An Information State-Based Dialogue Manager for Making Voice Web Smarter

Marta Gatius   Meritxell González   Elisabet Comelles

Technical University of Catalonia, Software Department
Campus Nord UPC, Jordi Girona, 1-3,
08034 Barcelona, Spain
34-93-4137797

{gatius,mgonzalez,comelles}@lsi.upc.edu

ABSTRACT

In this paper we propose the integration of intelligent components technologies (natural language and discourse management) in voice web interfaces to make them smarter. We describe how we have integrated reusable components of dialogue management and language processing in a multilingual voice system to improve its friendliness and portability. The dialogue management component deals with complex dialogue phenomena, such as user-initiative dialogues, and follows the information state-based theory. The resulting dialogue system supports friendly communication (through the telephone and the web) in several languages: English, Spanish, Catalan and Italian. The dialogue system has been adapted to guide the users to access online public administration services.

Categories and Subject Descriptors

H.5.2. [Information Interfaces and Presentation]:User Interfaces - *Natural language, Voice I/O,User interface management systems.*

General Terms: Management, Design, Human Factors.

Keywords: Voice web interfaces, dialogue management, multilinguality.

1. INTRODUCTION

Dialogue systems (DSs) which guide the users to access web services and information can improve the usability and accessibility of the web content. Although voice interfaces providing access to the web already exist, most of them only support simple dialogues.

One of the most relevant works in the area of commercial voice systems is the definition of the VoiceXML language, the standard widely adopted to provide telephone access to the web content. The VoiceXML language is appropriate to define simple dialogues, which ask the users to give the specific information the service needs. However, VoiceXML systems only support very limited user initiative (the user can only choose the order in which the information asked by the system is given) and they do not support complex dialogue phenomena, such as clarification. Besides, in VoiceXML systems all possible sequences of interactions have to be defined for each service. Furthermore, VoiceXML systems only support voice mode.

On the other hand, there are research DSs which use domain and dialogue management models and have reusable components of discourse management and language (such as [1,4,5]). Several of those complex DSs achieve a friendly communication. However, the cost of developing those DSs from scratch is high and their adaptation to support telephonic access to different types of web services requires some efforts.

In this paper, we are concerned with the integration of a dialogue manager (DM) component into a multilingual DS ([2]), based on VoiceXML, which supports access to online public administration services. The incorporation of the generic DM component, which supports rich communication, improves the friendliness and portability of the DS. Additionally, a component for text processing has also been incorporated into the DS to support text mode and to enhance improve the voice recognition module.

2. INTEGRATION OF THE LANGUAGE AND DIALOGUE TECHNOLOGIES

In this section we describe how language and dialogue management technologies have been integrated in the DS we developed to access the web content. The architecture of the DS is shown in Figure 1.
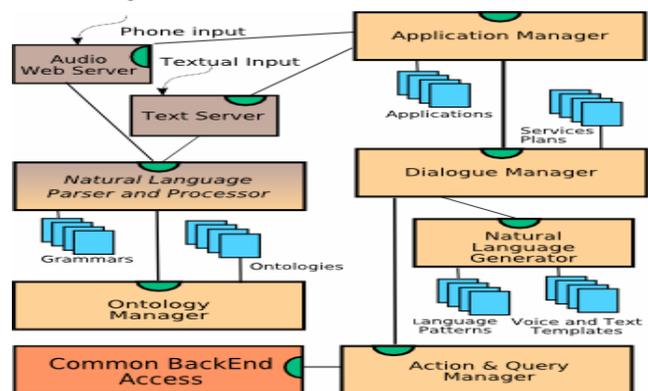

Figure 1. The architecture of the web dialogue system.

2.1 Dialogue Management

The DM component controls the dialogue flow. In order to achieve a friendly communication this component uses an explicit dialogue model which defines general dialogue mechanisms, such as feedback strategies. The DM follows the issue-based approach explained in [3], which describes dialogues in terms of issues being raised and resolved. In our system, these issues basically consist of the service tasks and their parameters.

The DM uses communication plans to determine which is the issue raised by the user and how to solve it. The DM uses these plans to recognize when the user asks for a specific service task (from a set of possible tasks) and when he provides the information necessary to perform a service task, even when no question about it has yet been raised.

Let us consider, for example, the dialogue interaction shown in Figure 2. The system asks the user to choose between one of the two services supported: the transactional service for large objects collection and the informational service about cultural events. The user does not answer the question, instead, he asks for information about movies. The DM gets the interpretation of the user's intervention *([ask, [event-type, cinema],[location, filmoteca]]*) and finds that this information is the answer to questions in the communication plan for the cultural agenda.

S1: Welcome to the automatic platform of Barcelona. Choose one service available: large objects collection service or cultural agenda
U1: I'm looking for movies in the Filmoteca
S2: Ok, you are interested in the title of the event. Ok, the event type is cinema. The place is Filmoteca.
　　　*** <database consultation>
　　　…

Figure 2. Example of a dialogue interaction.

Communication plans can be decomposed into actions and subplans. Possible plan actions are *ask*, *answer* plus the system actions to access the web services. To reduce complexity in dialogue management the plans are generated statically, when a new service is incorporated. To facilitate the generation of plans we have defined templates which describe general plans for two different types of web services: transactional and informational.

We have followed the information state-based theory to implement the DM component. This theory is based on a rich representation of dialogue context (the information state). In our DM, the information state consists of two parts:

- The information shared with the user. It includes questions under discussion (not resolved yet) and commitments (answers to previous questions).

- The information private to the system. It includes the actions to do and the set of believes (the results obtained from previous access to the web services).

There is a set of rules which govern the way the information state is updated and the next system dialogue actions.

2.2 The Language Components

In order to support a friendlier communication the speech recognition component (from the Loquendo VoiceXML platform) has been adapted to recognize a broader range of users' interventions. The automatic speech recognition uses grammars (in the standard SRGS formalism) to model possible user's input. Recognition grammars limit the user's interventions which can be understood to what the grammars themselves allow. In the previous prototype of the DS the voice grammars modelled the possible user's answers to the last system message. In order to cope with other possible user's interventions, these voice grammars have been extended.

We have incorporated a natural language parser and processor (NLPP) to enhance the capabilities of the voice recognition module. The recognized input (transformed in text) is passed to the NLPP which performs a deep syntactic and semantic analysis. The NLPP uses domain independent linguistic resources as well as domain-restricted lexicons and ontologies.

When a new web service is incorporated into the system, the appropriate system's prompts are generated automatically in the four languages supported by the system: English, Spanish, Catalan and Italian. In order to obtain the most appropriate system's prompts for a specific service, the generator component uses a syntactic-semantic taxonomy which relates the specific service tasks and parameters to the linguistic structures needed for their expression.

We have distinguished two types of users: novices and experts (they have used the system before). System's prompts for novice users guide them to give all the data the service needs. System's messages for expert users are more open. For example, in the dialogue shown in Figure 2, the user has been considered a novice. If the user is considered an expert, the first system's message is "*Welcome to the automatic platform of Barcelona. May I help you?*".

3. CONCLUSIONS AND FUTURE WORK

The evaluation of the performance of the DM component using only the text mode has proven that simple dialogues in which the system asks the user for specific information are appropriate for transactional services but more flexible dialogues are required for informational services (when user searches for different information). The evaluation of the performance of the definitive prototype of the system is planned for the following months. Future work will also include the adaptation of the DS to other types of web information.

4. ACKNOWLEDGMENTS

5. REFERENCES

[1] Allen, J., Byron,D., Dzikovska, M., Ferguson, G., Galescu, L., Stent, A., Toward Conversational Human-Computer Interaction. AI Magazine,v. 22,no. 4,(Winter, 2001),27-38.

[2] Gatius, M., Gonzalez, M., Militello, S. and Hernández, P. Integrating Semantic Web and Language Technologies to Improve the Online Public Administrations Services. In the Proceedings of the WWW'06 Conference, (May,2006).

[3] Larsson, S. Issue-based Dialogue Management. PhD Thesis, Goteborg University, 2002.

[4] Polifroni, J. Chung, G. and Seneff, S. Towars the Automatic Generation of Mixed-Initiave Dialogue Systems from Web Contents. In the Proceedings of the EUROSPEECH'03 Conference, (Geneva, 2003).

[5] Traum, D., Bos, J., Cooper, R., Larsson, S., Lewin, I., Mathesson, C., Poesio, M. A model of Dialogue Moves and Information State Revision. Trindi Technical Report D2.1, 1999. http://www.ling.gu.se/projekt/trindi//publications.html