

Architecture of a multi-modal dialogue system oriented to multilingual question-answering

Rafael M. Terol and Patricio Martínez-Barco and Manuel Palomar

Departamento de Lenguajes y Sistemas Informáticos

Universidad de Alicante

Carretera de San Vicente del Raspeig - Alicante - Spain

Tel. +34965903653 Fax.+34965909326

{rafamt, patricio, mpalomar}@dlsi.ua.es

Abstract

In this paper, a proposal of a multi-modal dialogue system oriented to multilingual question-answering is presented. This system includes the following ways of access: voice, text, avatar, gestures and signs language. The proposal is oriented to the question-answering task as a user interaction mechanism. The proposal here presented is in the first stages of its development phase and the architecture is presented for the first time on the base of the experiences in question-answering and dialogues previously developed. The main objective of this research work is the development of a solid platform that will permit the modular integration of the proposed architecture.

1 Introduction

In this paper we present the arguments that have motivated us to develop a multi-modal dialogue system oriented to multilingual question-answering, as well as its architecture. This architecture is going to permit, on the one hand, the information access to determined collective social that previously had impossibilities to do it, and on the other hand, the question-answering of digital documents.

Following sections present the state of art of Dialogue Systems, the system architecture proposal, some results obtained, and future works.

2 Background in Dialogue Systems

This section tries to synthesize the state of the art of the currently existing dialogue systems before to present our multi-modal dialogue system oriented to multilingual question-answering.

2.1 Dialogue Systems

A dialogue system normally interacts with other kind of system. According to the classification of dialogue systems proposed by James Allen (Allen 97), it fits to distinguish between two kind of dialogue systems depending on the main goal of the global system. These are:

- *Dialogue systems oriented to tasks (DSOT).*

This dialogue systems interacts with other systems whose main goal is to perform a concrete task such as evacuating people from an island (Ferguson & Allen 98), coordinating responses to emergencies (Allen *et al.* 00a), etc.

Some examples of these kind of systems have been developed by the Department of Computer Science of University of Rochester: In (Allen *et al.* 96) a system that allows the users to find efficient routes for a group of trains is presented. This system avoids problem areas as they were discovered. The definition of (Ferguson & Allen 98) treats about a system whose main task is to evacuate people from an island. (Allen *et al.* 00b) define a system that make the deployment of troops in a military situation. (Allen *et al.* 00a) define a system that coordinates emergency vehicles in response to simulated 911 calls in Monroe County, NY.

- *Dialogue systems oriented to information retrieval (DSOIR).* These dialogue systems interact with other systems whose main goal is information searching as well as information retrieval or question answering.

Here, some existent examples of this kind of systems are going to be presented. (Gurzki 01) presents a human-computer dialogue system that allows the user to get information about products on a shop. (Johansson *et al.* 02) define a human-computer dialogue system for information retrieval about TV programs. The definition of (Rouillard 99) is about a human-computer dialogue system for information retrieval about the documents of a digital library via the World Wide Web. In (Bagga *et al.* 00), a human-computer dialogue system that allows users to retrieve identification numbers of parts for medical

systems is presented.

2.2 Multi-modal Dialogue Systems

It is a fact that in the last and next years the use of human-computer multi-modal conversational dialogue systems allows a more easily access to information to people having problems to do it. Nowadays, the main task of many researchers is to work day to day in the construction of systems that allows a flexible multi-modal communication between humans and computers. Fruit of all this effort different systems have arisen making it feasible, such as:

- (Gustafson *et al.* 00) present a human-computer multi-modal conversational dialogue system in the domain of information seeking about apartments in Stockholm. Apart from spoken input, users have the possibility of providing the system with additional information by clicking on an apartment icon or marking areas on an interactive map of Stockholm.
- One of the goals of the August project (Gustafson *et al.* 99) was to be able to analyze how novice users interact with a multi-modal information kiosk, placed without supervision in a public location. Another goal was to demonstrate how the speech technology modules developed at the department could be put together to rapidly prototype a multi-modal spoken dialogue system. The system was designed with a number of domains, instead of one single complex domain.
- (Chai 02) defines a human-computer multi-modal conversational dialogue system in the domain of information searching about houses in Irvington Town. In this system, user multi-modal inputs exhibit a wide range of varieties. They could be abbreviated, ambiguous or complex. Fusing inputs together often cannot reach a full understanding. To process these inputs, contexts are important.
- In (Johnston *et al.* 02) the authors describe a multi-modal application architecture which combines finite-state multi-modal language processing, a speech-act based multi-modal dialogue manager, dynamic multi-modal output generation, and used-tailored text planning to enable rapid prototyping

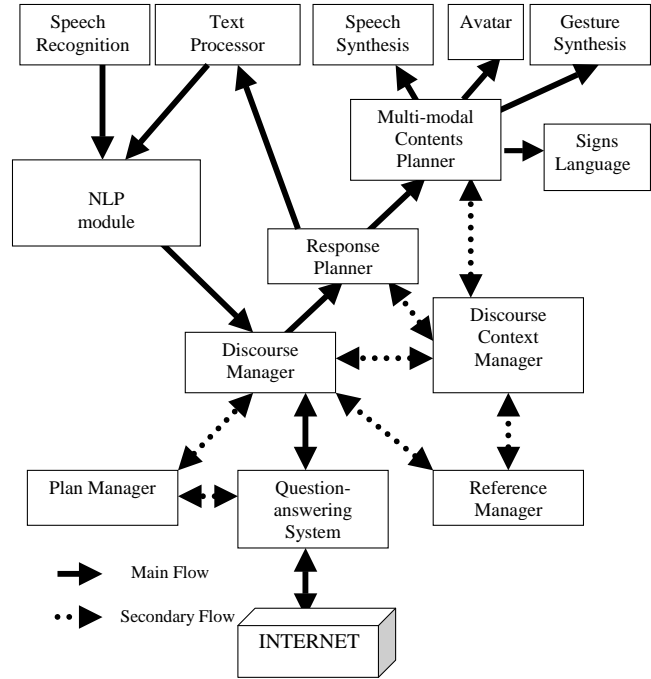


Figure 1: Architecture of a multi-modal dialogue system oriented to multilingual question-answering

of multi-modal interfaces with flexible input and adaptive output. The tested application MATCH (Multi-modal Access To City Help) provides a mobile multi-modal speech-pen interface to restaurant and subway information for New York City.

3 Proposed System Architecture

Following subsections present our proposal of a multi-modal dialogue system oriented to multilingual question-answering. In this paper has been focused on its main features and the arguments causing our interest in this area.

3.1 System Architecture

Figure 1 shows the core set of modules that integrate the system architecture, and Table 1 gives a brief description of each one.

3.2 Multi-modal Dialogue System

As Figure 1 shows, this technological proposal tries to combine the following features:

- multi-modality,
- great coverage accessibility,
- access to multilingual information,

Module	Functionality
Speech Recognition	Converts speech input into a word stream
Text Processor	Allows the user to input the word stream
Speech Synthesis	Transforms the word stream output into speech
Avatar	Transforms the word stream output into avatar movements
Signs Language	Transforms the word stream output into signs language movements
Gesture Synthesis	Transforms the word stream output into facial gestures
PLN Module	Obtains interpretations from input word stream
Response Planner	Determines the best communicative acts in function of the information provided by the Discourse Manager
Multi-modal Contents Planner	Determines the best multi-modal communicative acts
Discourse Manager	Coordinates effectively the dialog
Discourse Context Manager	Manages the discourse context
Plan Manager	Coordinates the information requests to start the question-answering process
Reference Manager	Identifies the referents for referring expressions
Question-answering	Realizes effectively the question-answering process

Table 1: Functionality of the modules in the Architecture of a multi-modal dialogue system oriented to multilingual question-answering

- question-answering task

The system will take like input a user request that will be able to be formulated through voice or text. This consultation will be analyzed by a parser (NLP module) whose result will be a conceptual representation of the user request.

This representation will be received by the Discourse Manager module which its main task is to coordinate the dialogue.

Discourse Manager will give information to the Plan Manager that will decide if it has sufficient information to perform the search, and in affirmative case it will provide a search start order to the Question-answering module.

This Question-answering module will launch the query in different languages. If the result of the search was ambiguous or the Plan Manager considers that there is not sufficient information to perform the search, it will generate a feedback order to the Dialogue Manager. The Dialogue Manager will order to the Response Planner the elaboration of an answer asking more information to the user. This answer would be shown in all the linguistic modes considered by the system: text, speech, avatar, gestures, etc, whose private contents would be determined by the Multi-modal Contents Planner.

The Discourse Manager will obtain information about the context from the Discourse Context Manager as co-reference resolution, ellipsis and other linguistic facts.

When the Plan Manager considers it has enough information to answer, it will generate a order to the Response Planner to show the infor-

mation in the considered ways.

3.3 Question-answering systems in non-restrict domains

The question-answering systems are defined like capable tools to obtain concrete answers to very precise needs of information from the writings documents analysis in natural language. These systems locate and extract the answer of those zones of the documents of whose content is possible to infer the information required in each question. The main components of a question-answering system are Question Analysis, Document Retrieval, Relevant Passages Selection and Response Extraction.

These components are related among themselves processing the textual information available in different levels until completing the question-answering process.

The questions formulated to the system are processed initially by the question analysis module. This process is very important since of the quantity and quality of the information extracted in this analysis will conditions in great measure the performance of the remaining modules and therefore, the final result of the system.

A part of the information obtained from this question analysis process is used by the document retrieval module to perform a first selection of pieces of text. By the great volume of documents to treat for these systems and the answer time limitations, this task is performed using information retrieval systems. The obtained result is a very reduced subset of the documentary database. Subsequent processes are going to be applied on

this reduced subset.

Subsequently, the relevant passage selection module takes charge of performing a more detailed analysis of the relevant text subset with the objective of detecting those reduced text fragments that are susceptible of containing the search answer.

Finally, the answer extraction module processes the small text fragments set obtained from the previous process with the purpose to locate and to extract the search answer.

4 Discussion and results

In this section fits to do mention to the results obtained by the system (Vicedo 02) as a module of our architecture. The final evaluation of this system was performed in the “Third Question Answering Track” organized in the TREC-10 conference.

In spite of the fact that the proposal of TREC-10 is composed of three different subtasks (main task, list task, and contextual task), by the given characteristics of this system, it was decided to participate only in the main task assuming, besides, an important limitation: SEMQA does not arrange of no process that permits to value the non-existence of answers to a question in the collection. Therefore, the system will obtain answers for all the proposed questions without keeping in mind this circumstance. This fact supposes that all those questions without answer in the document collection will be answered incorrectly.

With the purpose that it was evaluated separately the incident that the resolution techniques application of the pronominal anaphora could have in the results of SEMQA, two different tests were performed called ALIC01M1 and ALIC01M2:

1. ALIC01M1. This test is designed to evaluate the system performance without utilizing any kind of co-reference resolution.
2. ALIC01M2. This test is similar to the previous one although incorporates the pronominal anaphora resolution module.

Table 2 shows the general features of the test TREC-10 for the main task.

4.1 Evaluation Results

Table 3 shows the results obtained by the system in comparison with the result media obtained

Features	TREC-10
Number of documents	978.952
Documents in megabytes	3.033
Number of proposed questions	500
Number of evaluated questions	492
Non response questions in the collection	49

Table 2: Features of the Trec-10 test

by the participating remainder of systems in the main task.

In Table 3 is detailed, for each evaluation measures (strict and permissive), the reciprocal average (MRR), the percentage of answers that the system responds correctly (% Corr) and the percent of increment that this value supposes on the TREC-10 average (% Δ).

5 Conclusions and Future Works

As above mentioned, the objective of this system, is to bring the information to determined collective social that previously had impossibilities to obtain it.

At present, we dedicate remarkable research efforts, among others, to the developing of a definition language for dialogue systems whose first pilot experience is going to be the definition of a dialogue system oriented to obtain the price of real estate as flats, garages, houses, bungalows, etc. Also, we are designing a language that permit ourselves to implement multi-modal dialogue systems. This language is a VoiceXML (Consortium 03) higher level abstraction.

6 Acknowledgements

This poster has been partially supported by the TABIMED Project (TABIMED 03).

References

- (Allen 97) J. Allen. Dialogue Modelling by James Allen. ACL/EACL, 1997. <http://acl.ldc.upenn.edu/W/W97/W97-0600.pdf>.
- (Allen *et al.* 96) J. Allen, Bradford W. Miller, Eric K. Ringger, and Teresa Sikorski. A Robust System for Natural Spoken Dialogue. In *Proceedings of 34th Annual Meeting of the Association for Computational Linguistics*, pages 62–70, Santa Cruz, California, 1996.
- (Allen *et al.* 00a) J. Allen, D. Byron, D. Costello, M. Dzikovska, G. Ferguson, L. Galescu, and A. Stent. TRIPS-911 System Demonstration. In *Proceedings of the NAACL/ANLP 2000 Workshop on Conversational Systems*, pages 32–35, Seattle,

TREC-10 Results						
Probe	Strict			Lenient		
	MRR	% Corr.	% Δ	MRR	% Corr.	% Δ
TREC-10 Average	0,234	33,0	0,0	0,246	34,6	0,0
ALIC01M1	0,296	39,2	18,8	0,302	40,0	15,6
ALIC01M2	0,300	39,6	20,0	0,306	40,4	16,8

Table 3: Evaluation Results

- WA, 2000. Association for Computational Linguistics.
- (Allen *et al.* 00b) J. Allen, D. Byron, M. Dzikovska, G. Ferguson, L. Galescu, and A. Stent. An Architecture for a Generic Dialogue Shell. *Natural Language Engineering, special issue on Best Practices in Spoken Language Dialogue Systems Engineering*, 6(3):1–16, December 2000.
- (Bagga *et al.* 00) Amit Bagga, Tomek Strzalkowski, and G. Bowden Wise. A Dialog-Based System for Identifying Parts for Medical Systems. In *Proceedings of the Conference on Applied Natural Language Processing and the North of American Chapter of the Association of Computational Linguistics (ANLP/NAACL2000)*, pages 29–36, Seattle, WA, 2000. Association for Computational Linguistics.
- (Chai 02) Joyce Chai. Semantics-based Representation for Multimodal Interpretation in Conversational Systems. In *Proceedings of 19th International Conference on Computational Linguistics (COLING'2002)*, Taipei, Taiwan, 2002.
- (Consortium 03) World Wide Web Consortium. Voice Extensible Markup Language (VoiceXML) Version 2.0. <http://www.w3.org/TR/voicexml20/>, February 2003.
- (Ferguson & Allen 98) G. Ferguson and J.F. Allen. TRIPS: An integrated intelligent problem-solving assistant. In *Proceedings of the Fifteenth National Conference on AI (AAAI'98)*, pages 567–573, 1998.
- (Gurzki 01) Throsten Gurzki. An Architecture for Integrating Virtual Sales Assistant Technology into E-Business Environments. In Brian Stanford-Smith and Enrica Chiozza, editors, *E-work and E-commerce: Novel Solutions and Practices for a Global Networked Economy (Vol 1 and 2)*, Venice, Italy, October 2001.
- (Gustafson *et al.* 99) Joakim Gustafson, Magnus Lundberg, and Johan Liljenkrants. Experiences from the development of August - a multimodal spoken dialogue system. In *Proceedings of IDS'99*, 1999.
- (Gustafson *et al.* 00) Joakim Gustafson, Linda Bell, Jonas Beskow, Johan Boye, Rolf Carlson, Jens Edlund, Björn Granström, David House, and Mats Wirén. AdApt - a multimodal conversational dialogue system in an apartment domain. In *Proceedings of the International Conference on Spoken Language Processing*, Beijing, China, 2000.
- (Johansson *et al.* 02) Pontus Johansson, Lars Degerstedt, and Arne Jönsson. Iterative Development of an Information-Providing Dialogue System. In *Proceedings of the 7th ERCIM Workshop 'User Interfaces for All'*, Paris, France, 2002.
- (Johnston *et al.* 02) Michael Johnston, Srinivas Bangalore, Gunaranjan Vasireddy, Amanda Stent, Patrick Ehlen, Marilyn Walker, Steve Whittaker, and Preetam Maloor. MATCH: An Architecture for Multimodal Dialogue Systems. In *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics*, pages 376–383, Philadelphia, USA, July 2002.
- (Rouillard 99) José Rouillard. HALPIN: A Natural Language Information Retrieval System for a Digital Library on the World Wide Web. In P. De Bra and John J. Leggett, editors, *Proceedings of WebNet 99 - World Conference on the WWW and Internet*, pages 937–942, Honolulu, Hawaii(USA), 1999. Association for the Advancement of Computing in Education (AACE).
- (TABIMED 03) TABIMED. Simulador de precios de activos inmobiliarios. <http://www.dlsi.ua.es/proyectos/proyected.cgi?id=cas&proyected=15/>, Febrero 2003.
- (Vicedo 02) Jose Luis Vicedo. *SEMQA: Un modelo semántico aplicado a los sistemas de Búsqueda de Respuestas*. Unpublished PhD thesis, Departamento de Lenguajes y Sistemas Informáticos, Universidad de Alicante, Alicante, España, 2002.