aware, mobile and multimodal interface to ontology servers, composed Web Services and open-domain question answering (QA) systems. In the main application scenario, the user carries a PDA and is able to pose multimodal questions about football games, teams, and players at a visit to the football World Cup in Germany — using speech, pen, and gesture as input modalities. The displays of these mobile devices are small (320*240 pixel for T-Mobile's MDA3 or 480*640 pixel for the MDA4), and the pocket computer has very limited computational power. Nonetheless, the user should be able to interact with the system in different modalities such as speech and gesture and refer to the displayed results for further inspection or posing a new query.

In SMARTWEB dialogue objectives and hence the dialogue reaction behaviour is governed by the general QA scenario, which means that almost all dialogue and system moves relate to questions, follow-up questions, clarifications, or answers. As these dialogue moves can be regarded as adjacency pairs, a standard dialogue behaves according to some finite-state grammar for QA, which makes a basic FSA appear reasonable for dialogue management. A finite state approach generally enhances robustness and portability and allows to demonstrate dialogue management capabilities even before more complex information states are available to be integrated into the reaction and presentation decision process. The paper is organised as follows: in section 2. the interaction requirements are discussed, followed by the general system architecture. In section 3. the reaction and presentation module design is introduced, how the FSA for QA looks like, what kind of ontology structures are used, and what kind of meta data can be made available for automatic adaptation. In section 4. we give concluding remarks.

2. Mobile Interaction Requirements

Interaction requirements are discussed in terms of reaction and presentation requirements to provide a basis for implementing a multimodal mobile human-computer-interface (HCI).

2.1. From Storyboard to HCI Implementation

Basically SMARTWEB allows the user to send multimodal requests to various services linked by a Semantic Web framework. The partners in the project share implementation experience from earlier multimodal intreaction projects like Verbmobil and SmartKom (Wahlster, 2000; Reithinger et al., 2003). Like others, we used some guidelines (Oviatt, 1999; Alexandersson et al., 2004) in the development of the storyboard and the specification of interaction possibilities.

The user should be able to

- ask simple factoid and enumeration questions, and inspection questions or commands (search, explore, inspect).
- control the system. She can ask for status information, or cancel a running query.

On the other hand, the system can take the initiative to

- clarify or cancel user requests.
- add and replace results.
- provide status information and hints.

Interesting decisions in dialogue management are concerned with these system initiatives.

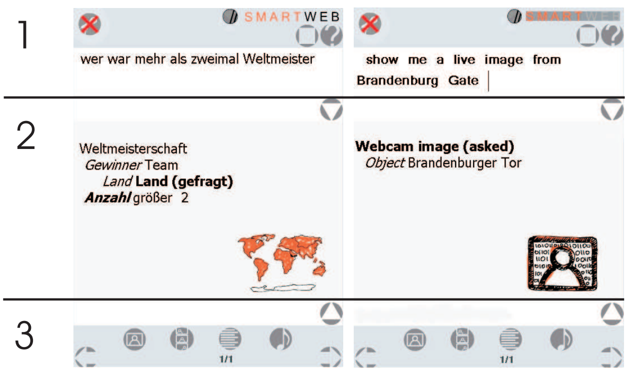

Figure 1: In (1) we display the output of the automatic speech recogniser, (2) shows the corresponding multimodal semantic paraphrase. The query paraphrase can also be listened to by *on ear* audio output. While audio repetition plays, barge-in is possible which leads to correction modes for single words or the complete query.

Paraphrasing a semantic query and displaying it to the user is one of the key elements for implicit user feedback (figure 1). The ontological structures resemble typed feature structures (TFS) (Carpenter, 1992) common in formal NLP. The paraphrase, which is presented to the user and sent to the Semantic Web in RDF representation, is constructed on the basis of the results of the question analysis. Ontology query instances are communicated between the dialogue server and the Semantic Web knowledge bases (figure 2). The nested predicate-argument structure shows the interpretation of the user utterance *Who won the football World Cup more than twice?* Note that the paraphrase is fully specified and contains the unfilled template slot *varName* for the winner name and the expected focus type (Team in text medium). Team itself is an underspecified concept which can be instantiated by a more specific instance according to the domain, e.g., a *FootballNationalTeam* instance. If the user completed his utterance by *Are pictures there?*, the *mediaTypes* slots would have changed to comprise text and image media. In this way we establish multimodal access to the Semantic Web.

```
[ discourse#Query
      text: "wer war mehr als zweimal Weltmeister"
      dialogueAct: [ InterrogQuestion ]
      focus: [ Focus
         focusMediumType:[ mpeg7#Text ]
         ...
         contextObject: [ FIFAWorldCup
                          winner:Team
                          origin: [ sumo#Country  ... ]
                        ]
         contextObject: [ GreaterThan
                          constraintRightArg: "2"
                        ]
         varName: ?X
         ]
]
```

Figure 2: Semantic queries serve as input to the Semantic Web knowledge bases.

2.2. Dialogue System Architecture

A flexible dialogue system platform is required to support audio transfer and other data connections between the mobile device and a remote dialogue server. We developed a new framework complementing other approaches (Cheyer and Martin, 2001; Herzog et al., 2004; Bontcheva et al., 2004) for Semantic Web based data structures for both dialogue system-internal and system-external communication. The dialogue system instantiates and sends requests to the so-called *Semantic Mediator*, which provides the umbrella for all different access methods to the Semantic Web we use: a knowledge server, a Web Service composer, semantically wrapped Web pages, and a QA system. To integrate the dialogue components we developed a Java-based hub-and-spoke architecture (Reithinger and Sonntag, 2005). The speech interpretation component (SPIN) (Engel, 2005), the modality fusion and discourse component (FADE) (Pfleger, 2005), the context-module SITCOM to resolve GPS coordinates, the natural language generation module (NIPSGEN), and the system reaction and presentation component (REAPR) are attached to it (figure 3). An exemplary data flow is $SPIN \rightarrow FADE \rightarrow REAPR \rightarrow SemanticMediator \rightarrow REAPR \rightarrow NIPSGEN$, which gets more complicated if, e.g., misinterpretations or clarifications are involved.

Having received a result list of multimodal items as answers to a question after query processing, we have to decide which responses are appropriate to be presented. The last point concerns content selection, medium selection, and selection of the visual presentation metaphors engaged. In this contribution we focus on the reaction behaviour, including the decisions of accepting a proposed semantic paraphrase we coded into a FSA structure. We put emphasis on the REAPR component in the remainder of the text.

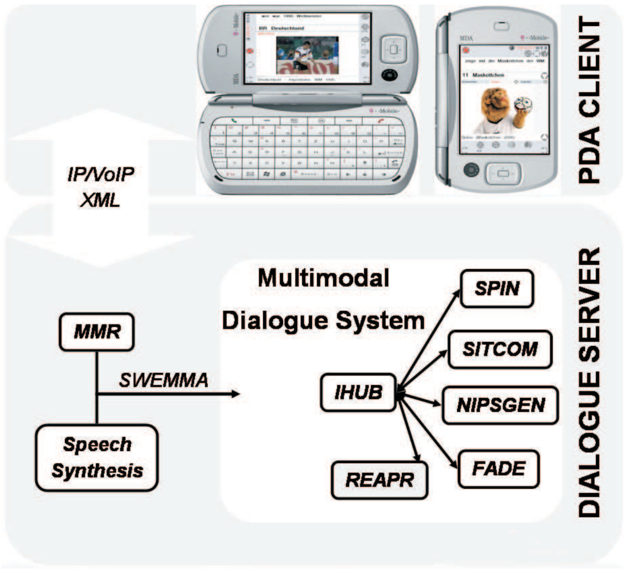

Figure 3: SMARTWEB's mobile dialogue system architecture: the PDA client and the dialogue server which comprises the dialogue manager. MMR stands for multimodal recogniser.

3. The Reaction and Presentation Module

REAPR manages dialogical interaction, i.e., the reaction and presentation behaviour, for the supported dialogue phenomena such as flexible turn-taking, incremental processing, and multimodal fission/fusion of system output. REAPR is based on the FSA shown in figure 4

3.1. General Discourse Obligations and Structures

1. The primary role to fulfill in information-providing dialogue systems is to elicit all relevant information from the user to pose a very specialised query for which getting the right answer is very probable. This role gains even more importance if the queries must be transformed into explicit semantic representations, i.e., ontological query instances.

2. Whenever users have the freedom to formulate statements, understanding may be difficult. In such cases the strategy is, for first, to produce useful reactions, and for second, to give hints or examples to the user on how to reformulate the question.

The general discourse obligations towards mixed and system initiative dialogue system behaviour are coded into the non-deterministic FSA structure, the multiple outgoing arcs at important input processing dialogue nodes, such as *query completion*.

3.2. User Correction Model

One important question in the user interaction model with respect to dialogue management decisions is how to correct invalid user input stemming from speech recognition errors or from errors that occur while interpreting user utterances. This becomes even more relevant in the context of composite multimodality, where the dialogue system must understand and represent the multimodal input.

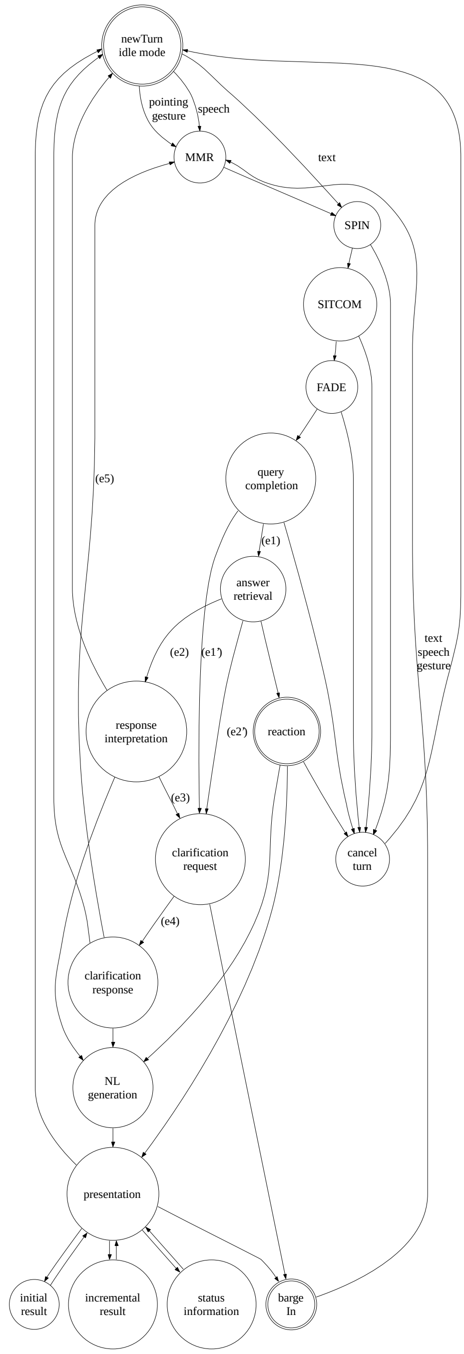

Figure 4: FSA structure as REAPR's ground control

When the system displays the query paraphrase, the user should have the possibility to interrupt the audio output and edit the query: A click on a word or word group directly enables the correction mode of the word(s), whereby the navigation through the displayed query is provided via keyboard or pen. Pen is preferred because it allows intuitive word selection on screen. For example, the user could simply click on or underline an incorrect word. The queries for a Semantic Web search have to be as accurate as possible, and correcting flawed speech recognition output is of paramount importance in the Semantic Web context. Practically this means a lot of manual corrections to be done by the user. In order to minimise the number of corrections to be done by the user, two system initiative strategies can be explored: To induce missing parts at the NLU stage and/or to apply an independent (binary) classificator for dialogue management to find out the underspecified queries with high or low success chances. Inducing missing parts at NLU is a language and domain-dependent task for the language and domain experts developing the NLU module. Applying a classificator depends on the suitability of meta data that can be extracted during dialogue processing. The design of REAPR is tailored toward an decision making process using automatic classifiers (section 3.2.).

FSA Structure as REAPR's Ground Control The FSA makes up the integral part of the dialogue managment decisions in the specific QA domain we model. The dialogue structure that is embedded and committed by the transitions of the FSA allows for a declarative control mechanism for reaction and presentation behaviour.

The initial node of the FSA is $N_{newTurn}$ which represents the system's idle mode while awaiting user input. The second starting node is $N_{reaction}$, the system takes initiative and informs the user or cancels the current turn. The third starting node is $N_{bargeIn}$ which is the user-initiative action as counterpart to system-initiative action. Every user action while processing a query can be seen as barge-in and is interpreted by the MMR component. This concerns the speech input, the selection of result words and sentences, and other gesture input such as poiting gestures on images, whereas new textual queries directly go to SPIN. The reason for that is simple, we do not fuse textual and image pointing gestures, since cross-modal fusion (speech and gesture) is much more convenient.

Starting with a new speech input, the example flow: $SPIN \rightarrow FADE \rightarrow REAPR \rightarrow SemanticMediator \rightarrow REAPR \rightarrow NIPSGEN$ can be mapped onto the actual FSA states: After $N_{queryCompletion}$ the answer is processed and retrieved $N_{answerRetrieval}$, generated $N_{NLgeneration}$, and presented $N_{presentation}$. However, we focus on the interesting part around the possibly cyclic paths containing $N_{clarificationRequest}$ (table 1). All cyclic paths containing $N_{clarificationRequest}$ are designed to elicit additional query information by feedback from the user to recover from query problems.

Path Name	Path Nodes
$P_{interpretation}$	$(e1) \cdot (e2) \cdot (e3) \cdot (e4) \cdot (e5)$
$P_{annotation}$	$(e1) \cdot (e2') \cdot (e4) \cdot (e5)$
$P_{induction}$	$(e1') \cdot (e4) \cdot (e5)$

Table 1: Cyclic paths containing $N_{clarificationRequest}$

$P_{interpretation}$ is the simplest QA dialogue processing path in which each retrieved answer is tested for semantic correctness (e.g. correct answer type). According to the meta data obtained from the answer status (figure 5), REAPR decides whether the answer is appropriate for presentation or not. An empty answer or false answer type initiates a system response, to either ask a different question or reformulate the question.

$P_{annotation}$ relies on the semantic annotation of the Semantic Web access. For example, the composed Web Service module can question missing parts matched to input descriptions of individual web services. In this case, the decision to pose a clarification question is dynamically deflected to the answering services, in this case the Web Services. The entity to be asked for is explicitly marked-up in the result obtained from the Web Services. The composed service module automatically sets a *ClarificationResponse* dialogue act instead of an *Answer* into the result structure.

$P_{induction}$ corresponds to the attempt to adapt dialogue processing towards recoverability decisions at a very early stage of processing. REAPR itself is responsible to decide if a query is valid and to be transferred to the Semantic Mediator. Although the right choice among the different strategies can be predicted by adaptable systems in many different ways, the $P_{induction}$ is a very interesting one: It is possible to infer patterns from the available meta data material, or abstractions from domain-ontological question-answer instances to judge a query as yet unsuitable, i.e., too underspecified, too specific, unsupported, untrusted, hence with little change of success. Effective selection and mining of ontological process data is a precondition and includes the question what kind of ontological meta data is available and suitable for feature spaces in machine learning environments (cf. section 3.3.).

The simple but effective FSA ground structure allows for adding new knowledge-driven functionality if necessary without the need for expensive training data following a fully empirical approach. On the other hand, tuning and adaptation can be easily integrated by casting the decision, which path in the undeterministic FSA to follow, into a classification problem to be solved by any suitable supervised machine learning classifier. Since the ontology structures are tailored toward semantically-rich information items, a unsupervised classification experiment can be complemented by association rule mining.

Ontological result structures can be seen in figure 5. The features obtained from the *AnswerStatus* resemble the features used in previous experiments to classify user models (Komatani et al., 2003) for dialogue system adaptivity. The feature *finishedSearchComponent* reveals the source of the obtained results. Additional scores for single utterances can be obtained from the recogniser (recogniser score), SPIN (speech interpretation score) and ratios of correct answering processes. The number of filled slots in the query and their names can also be added to the IS to extract patterns from. A semantic ontological result delivers meta data about the current dialogue state to be incorporated into the IS.

According to the undeterministic FSA, the dialogue management component has to decide on-the-fly whether a clarification dialogue is to be initiated, or a confirmation is needed, or the query is being sent without any confirmation. In this way we try to obtain optimised dialogue prompts in specific data situations. This optimisation toward more natural interaction should be obtained by mining the IS.

3.3. Information States for QA

Information state theory of dialogue modelling consists basically of a description of informal components (e.g.,

```
[ Result
    status: [ AnsweringStatus
           derivedFromQuery: discourse#Query
           elapsedTime:  "7"
           resAmount: "1"
           resForm: "nonincremental"
           finishedSearchComponent: "knowledgebase"
            ]
    ...
    content: [ FootballNationalTeam ... ]
    answerType: "FootballNationalTeam"
]
```

Figure 5: Semantic result: answer status, content, and answer type information

obligations, beliefs, desires, intentions) and their formal representation (Larsson and Traum, 2000). IS states as envisioned here do not declare update rules and an update strategy (for e.g. discourse obligations (Matheson et al., 2000)) because the data-driven approach is pattern-based, using directly observable processing features, which complements an explicit manual formulation of update rules. Since the dialogue ontology is a formal representation model for multimodal interaction, multimodal MPEG7 result representations (Sonntag and Romanelli, 2006), result presentations (Sonntag, 2005), dialogue state, and (agent) communication with the backend knowledge servers, large information spaces can be extracted from the ontological instances describing the system and user turns in terms of realised dialogue acts.

The turn number represents our first FSA extension to IS with the result of more flexibility to user replies. Replies which are not specified in a pathway, are not considered erroneous by default, since the IS now contains a new turn value. Ontologicial features for IS extraction under investigation are summarised in table 2.

Feature Class	IS State Features
MMR	*Listening, Recording, Barge-in, Last-ok, Input dominance (text or voice)*
NLU	*Confidence, Domain relevance*
Query	*Dialogue act, Focus medium, Complexity, Context object, Query text*
Fusion	*Fusion act, Co-reference resolution*
Answer	*Success, Speed, Answer streams, Status, Anser type, Content, Answer text*
Manager	*Turn/Task numbers, Idle states, Waiting for Results, User/system turn, Elapsed times: input/output, Dialogue act history (system and user) e.g. reject, accept, clarify*

Table 2: IS Feature Classes and Features

In previous existing work on dialogue management adaptations (Walker et al., 1998; Singh et al., 2002; Rieser et al., 2005), reinforcement learning was used for which large state spaces with more than about five non-binary features a hard to deal with. As seen in table 2, more than five relevant features can easily be declared. Since our optimisation problem can be formulated at very specific decisions in dialogue management due to the FSA ground control, less training material for larger feature extractions is to be expected. Relevance selection of ontology-based features is the next step for ontology-based dialogue managment adaptations.

Ontological Infrastructure SMARTWEB's ontological infrastructure is realised by merging concepts from two established foundational ontologies, DOLCE (Gangemi et al., 2002) and SUMO (Niles and Pease, 2001) into a new one (SWIntO) (Cimiano et al., 2004). Domain specific knowledge is modelled in sub-ontologies. SWIntO integrates question answering specific knowledge, interpretations of user utterances (modelled by the EMMA[1] extension SWEMMA), dialogue acts, and HCI concepts in a discourse ontology (DISCONTO). The DISCONTO also contains concepts for the communication between the *Dialogue Server* and the *Semantic Mediator*. The SWIntO and DISCONTO provide semantic representation structures for natural language understanding, generation, and dialogue management.

4. Concluding Remarks

We presented the interaction requirements and intermediate development steps of the reaction and presentation module REAPR for the second demonstrator of the SMARTWEB system[2]. The current dialogue model is FSA-based with user barge-in capabilities. SMARTWEB as multilingual (German and English), multimodal QA system was successfully demonstrated in the context of the football World Cup 2006 in Germany. The knowledge base (Swinto ontology) comprises 2308 concept classes, 1036 slots, and 90522 instances.

Semantic Web-based dialogue and data models are convenient models towards language-independence and multilinguality of HCI technologies. Operating on semantic ontological instances, knowledge-intensive processing modules within the dialogue system, such as REAPR, can be language-independent. Language-dependent modules (SPIN, FADE) operate on the same ontological instances, but exploit further language-dependent information provided by a multilingual lexicon model LingInfo (Buitelaar et al., 2005; Buitelaar et al., 2006). REAPR's FSA model works completely language-independent, for $P_{induction}$ our next step is to include linguistic features into the IS model. Currently, 19992 instances of the 90522 instances such as games, players, goals support linguistic information, and 6002 LingInfo instances have been created for German and English, which motivates machine learning experiments to incorporate linguistic, probably language-dependent, features into REAPR's IS dialogue model for adaptable dialogue management on top of a more structured but robust language-independent non-deterministic FSA-based dialogue management model.

5. Acknowledgments

The research presented here is sponsored by the German Ministry of Research and Technology (BMBF) under grant

[1]http://www.w3.org/TR/EMMAreqs
[2]http://www.smartweb-project.org

01IMD01A (SMARTWEB). We thank our student assistants and the project partners. Norbert Reithinger provided valuable comments on earlier versions. The responsibility for this papers lies with the author.

6. References

Jan Alexandersson, Tilman Becker, Ralf Engel, Markus Löckelt, Elsa Pecourt, Peter Poller, Norbert Pfleger, and Norbert Reithinger. 2004. Ends-based dialogue processing. In *Proceedings, Second International Workshop on Scalable Natural Language Understanding*, Boston.

Kalina Bontcheva, Valentin Tablan, Diana Maynard, and Hamish Cunningham. 2004. Evolving GATE to Meet New Challenges in Language Engineering. *Natural Language Engineering*, 10. Special issue on Software Architecture for Language Engineering.

Paul Buitelaar, Michael Sintek, and Malte Kiesel. 2005. Feature representation for cross-lingual, cross-media semantic web applications. In Thierry Declerck and Siegfried Handschuh, editors, *Workshop on Knowledge Markup and Semantic Annotation (SemAnnot2005)*, 11.

Paul Buitelaar, Thierry Declerck, Anette Frank, Stefania Racioppa, Malte Kiesel, Michael Sintek, Ralf Engel, Massimo Romanelli, Daniel Sonntag, Berenike Loos, Vanessa Micelli, Robert Porzel, and Philipp Cimiano. 2006. Linginfo: Design and applications of a model for the integration of linguistic information in ontologies. In *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC 2006)*, Genova, Italy, May 24–26.

B. Carpenter. 1992. The logic of typed feature structures.

Adam J. Cheyer and David L. Martin. 2001. The Open Agent Architecture. *Autonomous Agents and Multi-Agent Systems*, 4(1–2):143–148.

Shiu-Wah Chu, Ian ONeill, Philip Hanna, and Michael McTear. 2005. An approach to multi-strategy dialogue management. In *Proceedings of INTERSPEECH*, pages 865–868.

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.

Ralf Engel. 2005. Robust and efficient semantic parsing of free word order languages in spoken dialogue systems. In *Proceedings of 9th Conference on Speech Communication and technology*, Lisboa.

Dieter Fensel, James A. Hendler, Henry Lieberman, and Wolfgang Wahlster, editors. 2003. *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*. MIT Press.

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.

Gerd Herzog, Alassane Ndiaye, Stefan Merten, Heinz Kirchmann, Tilman Becker, and Peter Poller. 2004. Large-scale Software Integration for Spoken Language and Multimodal Dialog Systems. *Natural Language Engineering*, 10. Special issue on Software Architecture for Language Engineering.

Helmut Horacek and Magdalena Wolska. 2005. A hybrid model for tutorial dialogs. In *Proceedings of the 6th SIGdial Workshop on Discourse and Dialogue*.

Kazunori Komatani, Shinichi Ueno, Tatsuya Kawahara, and Hiroshi G. Okuno. 2003. Flexible guidance generation using user model in spoken dialogue systems. In Erhard Hinrichs and Dan Roth, editors, *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 256–263.

Staffan Larsson and David Traum. 2000. Information state and dialogue management in the TRINDI dialogue move engine toolkit. *Natural Language Engineering, Cambridge University Press*.

C. Matheson, M. Poesio, and D. Traum. 2000. Modelling grounding and discourse obligations using update rules. In *Proceedings of NAACL 2000*, May.

Michael F. McTear. 2002. Spoken dialogue technology: Enabling the conversational user interface. *ACM Computing Survey*, 34(1):90–169, March.

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

Sharon Oviatt. 1999. Ten myths of multimodal interaction. *Communications of the ACM*, 42(11):74–81.

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

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, Jan Alexandersson, Tilman Becker, Anselm Blocher, Ralf Engel, Markus Löckelt, Jochen Müller, Norbert Pfleger, Peter Poller, Michael Streit, and Valentin Tschernomas. 2003. SmartKom: Adaptive and Flexible Multimodal Access to Multiple Applications. In *Proc. of the 5th Int. Conf. on Multimodal Interfaces*, pages 101–108, Vancouver, Canada. ACM Press.

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.

Verena Rieser, Kruijff Kruijff-Korbayova, and Oliver Lemon. 2005. A framework for learning multimodal clarification strategies. In *Proceedings of the International Conference on Multimodal Interfaces (ICMI)*.

Satinder Singh, Diane Litman, Michael Kearns, and Marilyn Walker. 2002. Optimizing dialogue management with reinforcement learning: Experiments with the njfun system. *Journal of Artificial Intelligence Research (JAIR), Volume 16, pages 105-133*.

Daniel Sonntag and Massimo Romanelli. 2006. A multimodal result ontology for integrated semantic web dialogue applications. In *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC 2006)*, Genova, Italy, May 24–26.

Daniel Sonntag. 2005. Towards interaction ontologies for mobile devices accessing the semantic web - pattern languages for open domain information providing multimodal dialogue systems. In *Proceedings of the workshop on Artificial Intelligence in Mobile Systems (AIMS). 2005 at MobileHCI*, Salzburg.

Wolfgang Wahlster, editor. 2000. *VERBMOBIL: Foundations of Speech-to-Speech Translation*. Springer.

Wolfgang Wahlster. 2004. Smartweb: Mobile applications of the semantic web. In Peter Dadam and Manfred Reichert, editors, *GI Jahrestagung 2004*, pages 26–27. Springer.

M. Walker, J. Fromer, and S. Narayanan. 1998. Learning optimal dialogue strategies: A case study of a spoken dialogue agent for email.