Chapter 1

Introduction

Where is the life we have lost in living? Where is the knowledge we have lost in information? Where is the wisdom we have lost in knowledge? T. S. Eliot

1.1. Motivation—Why this book?

Dialogue System Research, Adaptable Question Answering, and Ontologies Toward Intelligent User Interfaces

Dialogue System Research

Dialogue systems, used to communicate with computers, are an important testbed for artificial intelligence and machine learning methods, and aim at natural communication with machines. Who would not like to speak freely to computers and ask questions which can be answered in real-time with the help of search engines on the World Wide Web or other information repositories? With increasing performance, the role of dialogue systems may shift from performance enhancers (e.g., voice input is fast and convenient on mobile devices) toward guides, educational tutors, or adaptable interfaces in ambient intelligence environments. Maybury and Wahlster (1998) define *intelligent user interfaces (IUIs)* as human-machine interfaces that aim to improve the efficiency, effectiveness, and naturalness of human-machine interaction. In order to realise these properties, explicit models of the discourse of the interaction, the available information material, the domain of interest, the task, and/or models of a user or user group have to be employed, at best in ontological form. How intelligent a user interface appears to the user depends on how appropriately the system chooses its actions. In information-seeking dialogue, these actions can be the response to questions, or feedback on queries, for example. In any case, error handling (preferably before errors occur) can be seen as a key factor to increasing general acceptance and usability of the dialogue-based interface. Consider the following example:

Imagine you are visiting a football match in Berlin and you take a mobile mini computer with you which is able to answer questions in real-time. If you ask *"Who was world champion in 1990?"* state-of-the-art QA systems for this specific domain with a natural language understanding component and access to an ontological knowledge base should be able to answer with great accuracy: *"That was Germany"*. Later, since you are new to the city, you are on a sightseeing tour. During the bus ride, you pass Castle Charlottenburg which arouses your curiosity: *"I wonder who might have built Castle Charlottenburg?"*

Unfortunately, at this point, most of the specific domain QA systems would return *"No Answer"* after checking the knowledge base where the answer cannot be found (a process which might also consume a lot of time). In this situation, the user would be very dissatisfied with the system. In order to enhance the range of possible questions to be answered, open domain access using a search engine could be realised as a fallback strategy. This would potentially enhance recall, but would on the other hand mean a loss of precision combined with problems of result provenance—the results are less reliable. Consider the fact that one of the first hits returned by a conventional search engine is often not necessarily a reliable source of information. Further problems occur if different information sources have different access characteristics, e.g., Web Services answering questions such as *"What's the weather like tomorrow?"* could be temporarily unavailable. This leads to a situation where an efficient, effective, and natural question answering dialogue is hard to achieve. Therefore, dialogue researchers should adopt two related but different goals. The first is to develop a theory which determines useful and cooperative dialogue extensions to the currently observed dialogue and behaviour of the participants. The second research goal and computational part is to develop algorithms that support and implement cooperative, and adaptive, dialogue in dialogue systems.

Adaptable Question Answering

A useful and cooperative question answering dialogue would combine different topics, heterogeneous information sources, and processing feedback. In information-seeking dialogue, feedback actions can, for instance, be status information on queries or results to be expected. In the following user-system dialogue example, adequate question feedback is shown (as well as realistic system response times):

---

1. **U:** "When was Brazil world champion?"
2. **S:** "In the following 5 years: 1958 (in Sweden), 1962 (in Chile), 1970 (in Mexico), 1994 (in USA), 2002 (in Japan)." (**6000 ms**)
3. **U:** "And England?"
4. **S:** "In 1966 (in England)." (**4200 ms**)

Later ...

5. **U:** "What can I do in my spare time on Saturday?"
6. **S:** "Sorry, services for such questions were unavailable a short while ago, shall I continue? This may take a minute or so ..." (**600 ms**)
7. **U:** "Ah okay, I can wait."
8. **S:** "Unfortunately, this service only produces empty results at the moment." (**52000 ms**)

Later on the bus ride ...

9. **U:** "Now I would like to know who built Castle Charlottenburg."
10. **S:** "It's not in my knowledge base. I will search the Internet for a suitable answer. Empty results are not expected, but the results won't be entirely certain." (**600 ms**)
11. **S:** "I think it was built by Johann Nehring and Martin Grünberg." (**10000 ms**)

Later, in front of the stadium ...

12. **U:** The French team is just getting off the bus. "Can I have pictures and names of all French players?"
13. **S:** "I predict fifty seconds for the search and download of the image content. The knowledge base will be queried immediately." (**600 ms**)

In order to model this dialogue[1], the system must be able to predict empty results, answer times, and classify queries for the probability of success according to query features and specific access/quality properties of the answer services in a changing environment. In the open and closed domain question answering application context, dialogue strategies can be too complex to be modelled in advance; environmental conditions such as availability and quality of information material can heavily influence the current design. Error handling (preferably before errors occur) through *thinking ahead* by using predictive ML models can be seen as a key factor in increasing general acceptance, usability, and naturalness of the dialogue-based interface.

---

[1]The dialogue derives from application potentials of Dialogue and Semantic Web Technology provided by Deutsche Telekom, available at *http://smartweb.dfki.de/SmartWeb_FlashDemo_eng_v09.exe* (scene three).

In the context of question answering, many research projects aim to enhance the user's satisfaction in dialogue systems with new forms of adaptivity management which complement multimodality and multilinguality.

Ontologies and Empirical Dialogue Models

Since the aforementioned existing dialogue models are not well suited for special meta level dialogue phenomena in question answering applications (for example, providing processing feedback as in the example dialogue), alternative dialogue models are needed.

As anticipated, one of the key reaction constraints in the question answering scenario is to inform the user about the probability of query success because both parties who participate in a (question answering) dialogue are responsible for sustaining it. Question answering can be defined as the task of finding answers to natural language questions, meaning that question answering systems do not retrieve documents (like information retrieval systems), but instead provide short, relevant answers in an interactive setting.[2] Obviously, there are joint commitments among dialogue participants—these motivate clarifications, confirmations, and feedback behaviour frequent in natural human-human communication, especially in information-seeking dialogue. Maybury (2003) proposed a roadmap for question answering, dealing with resources to develop or evaluate question answering, methods, and algorithms. Interactive/dialogue-based, multimodal, and constrained question answering (in terms of resources and solutions) are among the longer term objectives. Although some question answering systems exist which employ dialogue with advanced technical approaches (i.e., empirical, linguistic, and knowledge based), methodologies dealing with increasing system complexity and changing resource availability have yet to be developed.

According to the new requirements, we motivate the use of ontologies to guide the implementation of an adaptation manager for dialogue-based question answering. We present contributions concerning the representation of, reasoning with, and acquisition of dialogue manager models by exploiting ontology structures for ma-

[2] In multimedia/multimodal question answering, queries can be expressed by a range of media (e.g., text, audio, image, video) and/or modalities (e.g., auditory, haptic). Accordingly, the responses are in the ranges of these different media and modalities.

chine learning methods. An *introspective mechanism* producing machine learning models brings together theoretical and practical aspects of ontological knowledge representation for the system introspection and optimisation of dialogue reaction and presentation behaviour in dialogue-based question answering.

1.2. Who should read this book?

Target Audience

Those interested in interdisciplinary AI research and the application potential of semantic technologies for difficult AI tasks such as working dialogue and QA systems should read this book. This book should interest researchers and professionals, i.e.,

- *Information Retrieval* experts who plan to integrate semantics into databases and work on interfaces for semantic multimedia retrieval.
- *Machine Learning* experts who seek practical solutions to practical problems of model generation, such as the lack of supervised training material and the integration of learned models into a (dialogue) manager application. The learning aspect of a complex AI system is shown by the exploitation of the Semantic Web knowledge structure (i.e., ontological queries/answers and dialogue abox structures).
- *Dialogue System* experts interested in new forms of dialogue adaptivity and the formulation of meta dialogue; we also provide dialogue integration examples showing (1) how to predict answer times, (2) how to provide question feedback based on (ontological) metadata, and (3) how to learn to present incremental results from different (semantic) answer streams.
- *Semantic Web* experts interested in the potentials of answering engines that combine information from knowledge servers, Web Services, and open-domain QA. Meta dialogue allows us to mitigate the negative effect of different quality characteristics. We also address answer merging and result provenance aspects from a dialogue engineer's point of view.

The learning aspect of a complex AI system, i.e., an intelligent, dialogue-based user interface, is achieved by exploiting the Semantic Web knowledge structure with a combination of the abovementioned fields. In order to better understand the relationships among the scientific fields, the basics of natural language processing, ontologies, and machine learning are introduced as far as needed. The reader can learn a great deal about the fields by just reading through and/or consulting the references.

An extensive bibliography is included to allow for further study especially in the case when the provided ideas are too detailed for the uninformed reader. Therefore, this book can be used by research scientists, as well as by students and practitioners who are particularly interested in the interdisciplinary nature of the subject matter, the application potentials of semantic technologies, and introspective mechanisms. The material is also suitable for a two-semester graduate course.

Interdisciplinary Research

Dialogue-based question answering is challenging because it lies at the intersection of multiple scientific fields including dialogue systems (in terms of multimodality and semantic interaction design), natural language processing (discourse analysis, information extraction, and language generation), and information/knowledge retrieval (query formulation, knowledge representation, and relevance feedback). Figure 1.1 illustrates the relation of the scientific fields and their intersection. As can be seen, adaptable dialogue-based QA includes even more scientific fields; it lies at the intersection of psychology (metacognition as monitoring and control theory), machine learning and data mining (dynamic model creation), and graphical user interfaces (reinforcement signal, feedback environment), as illustrated on a second level. Very often the requirements for intelligent interaction as a multimodal, tailored, cooperative, and mixed-initiative process cannot be put into a single dialogue model easily—increasing performance comes along with increased complexity and the need to control complex dialogue systems as in the case of adaptable dialogue-based question answering. The learning environment set by the demand to capture natural dialogue phenomena and environmental conditions reveals additional challenges, e.g., the allocation of large volumes of training data.

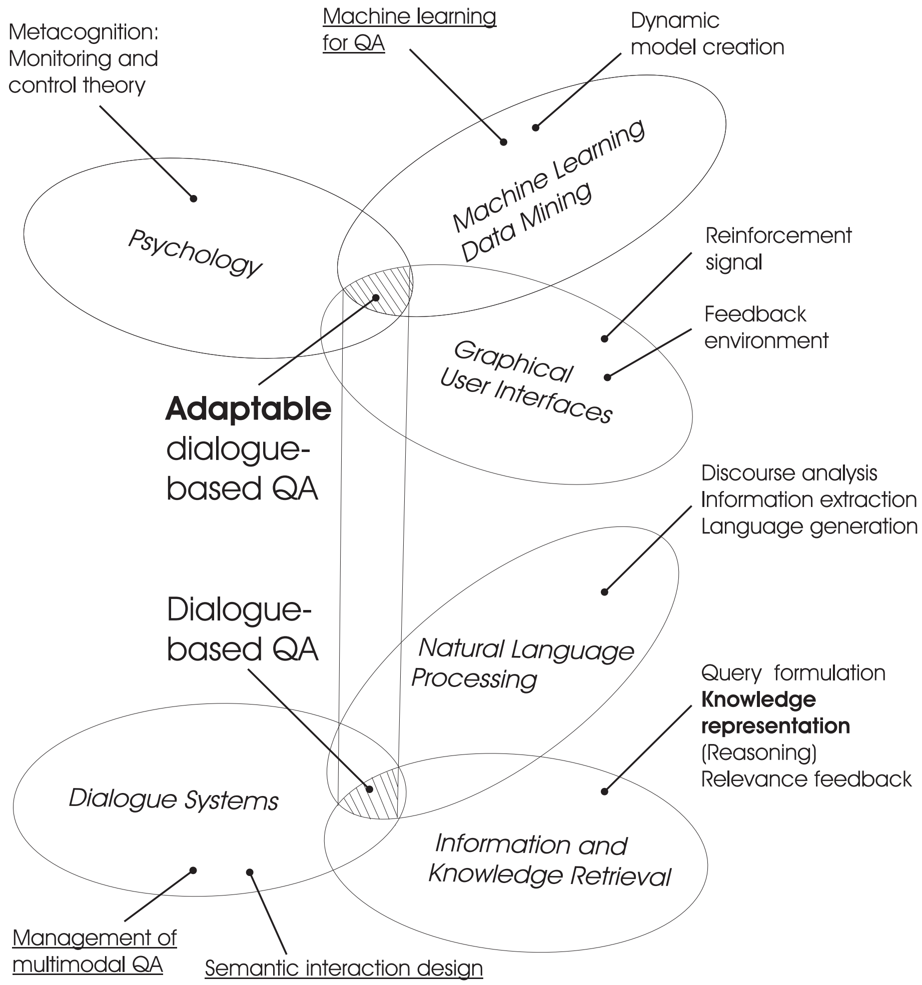

Figure 1.1. Disciplines for (Adaptable) Dialogue-Based Question Answering. Adaptation possibilities can be included by the additional integration of methods from psychology, machine learning and data mining, and graphical user interfaces.

Goals

Written from a computer science perspective, we provide the reader with state-of-the-art introductions to the representation and reasoning models of ontology-based dialogue processing (Part 1 and 2), the acquisition of adaptable dialogue models using ML, along with the selection and description of the most suitable methods (Part 2). We present contributions concerning the implementation of introspective machine learning models. We focus on the representation of, acquisition of, and reasoning with machine learning models. We will describe research on new forms of adaptivity for question answer based multimodal dialogue systems and formulate experiments to show that (semi-) automatic acquisition of optimised dialogical interaction behaviour for the *Management of Multimodal Question Answering* can be put into practice by mining dialogue processing structures and generating predictive models. These will be obtained by running dialogue sessions. Unlike traditional Wizard-of-Oz experiments where the expert needs to manually build appropriate modelling rules, we will try to automate the model creation and the integration process. The predictive models should provide a kind of *think-ahead* functionality to obey dialogue reaction and presentation constraints.

In particular, we address the following problems:

- *Data Annotation by integrating Semantic Technologies*:
  Knowledge management in dialogue systems often relies on primitive data structures. We aim to employ ontologies as rich knowledge representations. Ontology structures could be used not only for modelling domain and world knowledge, but also for extracting decision-relevant features. Decision-relevant features can in turn be used as new input spaces for machine learning experiments to induce domain rules and classify QA dialogue situations.
- *Dialogue Modelling by integrating Semantic Technologies*:
  We will propose a model for reaction and presentation decisions in terms of an adaptation manager that is able to dynamically integrate knowledge obtained from machine learning models. Thereby, data mining algorithms are seen as the instrument in this specific embedded machine learning environment. Data mining will be used for the purpose of obtaining decision models in manual and automatic model integration settings. The

adaptation is meant to increase the usability and robustness of question answering dialogue applications. The main objective is to find out which system-initiative question feedback can be given to the user. In this regard, the introspective mechanism should include a selection of suitable machine learning tasks and evaluation measures, and pinpoint a practical methodology. Thereby, the current system abilities play a major role, answering questions such as "*How can a dialogue system know what it cannot do at the moment?*"
- *Performance Enhancement over time*:

  We try to give a practical example which proves that "*Learning the behaviour online, dynamically, and automatically*" can be achieved for dialogue-based QA. *Automatic Model Operationalisation* meets the challenge how automatic dialogue adaptation can be executed and controlled. Automatic model creation and operationalisation will be implemented by a new methodology based on an introspective control and a control structure that injects the learned meta models into a dialogue manager.

The *introspective mechanism for dialogue-based question answering systems* is a theoretical framework and computational model of dialogue-based reaction and presentation decisions. The theory to be developed should make machine learning models possible to empower self-reflective dialogue systems with learning capacity. Introspective artificial intelligence systems dealing with adaptable dialogue can be implemented according to the model. We hypothesise that introspective system abilities correlate with more intelligent natural language processing systems; our aim is to assess the scope of the proposed mechanism's usefulness.

Relevance to Semantic Web Studies

The future of the Semantic Web relies heavily on the importance of explicit knowledge representation for complex AI systems. This book demonstrates that dialogue system research and knowledge representation have common research goals and implementations:

- the dialogue model's building blocks: ontological discourse, content representation, and dialogue frameworks based on semantic message transfer;

- the semantic integration framework and Semantic Web knowledge structures implemented in the context of a Semantic Web project (Some SmartWeb content is not published elsewhere in this combination or as an integrated application system.);
- the explicit knowledge representation structures (i.e., ontologies for knowledge management) in dialogue system management. International standards are introduced because we will use them to provide standardised and modularised interfaces for individual dialogue modules, multimedia representation structures, and the input to machine learning algorithms;
- the updates to the knowledge about the current system abilities by using data mining which answers the question *"How can a dialogue system know what it cannot do at the moment and act accordingly?"* whereby further applications include ontology evaluation, adaptivity in hypermedia generation, semantic mediation, and semantic navigation;

We stress that, apart from the research goals and implementations mentioned above, there remain other open research questions to be extensively investigated in the context of Semantic Web Studies.

1.3. SmartWeb System

SmartWeb aims to provide intuitive multimodal access to a rich selection of Web-based information services. In one of the main scenarios, the user interacts with a smartphone client interface to access the Semantic Web. The user is engaged in a natural language dialogue and can ask closed and open domain questions, for example, *"Who was world champion in 1954?"* (football domain), and *"Who invented the radio?"* (open domain), respectively.

An advanced ontology-based representation of facts and media structures serves as the central description for rich media content. Underlying content is accessed through conventional Web Service middleware to connect the ontological knowledge base and an intelligent Web Service composition module for external Web Services. These are able to translate between ordinary XML-based data structures and explicit semantic representations for user queries and system responses. The presentation module renders the media content and the results generated from the services, and provides a detailed description of the content and its layout to the fusion module. The user

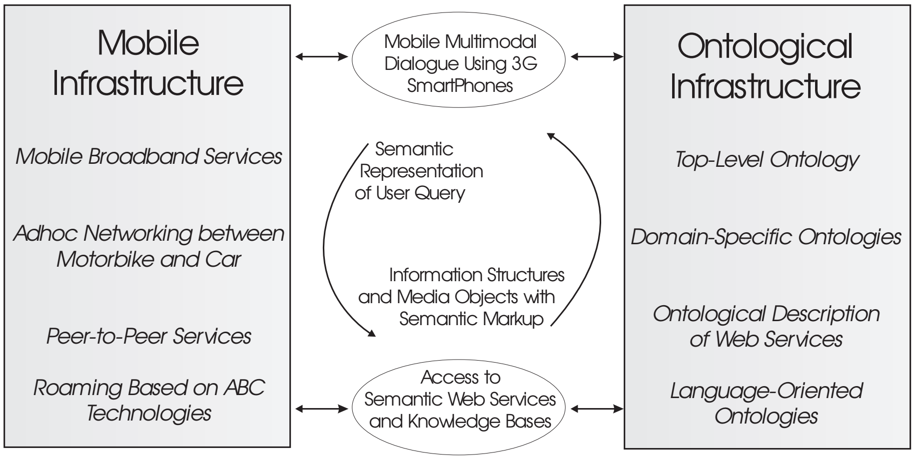

Figure 1.2. The Basic Architecture of SmartWeb

is then able to employ multiple modalities, like speech and gestures, to interact with the presented multimedia material in a multimodal way (Reithinger *et al.*, 2005; Sonntag *et al.*, 2007; Wahlster, 2007a). Figure 1.2 illustrates the basic architecture of SmartWeb. Technical realisations within the SmartWeb system, such as the dialogue framework, ontological representation structures, and the reaction and presentation module, constitute a considerable part of this book. The SmartWeb system is also used to generate the data sets used in the evaluation of the introspective methods.

Multimodal Mobile Interface to the Semantic Web

The development of a context-aware, multimodal mobile interface to the Semantic Web (Fensel, Hendler, Lieberman and Wahlster, 2003), i.e., ontologies and Web Services, is a very interesting task since it combines many state-of-the-art technologies such as ontology development, distributed dialogue systems, and standardised interface descriptions (cf. section 2.8). In our main scenario, the user carries a smartphone PDA, as shown in figure 1.3, and poses closed and open domain multimodal questions in the context of football games and a visit to a football Worldcup stadium. The PDA serves as an easy-to-use user interaction device which can be queried by natural language speech or handwriting, and which can understand social signalling—hand gestures on the PDA touchscreen and head move-

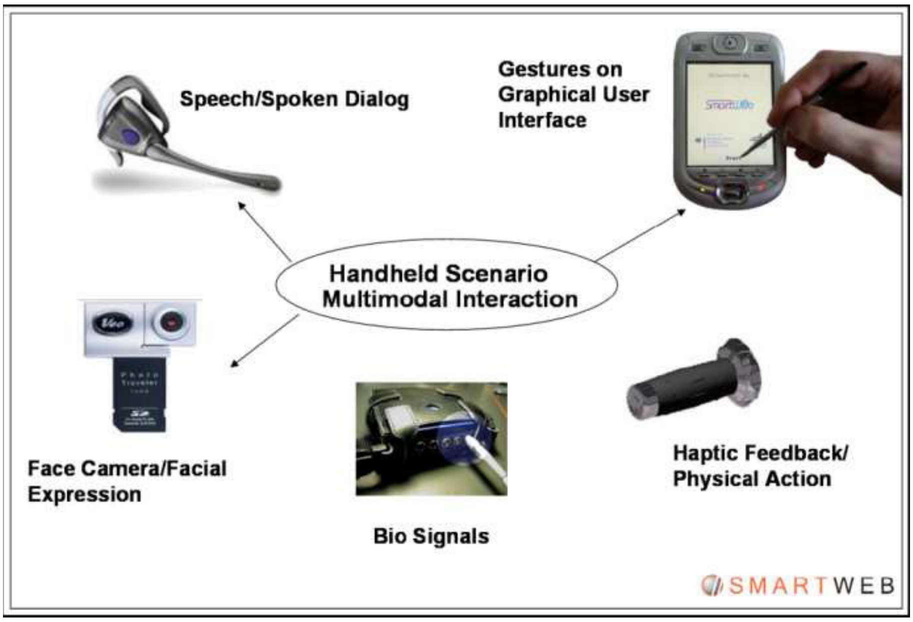

Figure 1.3. The multimodal dialogue handheld scenario comprises spoken dialogue recorded by a bluetooth micro, gestures on the graphical PDA touchscreen, and camera signals. In addition, the SmartWeb project uses the recognition of user states in biosignals to adapt system output in stressed car driving situations and haptic input from a force-feedback device installed on a motorbike.

ment as perceived by the PDA camera. With our multimodal dialogue interface we aim to provide natural interaction for human users in the Human Computing paradigm (Pantic, Pentland, Nijholt and Huang, 2006). One of the main contributions is the ontology-based integration of verbal and non-verbal system input (fusion) and output (system reaction). System-initiative clarification requests and other pro-active or mixed-initiative system behaviour are representative for emerging multimodal and embedded HCI systems. Challenges for the evaluation of emerging Human Computing applications (Poppe and Rienks, 2007) trace back to challenges in multimodal dialogue processing, such as error-prone perception and integration of multimodal input channels (Oviatt, 2003; Wasinger and Wahlster, 2006; Wahlster, 2003b). Ontology-based integration of verbal and non-verbal system input and output can be seen as groundwork for robust processing of multimodal user input.

Architecture Approach

The architecture consists of three basic processing blocks: the mobile PDA client, the dialogue server which comprises the dialogue manager, and the Semantic Web access system (cf. figure 1.4). The dialogue server system platform instantiates one dialogue server for each call and connects with the multimodal recogniser for speech and gesture recognition. The dialogue system sends the requests to the *Semantic Mediator* which provides the umbrella for all different access methods to the Semantic Web we use. It consists of an open domain QA system, a Semantic Web service composer, Semantic Web pages (wrapped by semantic agents), and a knowledge server. On the PDA client, a local Java-based control unit takes care of all input and output communication; the control unit is connected to the GUI controller. The local VoiceXML-based dialogue system runs on the PDA to allow for user interactions during link downtimes. The dialogue system consists of different, self-contained processing components. To integrate them we used the Java-based hub-and-spoke architecture (iHUB) (cf. section 3.1). The most important processing modules connected to the iHUB in the dialogue system are: a speech interpretation component (SPIN), a modality fusion and discourse component (FADE), a system reaction and presentation component (REAPR), and a natural language generation module (NIPSGEN). An EMMA Unpacker/Packer (EUP) component provides the communication with the dialogue server and Semantic Web subsystem external to the multimodal dialogue manager and communicates with the other modules of the dialogue server, the multimodal recogniser, and the speech synthesis system.

Dialogue Management

The integral part of dialogue management is the reaction and presentation module (REAPR) managing the dialogical interaction for the supported dialogue phenomena such as flexible turn-taking, incremental processing, and multimedia presentations of system output as explained in section 4.2.2. Our new approach differs from other IS approaches (e.g., Matheson, Poesio and Traum (2000)) by generating information state features from the ontological instances generated during dialogue processing (Sonntag, 2006), according to the introspective mechanism.[3]

It is important to mention that dialogue reaction behaviour within SmartWeb 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, the dialogue behaves according to some finite-state grammar for QA which makes up the automaton part (FSA) in REAPR. The finite-state approach enhances robustness and portability; accessorily it allows dialogue management to demonstrate capabilities even before the more complex IS states are available for integration into the reaction and presentation decision process in terms of, e.g., learned models for classification decision points, as demonstrated in section 7.6 by confirming the *Mining hypothesis*. The step to integrate the models obtained from the data providers is straightforwardly fulfilled by using information states and record states that provide the features for an introspective mechanism (described in section 6.3, also cf. figure 6.2 of the information state contributors).

[3] Information states are traditionally divided into global and local variables which make up the knowledge state at a given point of time. We also used a global and local representation. In addition, ontological structures that change over time vastly enhance the representation capabilities of dialogue management structures, or other structures like queries from which relevant features can be extracted.

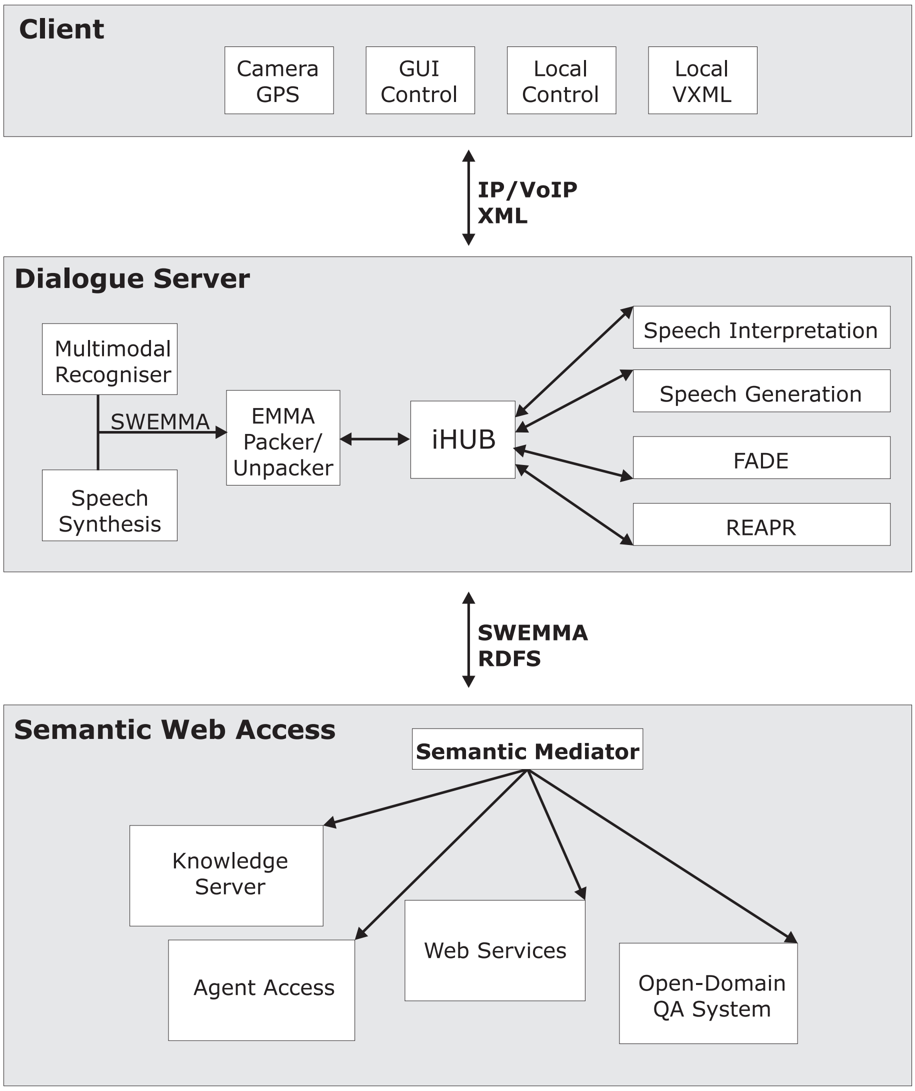

Figure 1.4. SmartWeb's mobile dialogue system architecture. The dialogue server includes the dialogue manager REAPR.

In SmartWeb, the responsibility for meaningful metadata, such as confidence values, lies with the delivering components. All information providers (DFKI-KM[4] for the Ontobroker access; EML[5] for the agent-based semantic access; DFKI-LT[6] for open-domain question answering; AIFB[7]&DFKI-KM[8]&DFKI-IUI[9] for the semantic-web-service access component; AIFB&DFKI-LT for semi-structured data access; and AIFB&DFKI-LT for the offline information extraction from free text) have been asked for delivering confidences and the explanation from what these confidences developed (confer sub-values). The empirical data of the data characteristics and data performances of information-providing modules (cf. chapter 7, in particular section 7.5) have been generated partly by the confidences and the explanations of the delivering components mentioned above. Metadata can be mined at the dialogue modeller's preference for reaction and presentation decisions as demonstrated in the evaluation. With the help of the generated models, the SmartWeb system could be tuned to detect and communicate uncertainties in the (open-domain) QA results. We will demonstrate how this knowledge can be turned into action rules in order to provide processing feedback and explanations. Especially in the context of mobile handheld applications, the source of presented information will play an increasing role.

[4] *http://www.dfki.de/web/research/km/*
[5] *http://www.eml-development.de/english/*
[6] *http://www.dfki.de/lt/*
[7] *http://www.aifb.uni-karlsruhe.de/*
[8] *http://www.dfki.de/km/*
[9] *http://www.dfki.de/iui/*

1.4. Roadmap

1. After introducing the scientific background (chapter 2), we will present a suitable integration framework for ontology-based dialogue processing (chapter 3) and dialogue management (chapter 4).

   We will focus on how to:

   - semantically model flexibility and functionality inherent to human-human communication;
   - automatically optimise conversational capability for question answering.

   We hypothesise that simple machine learning models could be employed on the meta-level that reason about the environmental input of a conversational agent. The optimisation is measured by the ML performance measures of the learning tasks. A learning task decides whether or not a suitable dialogue move can be initiated.
2. For the implementation of ontology-driven (semi-) automatic adaptation for dialogue systems, we provide the methods for system introspection and the optimisation of dialogue reaction and presentation behaviour (chapter 5).
3. What follows is the implementation of an adaptation manager model which induces relevant feature spaces and process logs for machine learning, data mining, and user feedback. This model guarantees adaptable dialogue management capabilities (chapter 6).
4. Finally, we show the applicability of the approach in the context of question answering scenarios (chapters 7 and 8).

Every chapter starts with some introductory words about the topics covered, and ends with related work where appropriate and a summary. The reader can survey any chapter by reading just the first and last section.