SMARTWEB - MOBILE DIALOGIC ACCESS TO THE SEMANTIC WEB

Norbert Reithinger

DFKI GmbH

norbert.reithinger@dfki.de

Abstract: SmartWeb provides 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. In this article we provide a short overview of the general approach of SmartWeb and of the three end-user systems.

1 Introduction

The appeal of being able to ask a question to a mobile Internet terminal and receive an answer 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. The size and the 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 provided by the so called Semantic Web [3] has an immediate appeal to be used for intelligent question-answering. Semantically annotated web pages, however, are still very rare, due to the time-consuming and costly manual mark-up.

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 Semantic Web content on various mobile devices. The SmartWeb consortium members — BMW, DaimlerChrysler, 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.

SmartWeb provides a context-aware user interface to diverse mobile 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 tourist knowledge, and of course information about topics of general interest. The user receives information provided as texts, pictures, and videos, depending on available sources.

---

[1] www.smartweb-project.org

Example user questions — with the respective English translation below — are

- "wann war Deutschland das letzte Mal Weltmeister"

  (When was Germany world champion fort he last time)

  The system presents the textual answer, and additional multimedia information like pictures, videos and texts. The user points to Andreas Brehme on a team photo.

- "welche Tore hat dieser Spieler ↗ geschossen"[2]

  (Which goals did this player score)

- "zeige mir Geldautomaten in Leipzig"

  (Show me ATMs in Leipzig)

  The system presents a map of Leipzig with ATMs which can be pointed at.

- "gib mir mehr Informationen dazu ↗"

  (Give me more information about that)

- "wie komme ich dahin "

  (How do I get there)

- "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 Leipzig und Saarbrücken"

  (Are there any traffic jams between Leipzig and Saarbrücken)

This small selection from our current set of over 500 generalized query patterns depicts the wide range of multimodal input the system is able to process.

2   General Approach

Today's commercial dialogue systems are typically realised by means of dialogue scripting using VoiceXML or 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, focuses 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

---

[2] The arrow depicts a pointing gesture with a stylus to a location on a picture displayed.

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 (SmartWeb Integrated Ontology). 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 (sport events, navigation, web services) is defined in dedicated ontologies modeled as sub-ontologies of SmartSUMO. In addition, a specific discourse ontology (DiscOnto) provides dialogue-related concepts within SWIntO. A media ontology (Smartmedia) provides representation constructs for multimodal information. Processing within SmartWeb makes use of theW3C standard EMMA (Extensible MultiModal Annotation markup language)[3] 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. The data exchange between system components at run-time is based on a RDF format.
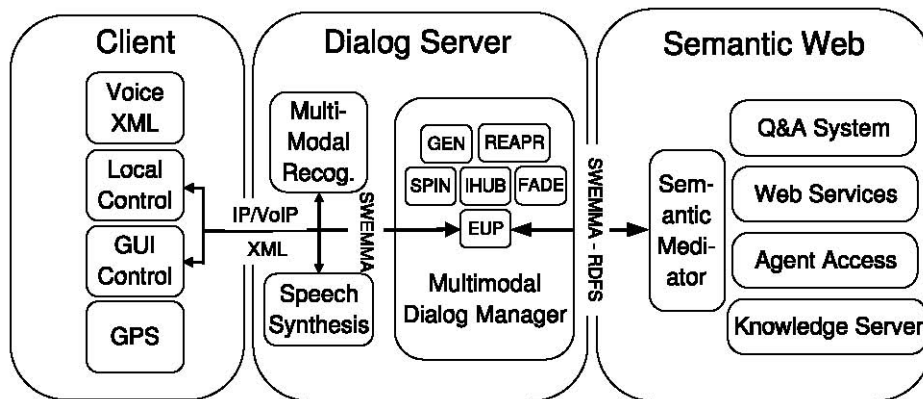
Figure 1 - The SmartWeb Architecture

Figure 1 shows a high-level architecture 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 (see below).

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 VoiceSolutions contributed their commercial speech dialogue system platform, which is being adapted and extended for multimodal interaction. The platform runs on Linux and contains all necessary functionalities for call handling over phone lines, VoIP and other data channels. It ships with client software for smartphones, which run Microsoft Windows

---

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

Mobile. The platform takes care of the basic infrastructure, providing IP connectivity between server and client. The server-side CTM (call and task manager) component coordinates multi-user operation, provides base-level barge-in processing, and includes interfaces for speech recognition, synthesis and dialogue manager functionalities. The speech recognizer is a research version of Sympalog's commercial speech recognition engine that was extended with capabilities to identify and to recognize out-of-vocabulary words.

3   Multimodal Dialogue Management and Access of the Semantic Web

The dialogue system consists of different, self-contained processing components (see [8] for more details). To integrate sub-modules for dialogue processing, 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 operates as ontology server. An EMMA Unpacker/Packer (EUP) component provides communication links to the Sympalog platform which connects to the multimodal recognizer, the speech synthesis system, and the Semantic Web services, which are external to the multimodal dialogue manager.

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 order-independent matching, i.e., the order of elements in the input stream can be ignored. 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 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 spaces [11]. 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. Input of the generation module are instances of the SWIntO ontology from REAPR 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.

After semantic analysis of the given user input, the dialogue server will forward the fully specified ontological representation of the request to the Semantic Mediator, which provides the umbrella for all different access methods to the semantic web that the system supports. To cover all information needs, SmartWeb can not rely on just one knowledge provider. The SmartWeb backend has access to rich knowledge from various sources:

- For historic world cup information, it relies on an ontology-based knowledge server that stores the factual knowledge concerning FIFA world cup tournaments. These annotated multimedia data were derived automatically from the relevant web sites during an offline pre-processing step. SmartWeb uses 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.

- An agent-based semantic wrapper component accesses web sites that provide reports of current games, which are not yet part of the knowledge base.

- The web services module [7] addresses multiple services, e.g., mobility and information services from Deutsche Telekom concerning maps, routes, point of-interests, and up-to-date information like weather data, traffic situation, cinema programs, etc. It realizes a generic approach for interactive service composition with a plan-based approach using semantic matching [10]. The novel feature is interactivity in the planning process: If some information is missing in the composition process, the access does not simply fail. Instead, missing information items are detected and requested from the user.

- Multimedia question-answering (Q&A) is responsible for open-domain questions to provide answers and images concerning persons, locations, and others named entities. This sub-system, which searches and analyses web pages on the fly, complements the other access components, which focus on specific application domains.

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.

4 The SmartWeb Scenarios

4.1 SmartWeb Handheld

In this scenario, the user carries a smartphone as interaction device (see fig.2 for two examples). Currently, we use smartphones with Microsoft Windows Mobile, like T-Mobile's MDA and AMEO series. The user can enter multimodal questions related to the different application contexts mentioned above. Speech is recorded using the built in microphone, pointing gestures use the built-in touch-screen.

Figure 2 - Two screenshots of the SmartWeb Handheld system

The interaction interface was designed carefully to enable the best possible user experience. Through a microphone icon (top left), the user has feedback about the recording status. Recognized speech input is presented in the top text line which offers correction possibilities

through the pen in case of recognition errors. Since the screen display size is extremely limited on the mobile device, we only display the most important information or functionality to the user. The main textual answer and the accompanying media objects (texts, pictures, and videos) are presented below. The user can navigate through the various media objects with the buttons at the bottom of the screen.

In the examples the user asked for Italian Restaurants (left screen shot) and web cams for Berlin's Potsdamer Platz (right screen shot). The user can select items on the left screen to enter follow-up multimodal requests. On the right hand side, the system detected four web cams. The user navigates through the results with the pen by clicking on the media navigation icons at the bottom of the screen.

4.2   SmartWeb Motorbike

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

The second scenario provides a novel interaction interface for motor bikers. 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 the left, 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 also allows multimodal interaction. Speech control is based on a small, embedded Java VoiceXML interpreter which can be addressed through a commercial on-board speech recognition engine for command-and-control utterances. It 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 left handle bar on the motor bike. In the left section of the screen, the GUI resembles a dial that the user can rotate with this device. The menu entries on the right hand side are attached to the wheel elements and can be selected/deselected by pushing or dragging the force-feedback element with the thumb. As direct manipulation using the force-feedback element can be tedious, the user can alternatively use speech to directly 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 (UMTS or WLAN), the local system can also connect to a remote SmartWeb server. The content being displayed and the operation of the server-side system is adapted to the specific needs and restrictions of a motorcyclist. The right-hand side screenshot shown in Figure 3 presents a more server-side interaction 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.

4.3 SmartWeb Automobile

SmartWeb Automobile is a new kind of mobile information system that interacts with drivers in natural language. It combines speech recognition technology from DaimlerChrysler, dialogue processing using Siemens' DIANE technology, and DAB data distribution provided by FhG FIRST. While you drive, the system searches the Internet for useful information and distributes data efficiently to all vehicles using digital broadcasting. Typical use cases from the application scenario include for example to ask for the cheapest gas station in a city, current speed camera locations, or how many goals Schalke 04 has scored in the recent match. SmartWeb automobile provides flexible voice dialogues to retrieve information from the different target applications.

Figure 4 - SmartWeb Automobile

The automobile demonstrator (see fig. 4 for a screenshot) can be divided into three main tasks: The first is the automatic retrieval of driver-relevant data from the Internet through a web crawler and its automatic translation into voice dialogue applications that can be deployed into the vehicle. The extraction is centred around selected services that proved to be useful while on the road. The creation of dialogue applications is based on a flexible approach for the combination of individual items of information from sources like web pages into utterances as they need to be understood and produced by the system. The dialogue information is then in a second step transferred via Digital Multimedia Broadcast (DMB) technology. This allows for easy distribution of new dialogues with new content to many cars in one broadcast. In the last step the user can address the new content in the car. SmartWeb Automobile is integrated in the already existing interaction environment in the car. It again incorporates a connection to the server-based SmartWeb system, comparable to the motor bike solution.

5 Summary

In this contribution 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.

The system was successfully presented on various occasions over the last year, e.g., at CeBIT 2006 and 2007, at accompanying events during the FIFA WorldCup, and also at the KI 2006 conference in Bremen.

Acknowledgments

References

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

[2] Oberle, D., Ankolekar, A., Hitzler, P., Cimiano, P., Sintek, M., Kiesel, M., Mougouie, B., Vembu, S., Baumann, S., Romanelli, M., Buitelaar, P., Engel, R., Sonntag, D., Reithinger, N., Loos, B., Porzel, R., Zorn, H.-P., Micelli, V., Schmidt, C., Weiten, M., Burkhardt, F., and Zhou, J.: DOLCE ergo SUMO: On Foundational and Domain Models in SWIntO (SmartWeb Integrated Ontology). Journal of Web Semantics. To appear.

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

[4] Gangemi, A., Guarino, N., Masolo, C., Oltramari, A. ,Schneider, L.: Sweetening Ontologies with DOLCE. In Proc. of the 13th International Conference on Knowledge Engineering and Knowledge Management, LNCS 2473. Springer. 2002.

[5] Niles, I., Pease, A.: Towards a Standard Upper Ontology. In Proc. of the 2nd International Conference on Formal Ontology in Information Systems. 2001.

[6] Op den Akker, R., Bunt, H., Keizer, S., van Schooten, B.: From Question Answering to Spoken Dialogue: Towards an Information Search Assistant for Interactive Multimodal Information Extraction¡ Proceedings of Interspeech 2005. Lisboa. 2005.

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

[8] Reithinger, N., Bergweiler, B., Engel, R., Herzog, G., Pfleger, N., Romanelli M., and Sonntag, D.: A Look Under the Hood – Design and Development of the First SmartWeb System Demonstrator. In: Proceedings the 7th International Conference on Multimodal Interfaces. Vancouver. 2005.

[9] Rosset, S., Galibert, O., Illouz, G., Max, A.: Integrating Spoken Dialog and Question Answering: The Ritel Project. In: Proceedings of Interspeech 2006. Pittsburgh. 2006.

[10] Sirin, E. and Parsia, B.: Planning for Semantic Web services. In: Semantic Web Services Workshop at 3rd International Semantic Web Conference. 2004.

[11] Traum, D. and Larsson, S.: The Information State Approach to Dialogue Management. In Smith, R. and van Kuppevelt, J.C.J. (ed): Current and New Directions in Discourse and Dialogue. Kluwer, 2003.

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