Natural Language Dialogue System for Information Retrieval

J. A. González-Bernal, **A. López-López**, **J. Muñoz-Arteaga**, **M. Montes y Gómez**, **C. A. Reyes-García**, **L. Villaseñor-Pineda**.
{jagonzalez, allopez, jaime, mmontesg, kargaxxi, villasen}@inaoep.mx

Coordinación de Ciencias Computacionales
Instituto Nacional de Astrofísica, Óptica y Electrónica
Apdo. Postal 51 y 216
Puebla, Pue. 72000 MEXICO

Abstract

The objective of our work is the development of a natural language dialogue system for information retrieval with multimodal input and multimedia output. Overall, the system consists of three phases: input analysis, information and knowledge management and output generation. The dialogue system is designed for consulting old Mexican historical documents. In this paper we describe the designed architecture of each of the three phases.

Introduction

The general goal of natural language dialogue systems is to support effective interaction between people and computer systems. To support effective communication, dialogue systems must facilitate users' understanding by incrementally presenting the most relevant information, and by adapting the interaction to address communication problems as they arise. The present project is a long-term initiative at the Natural Language Laboratory of the National Institute of Astrophysics, Optics and Electronics. This project involves different research areas: human-machine interaction (i.e. speech recognition, multimodal interfaces, multimedia presentation, dialogue systems), information retrieval and data mining. Among other features, our dialog system is designed to be a spoken system. In this sense, it will have both, speech recognition and speech synthesis capabilities. In the first version, our system enforces, like most other conversational systems, strict turn taking between the user and the system and processes each utterance sequentially. This means that the dialogue does not allow interruptions and a goal has to be completed before initiating the next one.

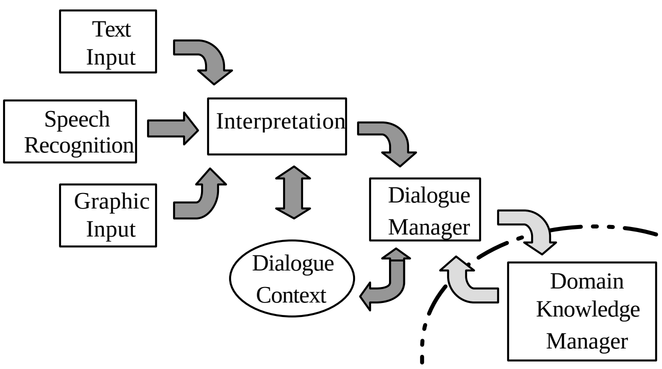

Figure 1. Input Analysis

The Input Analyzer

Figure 1 shows the section corresponding to the Input Analysis. When the user wishes to communicate with the system (either by speech, keyboard, mouse, or a combination of these), the input is interpreted and sent to the dialogue manager. The interpretation process involves a parsing step and then the resolution of anaphoric expressions and ellipsis. In order to achieve this, the interpretation process uses the dialogue and graphical contexts and also the deictic information of the last interaction. Finally, the result of this process is a grounded expression, which is sent to the dialogue manager.

The Speech Recognition Module

One of the modes selected to interact with the dialog system is through spoken input. In order to process speech to establish a dialog with the system, a speech recognition module will be implemented. The mayor task of this module consists in the transcription of speech into text. The text string is then passed to the input interpreter module, which after interpreting the components of the sentence passes the output to the Dialog Manager [1]. The ASR process is divided basically in two stages, namely, signal processing and pattern classification. In the signal processing phase, the speech wave is acquired via microphone, and after being adequately digitized, it is acoustically analyzed. The objective of the acoustical analysis is to extract acoustic feature vectors, which are the patterns of the speech units. The feature vectors are sent to the pattern recognizer to be classified to their respecting class [2]. One of the most important tasks, not only for the good performance of the speech recognizer, but also of the system as a whole, is the collection of a corpus. Since our system is directed to retrieve information from old Mexican historical documents, the corpus should contain enough samples of the words used in the documents, including the anachronic ones. For doing this, we will support on existing glossaries and related documents, besides having the advise of several experts in history.

The Dialogue Manager

The dialogue manager (DM) coordinates the system's conversational behavior. With the grounded expression received from the interpretation module, the DM tries to identify the intended speech act. This operation involves the evaluation of the grounded expression within the dialogue context (the speech act interpretations of the utterances in the previous turns) and eventually, with domain specific information (which involves generic domain knowledge, as well as, the current status of the task). Then, the dialogue manager passes the speech act to the Information and Knowledge Manager (IKM), which decides how to respond to this act. The solution calculated by the IKM is passed to the dialogue manager, which verifies the current discourse obligations. If needed, the dialogue manager may start clarification or rectification sub dialogues. Finally, the dialogue manager also provides the mechanisms to manage communication problems at different levels: speech recognition, anaphora resolution or conversational inconsistency [3,4].

The Information and Knowledge Manager

Figure 2 shows the section corresponding to the Information and Knowledge Management. The domain and knowledge manager is the main component of this section. It receives the query in a structured representation (for instance, KIF or CGIF), and then performs one of the following tasks based on the domain knowledge and the refinement information:

- Verifies the correctness of the query. For instance, it identifies whether the query is pertinent to the domain and possibly decides if it is feasible to answer.
- Enriches the query. Basically, it considers the following actions:
  - Transforming modern terms into outdated terms.
  - Separating the initial request into several more precise queries
  - Adding information such as dates or names to the query in order to restrict the searching space.
- Selects the most adequate searching engine, depending on the kind of query. One can request historical information or ask for an explanation about anachronic terms. Accordingly, we are considering engines for:
  - Word resolution
  - Document searching
  - Question answering

The Information and knowledge management module also includes a component for information refinement. This component keeps record of all searching events (i.e., the query, the results, the usefulness of the answer, the output format, etc.). This information will be explored with a data mining algorithm to find descriptive patterns that can be used to advice during the searching process and hopefully enhance the results. This component will also be used to learn patterns on which output format (i.e. graphs, text, voice, etc) works better for which types of queries (or users). In this way the component will be able to advice which output format should work better for the current query [5,6,7].

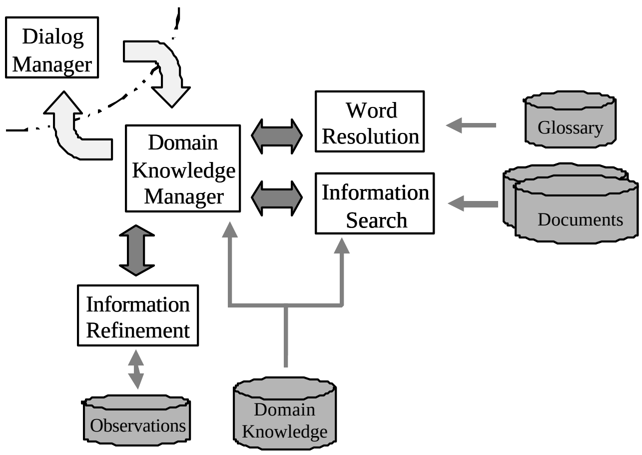

Figure 2. Information and Knowledge Manager

The Output Generation Module

The output generator provides multimedia information, integrating textual, audible and graphical expressions of historical documents. The dialogue manager passes the information to the output generator for the presentation in the interface (see Figure 3), which creates text/graphical displays or speech depending on the type of output [8].

In general, the system provides a text/graphical display for the content of the document and uses the speech synthesis to establish a negotiation with the user about more details of the user's requests. In addition, the output generator applies visualization techniques (e.g. fisheyes and hierarchical data [9]) to determine multiple forms to display the information, helping the user to have a better understanding about the contents of the historical documents.

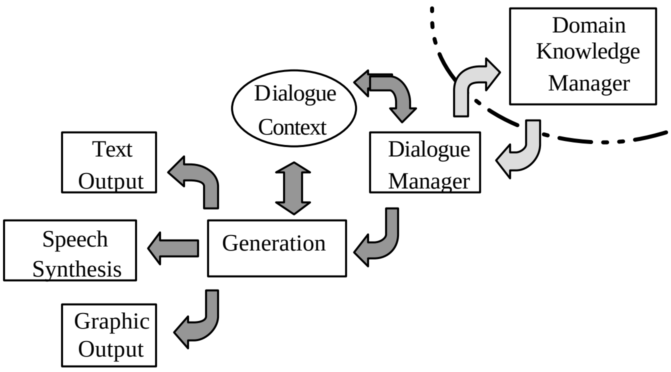

Figure 3. Output Generation Module.

References

1. Huang, X., Acero, A., y Hon, H-W, *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*, Prentice Hall PTR, New Jersey, 2001.
2. Ben Gold y Nelson Morgan, *Speech and Audio Signal Processing*, Ed. John Wiley & Sons, New York, 2000.
3. Allen, J., Byron, D., Dzikovska, M., Ferguson, G., Galescu, L. and Stent, A. "Towards Conversational Human-Computer Interaction", *AI Magazine*, 2001.
4. Xu, W. and Rudnicky, A. Task-based dialog management using an agenda. *ANLP/NAACL 2000 Workshop on Conversational Systems*, May 2000, pp. 42-47.
5. Montes y Gómez, M., López-López, A., Gelbukh, A. *Information Retrieval with Conceptual Graph Matching*, 11th International Conference on Database and Expert Systems Applications DEXA 2000, September 4 - 8, 2000, London, United Kingdom, Springer Verlag, Series: Lectures Notes in Computer Science, Vol. 1873, ISBN 3540679782, pp. 312-321.
6. López-López, A. and Myaeng, S. H. *Evidence Accumulation with Competition in Information Retrieval*, SPIRE'98 String Processing and Information Retrieval: A South American Symposium, Santa Cruz de La Sierra, Bolivia September 9-11, 1998, published by IEEE Computer Society, ISBN 0-8186-8664-2, pp. 41-49.
7. López-López, A. and Myaeng, S. H. *Extending the Capabilities of Retrieval Systems by a Two-Level Representation of Content*, Proceedings of the Australian Document Computing Symposium, Justin Zobel (Ed.), Part I, 1996, pp. 15-20.
8. Maybury, M. T. and Lee, J. "Multimedia and Multimodal Interaction Structure" in Taylor, M. M., Neel, F. and Bouwhuis, D. G. (eds.) *The Structure of Multimodal Dialogue* II. John Benjamins: Amsterdam/Philadelphia, 295-308, Chapter 15. Earlier version appeared in a pre-proceedings, ISSN 1018-4554, 2000.
9. Edited by Stuart Card, Jock Mackinlay, Ben Shneiderman, *Readings in information visualization, Using vision to think*, Morgan Kaufmann Publishers, San Francisco, January 1999, ISBN: 1-55860-533-9