SmartWeb – Mobile Broadband Access to the Semantic Web

Norbert Reithinger, Gerd Herzog, Anselm Blocher

SmartWeb aims to provide intuitive multimodal access to a rich selection of web-based information services. Available system prototypes include a smartphone client interface to the semantic web, an advanced motorbike human machine interface, and an automobile demonstrator with an on-board dialogue system, that is updated automatically with new content from the web using broadcast technology. An advanced ontology-based representation of facts and media structures serves as central description for rich media content. Underlying content is derived from existing web pages and provided by conventional web services. Semantic annotations are being generated automatically using different processing techniques. In addition, a dedicated knowledge-based service composition module is being employed to translate between ordinary XML-based data structures and explicit semantic representations for user queries and system responses. The different variants of the SmartWeb client enable the user to employ multiple modalities, like speech and gestures, to interact with the presented multimedia material.

1 Introduction

Recent progress in mobile broadband communication and web technology is enabling innovative Internet services that provide advanced features to access knowledge and services on the web. The goal of the SmartWeb project [1], funded by the German Ministry for Education and Research from 2004 to 2007, is to lay the foundations for multimodal user interfaces to distributed and composable semantic web services on various mobile devices. The SmartWeb consortium members—BMW, Daimler-Chrysler, Deutsche Telekom, EML, Ontoprise, Siemens, Sympalog, DFKI, FhG FIRST, ICSI, and university groups from Bremen, Erlangen–Nuremberg, Karlsruhe, Munich, Saarbrücken, and Stuttgart—bring together experts from various research communities: Mobile services, intelligent user interfaces, language and speech technology, information extraction, and semantic web technologies.

The appeal of being able to ask a question to a mobile Internet terminal and receive an answer immediately has been renewed by the broad availability of information on the web. Ideally, a spoken dialogue system that uses the web as its knowledge base would be able to answer a broad range of questions. Practically, the size and dynamic nature of the web and the fact that the content of most web pages is encoded in plain text and graphics makes this an extremely difficult task. Machine-understandable content, as to be provide by the so called *Semantic Web* [3] has an immediate appeal to be used for intelligent question-answering. Since semantically annotated web pages are still very rare, due to the time-consuming and costly manual markup, SmartWeb is also using advanced language technology and information extraction methods for the automatic annotation of traditional web pages encoded in HTML or XML as well as included images and graphics.

SmartWeb provides a context-aware user interface to scenarios, in order to be able to support the user in different roles, e.g., as a driver of a car, a motorcyclist, a pedestrian, or a sports spectator. The core scenario that influenced system development was the information needs of mobile visitors to the 2006 FIFA World Cup in Germany. To cover this application scenario, the system needs detailed knowledge concerning past and current tournaments, information about points of interests in the tournament cities, general touristic knowledge, and of course information about topics of general interests. The user receives information provided as texts, pictures, and videos, depending on available sources.

[1] `www.smartweb-project.org`

2 General Approach

Today's commercial dialogue systems are typically realised by means of dialogue scripting using VoiceXML and similar frameworks. Hence they are not yet able to support a context-aware semantic processing of user contributions. Novel features like multimodality or mixed initiative conversation can only be found in current research prototype systems, like SmartKom or TALK [12, 1]. Comparable research activities dealing with the realisation of speech-enabled user interfaces are for example Ritel and Imix, which focus on interactive question-answering. The Ritel project aims at integrating spoken language dialogue and open-domain information retrieval to allow the user to ask general questions and to refine previous requests interactively [9]. The IMIX system, a multimodal dialogue system providing access to a question-answering system, focusses on answering non-factoid questions in a large domain [6].

Key to the SmartWeb approach to multimodality in a question-answering context are interoperating ontologies and a common data model for all knowledge-aware system modules. SmartWeb is based on W3C standards for the Semantic Web, like Resource Description Framework (RDF/S) and Web Ontology Language (OWL) for representing machine interpretable content on the web. OWL-S ontologies support semantic service descriptions, focusing primarily on the formal specification of inputs, outputs, preconditions, and effects of web services.

The ontological infrastructure of SmartWeb relies on SWIntO (**S**mart**W**eb **Int**egrated **O**ntology). It is based upon an upper model ontology, which is realised by merging well chosen concepts from two established foundational ontologies, namely DOLCE [4] and SUMO [5], into our SmartWeb foundational ontology SmartSUMO [2]. Domain specific knowledge (sportevents, navigation, web services) is defined in dedicated ontologies modeled as sub-ontologies of the SmartSUMO. In addition, a specific discourse ontology (DiscOnto) provides question-answering related concepts within SWIntO and a media ontology (SmartMedia) provides representation constructs for multimodal information. Processing within SmartWeb makes use of the W3C standard EMMA (Extensible MultiModal Annotation markup language)[2] for multimodal dialogue management, that we extended for handling result feedback. Some of these extensions will be adopted for the upcoming revision of the EMMA standard. Data exchange between system components at run-time is based on RDF format.

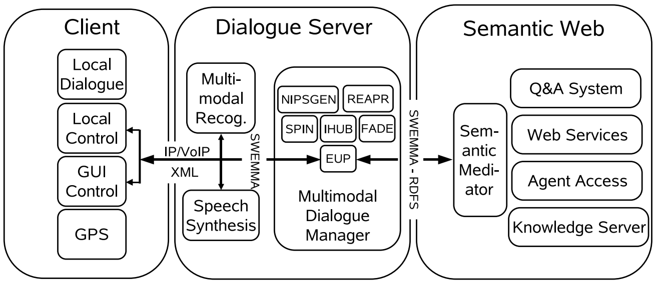

Figure 1: Overview of the architecture

Figure 1 provides a high-level overview of the main building blocks of the SmartWeb system. It comprises three parts: A scenario-specific client controller that runs on the end-user device, a dialogue server, and the server-side access components for semantic web services. Local dialogue and GUI capabilities differ from scenario to scenario. These variants will be described later on.

All three scenarios share a common client-side control component that connects the mobile application system to server-side components. Multimodal input, e.g., speech, haptics, or pen input, is recorded and transmitted to the back-end server using UMTS or wireless LAN. Components for multimodal recognition, dialogue management, and sub-systems for semantic web access are located on the server hosts. SmartWeb is designed for multi-session operation and a single SmartWeb server instance can handle multiple concurrent users. The response to a given question is delivered back to the mobile device for presentation of obtained results. Distributed dialogue processing relies on a server-side text-to-speech component for the generation of accompanying speech output. Sympalog Voice Solutions contributed their commercial speech dialogue system platform, which is being adapted and extended for multimodal interaction.

3 Multimodal Dialogue Manager

The dialogue system consist of different, self-contained processing components (see [8] for more details). To integrate dialogue sub-modules, we developed a Java-based hub-and-spoke architecture called iHUB. The main modules of the dialogue manager are connected via the iHUB, which also serves as ontology server. An EMMA Unpacker/Packer (EUP) component provides communication links to Sympalog platform and semantic web subsystem, which are external to the multimodal dialogue manager, and EUP communicates with the other modules of the dialogue server, i.e., multimodal recognizer, and speech synthesis system.

Input from (multimodal) recognition is first passed to the parsing module, which is based on the semantic parser SPIN. The parser lacks a syntactic analysis but builds up instances of the SWIntO ontology directly from word level. The outstanding feature of the parser is the possibility for order-independent matching, i.e., the order of elements in the input stream is ignored if order-independent matching is active. This simplifies the processing of free-word order languages like German and provides an increased robustness against speech recognition errors and disfluencies produced by the user.

Instantiated ontology expressions are then interpreted with respect to their discourse context by a the fusion and discourse engine (FADE). The task of FADE is to integrate verbal and nonverbal user contributions into a coherent multimodal representation to be enriched by contextual information, e.g., resolution of referring and elliptical expressions. The result is a contextually enriched representation of the user's utterance.

Core dialogue management is the task of the reaction and presentation module (REAPR). It manages the dialogical interaction for the supported dialogue phenomena such as flexible turn-taking, incremental processing, and multimodal fission of system output. REAPR is based on a finite-state-automaton and information space. Our new approach differs from other information space approaches (e.g. [11]) by generating features from the ontological instances generated during dialogue processing.

The text generation module uses the framework of the language understanding module in reverse order (NIPSGEN) together with a Tree Adjoining Grammar that takes care of the syntactic generation. The input of the generation module are instances of the SWIntO ontology that represent the search results. These results have to be verbalised in different ways, e.g., as heading, as a row of a table, or as text for speech synthesis.

4 Accessing the Semantic Web

After semantic analysis of the given user input, the dialogue server will forward the fully specified ontological representation of the request to the semantic services, which provide the umbrella for all different access methods to the semantic web that the system supports. These distinct parts include an open domain Q&A system, specialized web services, semantic web pages and a knowledge server.

The Semantic Mediator encapsulates all these services and provides a coherent interface to the background processes. Depending on the content of the query from the dialogue server, it preselects preferred access paths. If, for example, the system could not provide a deep semantic representation, chances are high that the request is in the open domain. Instead of sending the query to all processes, only the Q&A system is activated and the results are sent back. In the cases where more than one service is accessed, the semantic mediator also merges and scores the results to provide a better answer to the user.

For historic world cup information, it relies on an ontology-based knowledge server that stores the factual knowledge con-

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

Figure 2: SmartWeb Handheld system

Figure 3: SmartWeb Motorbike: On-board dialogue and server connection

Figure 4: SmartWeb Automobile system

cerning FIFA tournaments. These annotated data are derived automatically from the relevant web sites. An agent-based semantic wrapper component accesses web sites that provide reports of current games, which are not yet part of the knowledge base. A multitude of professional web services, provided by our project partner Deutsche Telekom, offer rich information concerning maps, routes, point-of-interests, and up-to-date information like weather data, traffic situation, cinema programs, etc. Multimedia question-answering (Q&A) is responsible for open-domain questions to provide answers and images concerning persons, locations, and others named entities. This subsystem, which searches and analyses web pages on the fly, complements the other access components, which focus on specific application domains.

5 The SmartWeb Scenarios

In the SmartWeb Handheld scenario, the user carries a smartphone as interaction device (see fig. 2). Currently, we use smartphones with Microsoft Windows Mobile, like T-Mobile MDA III and MDA Pro. The user can enter multimodal questions related to the different application contexts mentioned above. Speech is recorded using the built in microphone. Using speech and pen gestures, one can ask for information about World Cup facts, weather information and other services availlable through external web services, semantic web pages, and the plain web. The user can navigate through the given result using speech and gestures, and is also able to employ multimodality to browse through alternative results.

In the example, the user requested webcam pictures from Potsdamer Platz. The GPS signal tells the system that the user is located in Berlin and four web camera locations have been found. The user can navigate with the pen by clicking on the navigation icons at the bottom.

The SmartWeb Motorbike scenario provides a novel interaction interface for motorbikers. They are used to navigation and other services form their car environment. The motorbike client realizes first experimental steps with corresponding mobility-related services. Figure 3 presents the two GUI modes of the system. On top, the local human-machine interface is shown, where the driver can control on-board functions like radio, telephone and navigation system. The local dialogue system allows multimodal interaction. Speech control is based on a small, embedded Java VoiceXML interpreter which can be addressed through an on-board speech recognition engine for command-and-control utterances. The local dialogue system is connected to the audio gear in the driver's helmet via a Bluetooth link. The user can also use a rotating force-feedback element in the handle bar. Located in the left section of the screen, the GUI resembles a dial that the user can rotate. The menu entries on the right hand side are attached to the wheel. As direct manipulation using the force-feedback element can be tedious, the user can use speech directly to jump between menu entries. The local system is also connected to vehicle sensors and informs the driver about events, like low fuel, using multimodal output including speech, sounds, and graphical icons. Via a suitable data link (3G or WLAN), the local system can also be connected to a remote SmartWeb server. The content being displayed and operation of the server-side system is adapted to the specific needs and restrictions of a motorcyclist. The bottom screenshot shown in Figure 3 presents a more detailed display while the motorbike is standing. The driver can use the force-feedback element through a special haptic profile to navigate the interface similar to the handheld system. While driving, most of the available information is not displayed in order not to distract the driver.

SmartWeb Automobile is a new kind of mobile information system that interacts with drivers in natural language (see fig. 4). While you drive, the system searches the Internet for useful information and distributes data efficiently to all vehicles via radio signals based on the new Digital Multimedia Broadcasting (DMB) standard. Typical use cases from the application scenario include for example to ask for the cheapest gas station in Dortmund, current speed camera locations, or how many goals Schalke 04 has scored in the recent match.

Currently, we are in the process of integrating the server-support SmartWeb infrastructure into the vehicle as well. Hence, the driver will also have access to extended online services of SmartWeb. Of course, care needs to be taken to respect all safety relevant restrictions while driving, as it is the fact for the motorbike system.

Typical example input sentences—with the respective English translation below—are

- *"Wie ist das Finale 1954 ausgegangen"*
  (who won the 1954 final)
- *"Wo fand das Spiel Deutschland gegen Argentinien 2006 statt"*

(where did the 2006 match between Germany and Argentina take place)
- *"zeige mir das Tor im Spiel Frankreich gegen Senegal"* (show me the goal from the match between France and Senegal)
- *"was kommt heute abend in Saarbrücken im Kino"* (what is playing at the cinemas in Saarbrücken tonight)
- *"in welchem Kino läuft dieser Film"* (where is this movie playing)
- *"gibt es Staus zwischen Karlsruhe und Berlin"* (are there any traffic jams between Karlsruhe and Berlin)
- *"wo sind hier ↗ Restaurants"*[3] (are there some restaurants here ↗)
- *"zeige mir das Brandenburger Tor'"* (show me the Brandenburg gate)

This small selection from our current set of about 450 generalized query patterns depicts the wide range of input the system is able to process.

6 Summary

In this article we presented a short overview of the SmartWeb system and its realisation in the three end-user scenarios Handheld, Motorbike and Automobile. SmartWeb addresses the whole range of issues that must be solved to leverage the potential power of the semantic web for new information devices. The multimodal context-aware user-interface provides a smooth and seamless access to heterogeneous, rich information of the semantic and classic web, as well as to web services and dedicated knowledge sources. Interoperating ontologies and a common data model are the fundament for all knowledge-aware system modules and the server-based dialogue processing. The central ontology is available at `www.smartweb-project.org/ontology.html`. Current work deals with the integration of more contextual information and additional modalities like face-detection for on-/offtalk discrimination. The system was successfully presented on various occasions over the last year, e.g., at CeBIT 2006, at accompanying events during the FIFA World Cup, and also at the KI 2006 conference in Bremen.

7 Acknowledgments

We would like to thank our project partners in SmartWeb. This research was funded by the German Federal Ministry for Education and Research under grant number 01 IMD 01. The responsibility for this article lies with the authors.

References

[1] T. Becker, C. Gerstenberger, I. Kruijff-Korbayova, A. Korthauer, M. Pinkal, M. Pitz, P. Poller, J. Schehl. Natural and Intuitive Multimodal Dialogue for In-Car Applications: The SAMMIE System. In: Proc. of the 4th Europ. Conf. of Prestigious Applications of Intelligent Systems (PAIS 2006).

[2] P. Cimiano, A. Eberhart, P. Hitzler, D. Oberle, S. Staab, S. Studer. The SmartWeb Foundational Ontology. Technical report, Institute for Applied Informatics and Formal Description Methods (AIFB) University of Karlsruhe, SmartWeb Project, Karlsruhe, Germany, 2004.

[3] D. Fensel, J. A. Hendler, H. Lieberman, W. Wahlster. Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential. MIT Press, 2003.

[4] A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, L. Schneider. Sweetening Ontologies with DOLCE. In: Proc. of the 13th Int. Conf. on Knowledge Engineering and Knowledge Management (EKAW02), volume 2473 of Lecture Notes in Computer Science, Sigünza, Spain, 2002.

[5] I. Niles, A. Pease. Towards a Standard Upper Ontology. In: Proc. of the 2nd Int. Conf. on Formal Ontology in Information Systems (FOIS-2001), C. Welty and B. Smith, Ogunquit, Maine, 2001.

[6] R. op den Akker, H. Bunt, S. Keizer, B. van Schooten. From Question Answering to Spoken Dialogue: Towards an Information Search Assistant for Interactive Multimodal Information Extraction. Proc. of Interspeech 2005 - Eurospeech 9th Europ. Conf. on Speech Communication and Technology. Lisboa, Portugal, 2005.

[7] A. Pfalzgraf. Ein robustes System zur automatischen Komposition semantischer Web Services in SmartWeb. Masterthesis. Saarland University. 2006.

[8] N. Reithinger, S. Bergweiler, R. Engel, G. Herzog, N. Pfleger, M. Romanelli, D. Sonntag. A Look Under the Hood - Design and Development of the First SmartWeb System Demonstrator. In: Proc. the 7th Int. Conf. on Multimodal Interfaces (ICMI) 2005.

[9] S. Rosset, O. Galibert, G. Illouz, A. Max. Integrating Spoken Dialog and Question Answering: the Ritel Project. In: Proc. of the Int. Conf. on Spoken Language Processing (Interspeech - ICSLP) 2006.

[10] E. Sirin, B. Parsia. Planning for Semantic Web Services. In: Semantic Web Services Workshop at 3rd Int. Semantic Web Conference (ISWC2004), 2004.

[11] D. Traum, S. Larsson. The Information State Approach to Dialogue Management. In: Smith and Kuppevelt (eds.), Current and New Directions in Discourse and Dialogue, pages 325-353, Kluwer, 2003.

[12] W. Wahlster (ed.). SmartKom: Foundations of Multimodal Dialogue Systems. Cognitive Technologies, Springer, 2006.

Kontakt

Norbert Reithinger, Gerd Herzog, Anselm Blocher

DFKI GmbH
Stuhlsatzenhausweg3, D-66111 Saarbrücken, Germany
Tel.: +49 (0)681-3025346
Fax: +49 (0)681-3025020
Email: sw-info@dfki.de

[3] The arrow depicts a pointing gesture with a stylus to a location on a map displayed.

Bild

Dr. Norbert Reithinger is a Principal Researcher and Research Fellow at DFKI. He is project manager in the area of dialogue processing in the Intelligent User Interfaces department. He was and is involved in nationally and internationally funded projects on multimodal interaction like SmartKom, MIAMM, VirtualHuman and INSCAPE. For SmartWeb he is also module coordinator.

Bild

Gerd Herzog is Chief Software Engineer and has been involved in various European and national research activities as well as industrial projects with a main focus on intelligent multimedia systems. He was and is the lead software integrator for SmartKom and SmartWeb. In this role he is responsible for functional unit test and overall system evaluations of complex multimodal systems.

Bild

Dr. Anselm Blocher is Senior Researcher at DFKI. He was and is project manager of large-scale research projects like SmartKom, VirtualHuman and SmartWeb, all funded by the German Federal Ministry for Education and Research. He also participated in various EU-funded projects and was local coordinator for the project COMIC.