INTERACTIVITY AND MULTIMODALITY IN THE IMIX DEMONSTRATOR

Lou Boves[1], Els den Os[2]

[1]Centre for Language & Speech Technology, Radboud University Nijmegen
[2]Max Planck Institute for Psycholinguistics, Nijmegen

ABSTRACT

It is generally acknowledged that many experts and almost all lay persons have difficulty in formulating requests for information in such a manner that conventional off-line Information Extraction systems can find optimal answers. Therefore, it is increasingly evident that there is a need for an interactive dialog between information seekers and information extraction systems. In this paper we describe the demonstrator of an interactive and multimodal information extraction system that is under construction in the NWO funded research program IMIX.

1. INTRODUCTION

It is widely known that even domain experts have trouble in using search engines when they are preparing reports and papers, because they are unable to guess the index terms that are attached to the relevant documents. For lay persons the problem is even larger, and it keeps growing with the growth of the amount of data that is accessible through the Internet. Here too, the problem is in the specification of the queries: often these are either too general or too specific, more often than not because the person who seeks information does not know what can be asked, nor how requests should be phrased to maximize the probability of a useful answer. Replacing search terms by natural language questions cannot completely solve the problem. It is easy to formulate too vague and general questions, or questions that are very specific but use the wrong words. According to Zweigenbaum determining the meaning of questions, even in restricted domains, is often tantamount to machine translation [1].

Ambiguities that can easily lead to misunderstanding also occur very frequently in human-human interaction. In a situation where one person tries to obtain information from another person (perhaps an expert) there is a shared responsibility for detecting potential misunderstandings. There are at least two ways in which this can happen: the 'expert' knows that a question is ambiguous, perhaps because it contains expressions that have multiple meanings. Alternatively, the information seeker decides that the answer is not what she expected. In the first case one would expect some kind of clarification dialogue, initiated by the expert to resolve the ambiguities. In the latter case one should expect follow-up questions, most probably referring to some aspects of the returned answers, the original question, or both [2, 3].

A picture can be worth a thousand words. This is also true for the answers returned in Question-Answering (QA) settings. If the answer is found in a document that contains pictures along with text, including pictures may very well improve the quality of the answer. Alternatively, advanced answer generation technology might be able to locate useful pictures in other documents, or perhaps to generate drawings. When answers contain pictures, it should be possible to include those pictures in follow-up questions, by talking and preferably also by pointing or drawing.

The NWO funded research program Interactive Multimodal Information eXtraction (IMIX) intends to address the problems sketched above by developing QA technology that can be embedded in an interactive multimodal environment, or, in other words, a multimodal dialog system. The eventual goal of the IMIX program is to improve the quality of the answers provided by a QA system by solving problems with ambiguity and lack of specificity. IMIX integrates research in several disciplines, viz. automatic speech recognition (ASR) in combination with pen input recognition, information extraction and question-answering, multimodal rendering of information, fact mining and dialog management. Most of the results of the individual research projects will be integrated in a system that demonstrates the advantages of interactive multimodal information extraction. The specification and the initial implementation of the common IMIX demonstrator is described in this paper.

2. RESRICTED DOMAIN QA

Part of the research in IMIX is devoted to open domain QA. However, IMIX also covers research problems that are more appropriately addressed in the setting of restricted domain QA. Because the IMIX demonstrator is intended to integrate as many results of the program as possible, it is only natural that the demonstrator is focused on restricted domain QA.

0-7803-9332-5/05/$20.00 ©2005 IEEE

It is well known that restricted domain QA poses different challenges than open domain QA, in several respects [4]. One important difference is the fact that in restricted domain QA one cannot rely on the redundancy of the data in the Internet. Rather, it is necessary to analyze and interpret natural language expressions in the queries and the documents that contain potential answers in great detail. This pivotal role of natural language processing fits nicely with one of the goals of the IMIX program, viz. to investigate what principled linguistic analysis can contribute to the quality of applications of language technology. Another important difference is the methods and measures with which the performance and quality of QA systems should be evaluated. However, in this paper we will not address evaluation.

2.1. Choice of the domain

At the start of the IMIX program much time has been devoted to the selection of a domain that is small enough to handle, yet challenging enough to support a five year research program. Because multimodal rendering of answers is one of the research fields in the program, it should be natural to present at least part of the information in the domain in the form of a combination of text, pictures, and tables. Moreover, a sufficiently large and diverse collection of multimedia documents should be freely accessible. Last but not least, the domain and the characteristics of the users should lead to 'analytical' questions, which are often difficult to answer appropriately without some kind of interaction with the user to make the query more precise, or to explain and extend the initial answer [5]. It appeared that a domain that satisfies all our requirements was not easy to identify. Eventually, we have settled for the medical domain in general, and for the domain of Repetitive Stress Injury (RSI) in particular. The eventual system should behave as an intelligent agent that supports lay persons who seek encyclopedic information about medical issues in general and RSI in particular [6].

2.2. Information extraction

IMIX will compare different approaches to information extraction and question-answering that can be applied both in open and restricted domains. One promising approach is based on off-line fact mining, potentially in combination with automatic induction of a domain ontology. The fact mining approach will be compared with a machine learning approach, and an approach based on exploiting the results of dependency parsing applied to the questions and the documents. Since most of the research in IMIX will use a restricted set of documents, deep parsing can be accomplished off-line, in the same way as off-line fact mining [7].

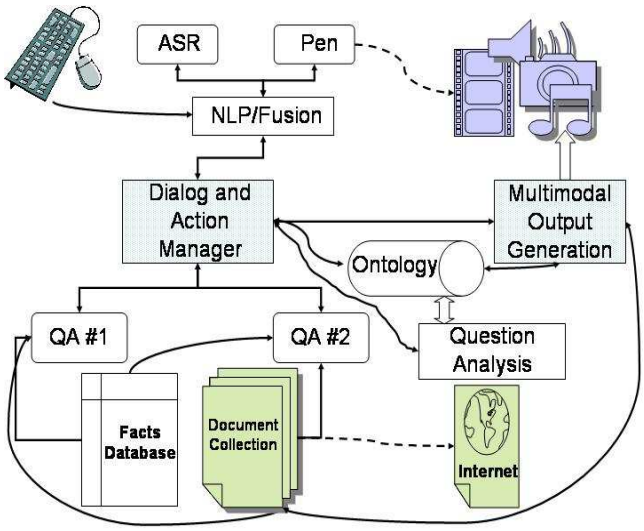

Figure 1 Functional design of the IMIX Demonstrator.

3. FUNCTIONAL SPECIFICATION OF THE IMIX DEMONSTRATOR

The functional design of the IMIX Demonstrator is shown in Fig. 1. The user can interact by means of a keyboard and mouse, or with a combination of speech and pen input. The restriction to the RSI domain is important for ASR (due to limitations of the vocabulary and language model), for multimodal output generation (because of its reliance on domain knowledge) and for the dialog and action manager (that relies on domain knowledge to be able to classify questions and decide on the most appropriate action). The architecture shown in Fig. 1 can also operate as an open domain QA system, by connecting the document collection to the Internet. For open domain QA the capabilities of the Dialog and Action Module (DAM) to engage in a dialog with the user is limited to cases where the analysis of returned candidate answers results in a small number of clusters that are easy to characterize.

User input will be analyzed to determine whether it constitutes a question that the QA modules in the system should be able to handle. Question Analysis will draw heavily upon a domain ontology, which must contain knowledge about natural language expressions in addition to information about the objects in the domain, their relations and possible actions that can be performed on or with these objects.

If the Question Analysis module detects a possible ambiguity, the DAM module will issue a request to ask for clarification to the Multimodal Output Generation module, which will convert the request into a natural language expression that can be printed on the screen or spoken through the Text-to-Speech module. An example of a

question of which the system might understand that it is ambiguous is

What can one do against RSI?

which can either mean

How can one prevent contracting RSI?

or

How can RSI be cured?

It is evident that the two interpretations should lead to quite different answers.

Another way for detecting that the query was ambiguous is by clustering the potential answers. It is evident that this is easier in open domain QA, where each query usually returns a large number of potentially relevant documents than in restricted domain QA, where the number of potentially relevant passages tends to be small, and the passages tend to be short. Therefore, in the IMIX demonstrator it will not be possible to see that a question like

Can you give me information about Java?

will probably return information on the Indonesian island as well as on the programming language. In the IMIX demonstrator more subtle analysis of the answer passages is needed to detect the ambiguity. Nevertheless, once it is detected, the DAM module can again issue a request to the Output Module to ask the user which 'Java' was meant. Here too, the fact that we are dealing with a restricted domain can be used to advantage. For example, if a user has been addressing remedies for some complaint in the previous turns, the remedy interpretation of an ambiguous query is more likely than the 'prevention' alternative. Thus, it should be advantageous to keep a record of the dialog history, not only to resolve ambiguities, but also anaphoric expressions in follow-up questions [8].

Since the IMIX demonstrator will render the answer to questions related to RSI in the form of a multimedia presentation, the user can refer to all objects on the screen in follow-up questions. If the response contains a picture of a (part of) the human body, the user may point to a specific part of that picture in a multimodal question such as

Can you which muscle this is?

It is evident that the Fusion and DAM modules must have access to the screen state to be able to interpret this type of expression.

As can be seen in Fig. 1 the Multimodal Output Generation module has also access to the internal document collection. The answer that the DAM module passes to Output Generation comprises references to the passages from which the answer was extracted. This enables the output module to extend the answer to make it more informative, by means of summarization techniques based on discourse structure [12] and syntactically and semantically correct fusion of sentences in the answer passages [13].

4. IMPLEMENTATION OF THE DEMONSTRATOR

Complex multimodal dialog systems such as the IMIX demonstrator can only be constructed by adapting existing modules such that they can inter-operate and be integrated in a single system. For IMIX we have decided to use the Multiplatform system developed by DFKI [9], partly because some of the partners in IMIX had prior experience with that integration platform [10]. This experience has facilitated the construction of the first version of the IMIX demonstrator considerably.

4.1 The architecture of version 1

The architecture of the first version of the IMIX demonstrator is depicted in Fig. 2, in the form of the GUI representation that is customarily used in connection with the Multiplatform system. The thin white lines represent data paths, while the thick white lines represent so called pools, which are comparable to blackboards. Modules must subscribe to pools for reading and writing.

The single most important goal of the first version of the first version of the IMIX demonstrator is to prove that the existing modules (ASR, QA systems and Multimodal Output Generation) can be integrated and communicate. In the first version there is no fully functional dialog management module. Consequently, this version is limited to a single query-response pair. However, the system can be –and will be- used to observe the kind of questions that user ask, and how they react to the responses. These observations will be used to bootstrap both the procedure for detecting ambiguity in the initial question analysis and users' reactions to the answers that are returned. Also, this version will allow first observations of users' reactions to the way in which the output is rendered.

Version 1 does not yet include pen input, but in future versions the combination of speech and pen input used in [10] will be integrated. This version of the demonstrator runs on a single CPU Linux computer. It is quite possible that future versions that also include dialog management and pen input will need at least two CPUs for transparent interaction, but the Multiplatform system can integrate modules that run on multiple computers in a network.

4.2 The operational modules in version 1

The NORISC.ASR module is the HTK-based speech recognition system that has already been used in [10]. For the function that ASR must fulfill in an interactive dialog,

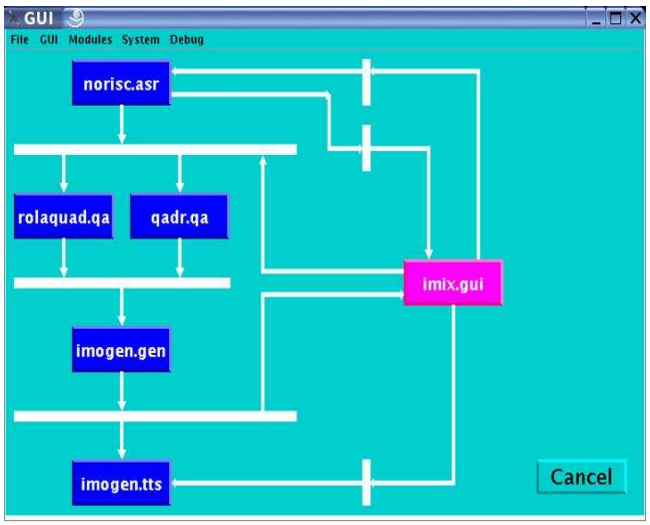

Figure 2 The architecture of the first version of the IMIX demonstrator.

the most important limitation is that the present version of HTK does not support rapid switching between vocabularies and language models.

Version 1 has two QA systems, QADR that is under development in Groningen [7], and ROLAQUAD (RObust LAnguage understanding in Question-Answer Dialogs) [11] that is being developed in Tilburg, which is based on machine learning of type, topic and content of questions and possible answers. For its development ROLAQUAD needs a substantial amount of annotated documents and questions. The annotated document collection is described in [7]. The present version does not yet use a facts database that is also described in [7]. Since there is not yet a dialog manager that can integrate the answers returned by the two QA systems, we decided to build this version in such a manner that the GUI can switch between the two QA modules. In the version that will be shown in the conference QADR can handle general medical questions (as long as the answer can be found in the internal document collection). The learning phase of ROLAQUAD has not been completed, so that this QA system can only handle queries about RSI.

The IMOGEN module is responsible for the generation of syntactically correct and semantically appropriate responses, which can be rendered by showing text and pictures on the screen, and by speaking the text by means of the NEXTENS text-to-speech system for Dutch.

5. ACKNOWLEDGEMENTS

The demonstrator system described in the paper is being constructed in the framework of the Interactive Multimodal Information eXtraction (IMIX) program, which is funded by Netherlands Organisation for Scientific Research (NWO).

6. REFERENCES

[1] P. Zweigenbaum "Question answering in biomedicine" *EACL workshop on Natural Language Processing for Question Answering*, 2003.

[2] J. Burger, et al. "Issues, tasks and program structures to roadmap research in question & answering (Q&A). Available at: http://www-nlpir.nist.gov/projects/duc/roadmapping.html

[3] L. Hirschman and R. Gaizauskas "Natural language question answering: the view from here", *Natural Language Engineering* 7 (4): pp 275-300, 2001

[4] A.T. Diekema, O. Yilmazel, and E. D. Liddy, "Evaluation of restricted domain Question-Answering systems", *Proc. ACL Workshop* "*Question Answering in Restricted Domains"*, 2004.

[5] S. Small et al., "HITIQA: Scenario Based Question Answering", *Proceedings of HLT*, Boston, Massachusetts, 2004.

[6] R. op den Akker, H. Bunt, S. Keizer, and B. van Schooten, "From Question Answering to Spoken Dialogue: Towards an Information Search Assistant for Interactive Multimodal Information Extraction", submitted to *Interspeech*, 2005

[7] E. Tjong Kim Sang, G. Bouma, and M. de Rijke, "Developing Offline Strategies for Answering Medical Questions", paper submitted to *AAAI-05 workshop on Question Answering in restricted domains,* Pittsburgh, Pennsylvania.

[8] Arne Jönsson, Frida Andén, Lars Degerstedt, Annika Flycht-Eriksson, Magnus Merkel, and Sara Norberg, "Experiences from combining dialogue system development with information extraction techniques", in: Mark T. Maybury (Ed), *New directions in Question Answering*, pp 153-164, AAAI/MIT Press, 2004.

[9] G. Herzog, H. Kirchmann, S. Merten, A. Ndiaye, P. Poller, T. Becker. (2003) MULTIPLATFORM Testbed: An Integration Platform for Multimodal Dialog Systems. *Proceedings, HLT-NAACL 2003 Workshop: Software Engineering and Architecture of Language Technology Systems (SEALTS)*, Edmonton, Alberta.

[10] E. den Os, L. Boves, S. Rossignol, L. ten Bosch, L. Vuurpijl "Conversational Agent or Direct Manipulation in Human-System Interaction", to appear in *Speech Communication.*

[11] P. Lendvai, A. van den Bosch, E. Krahmer, S. Canisius. "Memory-based Robust Interpretation of Recognised Speech". In: *Proceedings of SPECOM '04, 9th International Conference "Speech and Computer"*, St. Petersburg, Russia, pp. 415-422, 2004.

[12] W. Bosma. "Query-Based Summarization using Rhetorical Structure Theory", submitted to *Proceedings of CLIN 2004*.

[13] E. Marsi and E. Krahmer, "Explorations in Sentence Fusion", submitted to *10th European Workshop on Natural Language Generation*.