Griol D., Molina J.M., Sanchís de Miguel A. (2014) Developing Multimodal Conversational Agents: From the Use of VoiceXML to Android-Based Applications. In: Demazeau Y., Zambonelli F., Corchado J.M., Bajo J. (eds) Advances in Practical Applications of Heterogeneous Multi-Agent Systems. The PAAMS Collection. PAAMS 2014. Lecture Notes in Computer Science, vol 8473.
pp 335-338 Springer, Cham

DOI: 10.1007/978-3-319-07551-8_32

# Developing Multimodal Conversational Agents: From the Use of VoiceXML to Android-Based Applications

David Griol, José Manuel Molina, and Araceli Sanchis de Miguel

Computer Science Department
Carlos III University of Madrid
Avda. de la Universidad, 30, 28911 - Leganés, Spain
{david.griol,josemanuel.molina,araceli.sanchis}@uc3m.es

**Abstract.** The current industrial development of commercial conversational agents and dialog systems deploys robust interfaces in strictly defined application domains. However, commercial systems have not yet adopted new perspectives proposed in the academic settings, which would allow straightforward adaptation of these interfaces. In this paper, we propose two approaches to bridge the gap between the academic and industrial perspectives in order to develop conversational agents using an academic paradigm for dialog management while employing the industrial standards, like the VoiceXML language or the Android OS. Our proposal has been evaluated with the successful development of different spoken and multimodal systems.

**Keywords:** Human-agent interaction, User interfaces, Conversational agents, Spoken and Multimodal interaction, Statistical methodologies.

## 1 Introduction

Speech Technologies and Language Processing have made possible the development of a number of new applications which are based on conversational agents [1]. Speech access is then a solution to the shrinking size of mobile devices (both keyboards to provide information and displays to see the results). Besides, speech interfaces facilitate the access to multiagent systems [2], especially in environments where this access is not possible using traditional input interfaces (e.g., keyboard and mouse). It also facilitates information access for people with visual or motor disabilities.

In this paper we describe two approaches than can be used to bridge the gap between the academic and industrial perspectives in order to develop dialog systems using an academic paradigm based on a statistical dialog management technique [3] combined with the industrial standards, like the VoiceXML standard[1] or the Android OS [4]. This makes it possible to obtain new generation interfaces without the need for changing the already existing commercial infrastructures. The first approach is oriented to the development of spoken conversational agents, while the second approach also allows to develop systems dealing with multimodal inputs and outputs.

## 2 Main Purpose

Our first approach to integrate statistical methodologies in industry applications combines the flexibility of statistical dialog management with the facilities that VoiceXML offers, thus introducing statistical methodologies for the development of commercial (and not strictly academic) dialog systems. Our technique employs a statistical model based on neural networks that takes into account the history of the dialog up to the current dialog state in order to predict the next system response [3]. To learn the dialog model we propose the use of dialog simulation techniques. Our approach for acquiring a dialog corpus is based on the interaction of a user simulator and a dialog manager simulator [5]. In addition, the system prompts and the grammars for ASR are implemented in VoiceXML compliant formats, for example, JSGF or SRGS.

A VoiceXML-compliant platform (such as Voxeo Evolution[2]) is used for the creation of Interactive Voice Response (IVR) applications and the provision of telephone access. Static VoiceXML files and grammars can be stored in the voice server. We propose to simplify these files by generating a VoiceXML file for each specific system prompt, as can be observed in the bottom left corner of the figure. Each file contains a reference to a grammar that defines the valid user's inputs for the corresponding system prompt.

The conversational agent selects the next system prompt (i.e. VoiceXML file) by consulting the probabilities assigned by the statistical dialog manager to each system prompt given the current state of the dialog. This module is stored in an external web server and is implemented using a data structure to store the information that is provided by the user in each dialog turn. The result generated by the statistical dialog manager informs the IVR platform about the most probable system prompt to be selected for the current dialog state. The platform just selects the corresponding VoiceXML file and reproduces it to the user.

Our second approach is focused on the development of multimodal conversational agents for mobile devices operating with the Android OS [4]. Our proposal integrates the Google Speech API to include the speech recognition functionality in a multimodal conversational agent. The development of multimodal systems involves user inputs through two or more combined modes, which usually complement spoken interaction by also adding the possibility of textual and tactile inputs provided using physical or virtual keyboards and the screen. In our contribution, we also model the context of the interaction as an additional valuable information source to be considered in the fusion process. We propose the acquisition of external context by means of the use of sensors currently supported

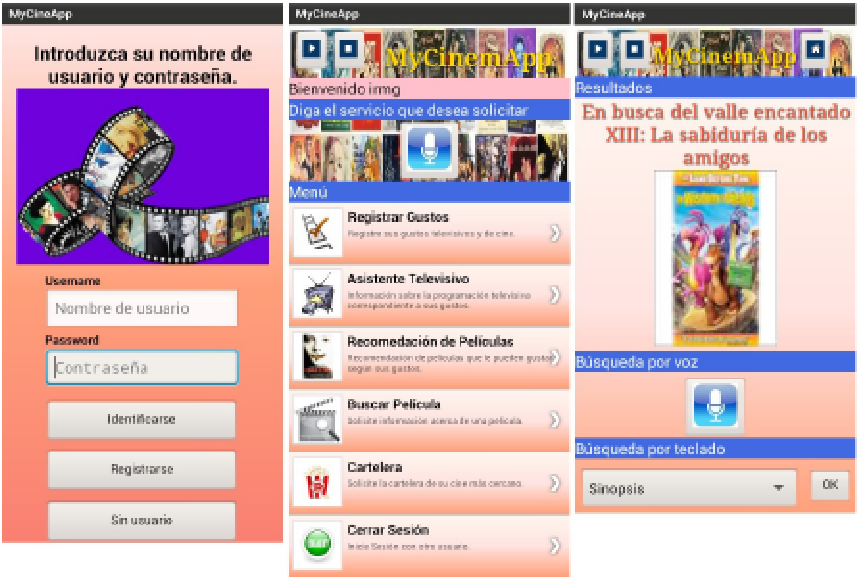
MyCineApp
Introduzca su nombre de usuario y contraseña.
Username
Nombre de usuario
Password
Contraseña
Identificarse
Registrarse
Sin usuario
MyCineApp
MyCinemApp
Bienvenido irmg
Diga el servicio que desea solicitar
Menú
Registrar Gustos
Registre sus gustos televisivos y de cine.
Asistente Televisivo
Información sobre la programación televisiva correspondiente a sus gustos.
Recomedación de Peliculas
Recomendación de películas que le pueden gustar según sus gustos.
Buscar Pelicula
Solicite información acerca de una película.
Cartelera
Solicite la cartelera de su cine más cercano.
Cerrar Sesión
Inicie Sesión con otro usuario.
MyCineApp
MyCinemApp
Resultados
En busca del valle encantado XIII: La sabiduría de los amigos
Búsqueda por voz
Búsqueda por teclado
Sinopsis
OK

**Fig. 1.** Different functionalities of the *MyCineApp* multimodal system

by Android devices. The Android sensor framework (*android.hardware* package) allows to access these sensors and acquire raw sensor data.

The dialog manager of the system is based on the previously described statistical methodology. The visual structure of the user interface (UI) is defined by means of layouts, which are defined by declaring UI elements in XML or instantiating layouts elements at runtime. Finally, we propose the use of the Google TTS API to include the text-to-speech functionality. The *android.speech.tts* package includes the classes and interfaces required to integrate text-to-speech synthesis in an Android application.

## 3 Demonstration

We have developed different conversational agents using the described approaches. As an example of the application of the first approach we present a conversational agent developed to provide information in Spanish about movies, current billboard, and awards in different festivals. This information has been extracted from the FilmAffinity movie recommendations website[3]. The application is internally divided into three main modules. The first module corresponds to the beginning of the interaction in which the user is welcomed and the system provides detailed instructions about the different functionalities. The second module includes the access to these functionalities, related to information about movies, festivals and current billboard. The third module includes different libraries developed in PHP to access and parse the information extracted from the filmaffinity website and generate the corresponding system prompts.

As an example of the application of our second approach we present a multimodal system developed for a similar domain, providing information about films and TV programs in Android-based devices. This information is adapted taking into account the specific preferences and suggestions selected by the users. The application is divided into different modules that allow application registration, complete a user profile, access the list of TV programs and the current billboard, or obtain adapted recommendations related to these information sources. Figure 1 shows different screens of the *MyCineApp* multimodal system.

## 4 Conclusions

In this paper, we propose two techniques for developing conversational agents using well-known standards and operative systems like VoiceXML or Android, and also including a statistical dialog manager automatically learned from a dialog corpus. The main objective of our work is to reduce the gap between academic and industry perspectives and take the best of both methodologies. On the one hand, the effort that is required for the definition of optimal dialog strategies is reduced. On the other, VoiceXML and Android-based implementations makes it possible to benefit from the advantages of using the different devices and platforms that are already available to simplify the development of conversational agents. The paper also describes two systems developed using the described techniques and respectively providing spoken or multimodal access to users' adapted information about movies and TV programs.

**Acknowledgements.** This work was supported in part by Projects MINECO TEC2012-37832-C02-01, CICYT TEC2011-28626-C02-02, CAM CONTEXTS (S2009/TIC-1485).

## References

1. Pieraccini, R.: The Voice in the Machine: Building Computers that Understand Speech. The MIT Press (2012)
2. Corchado, J., Tapia, D., Bajo, J.: A multi-agent architecture for distributed services and applications. Computational Intelligence 24(2), 77–107 (2008)
3. Griol, D., Hurtado, L., Segarra, E., Sanchis, E.: A Statistical Approach to Spoken Dialog Systems Design and Evaluation. Speech Communication 50(8-9), 666–682 (2008)
4. McTear, M., Callejas, Z.: Voice Application Development for Android. Packt Publishing (2013)
5. Griol, D., Carbó, J., Molina, J.: An Automatic Dialog Simulation Technique to Develop and Evaluate Interactive Conversational Agents. Applied Artificial Intelligence 27(9), 759–780 (2013)