Natural Language Dialogue in a Virtual Assistant Interface

Ana M. García-Serrano, Luis Rodrigo-Aguado, Javier Calle

Intelligent Systems Research Group
Facultad de Informática
Universidad Politécnica de Madrid
28660 Boadilla del Monte, Madrid
{agarcia,lrodrigo,jcalle}@isys.dia.fi.upm.es

Abstract

Dialogue management and language processing appear as key points in virtual assistants for e-shops, as their main goal is to assist the user both in his navegation through the shop product pages and in his search for the most appropriate product. In this paper we present the details of the semantic and pragmatic approach and the dialogue management done in the ADVICE project (IST-1999-11305) for a bricolage tools shop, as well as how this modules have been evaluated.

1. Introduction

In the last years, internet has popularized as one more of the possible communication ways available. As this happens, a growing number of people need to access the resources available on it, but many of them lack the required knowledge or abilities to handle the interfaces that computers present nowadays. It is necessary to fill that communicative gap, shortening the distance between users and computers. Natural language processing emerges as one of the key fields to make this possible.

The objective of processing a general, unconstrained natural language is far from being reached. The work being done in the field at the moment could be classified as either developing linguistic resources to aid in the general target of processing language or testing natural language processing techniques in restricted domains. The work described in this paper belongs to the second approach, as it describes how a virtual assistant located in an electronic shop can interact with the users and potential buyers. In this environment, it happens that the knowledge involved is naturally limited to the products domain of the electronic shop and the system only has to deal with language present in a buyer-seller relation, so no artificial restrictions are needed.

A second main aspect of the buying-selling interaction in the web is the capability of the web site of generating some kind of trust feeling in the buyer, just like a human shop assistant would do in a person-to-person interaction. Things like understanding the buyer needs, being able to give him technical advice, assisting him in the final decision are not easy things to achieve in a web selling site. Dialogue management play a crucial role in providing this kind of enhancements.

In the work presented, the dialogue management is performed taking into account three different knowledge bases: the session model, the dialogue patterns and some features defining the user model. This knowledge is located in the interaction agent in the agent-based architecture of the developed system.

First, the general architecture of the system will be presented, followed by a brief description of the modules involved in the natural language processing and dialogue management. The second half of the paper will be devoted to the evaluation of the system.

2. System Architecture

The system has been designed following an agent based approach. There are three agents, responsible for different tasks in the system. The interface agent manages the information that is received from and presented to the user. The interaction agent is in charge of managing the dialogue. Finally, the intelligent agent contains the specific knowledge of the domain and the appropriate problem solving methods to apply that knowledge.

When the system receives any kind of input from the user, the interface agent processes it and translates the input into communicative acts (Searle, 1969), that are sent to the interaction agent. This agent builds a coherent response to the user input. If this answer needs intelligent contents that have not been previously used in the conversation, it can be provided by the intelligent agent. Finally, the interface agent will make use of all the available modes to convert the communicative acts generated by the interaction agent in a coordinated multimodal output, as can be seen in Figure 4.

In the following, the focus of the paper will be the natural language processing elements of the system, that is, the natural language interpreter and generator present in the interface agent aa well as the interaction agent.

3. Natural Language Understanding

The main objective of a natural language processing system embedded in a dialogue system is to identify and transmit the content and the intentionality of the speaker (or writer), so that the dialogue manager can make use of all the information that the user has provided.

The design of the natural language interpreter will be exposed after a short description of the corpus study carried out.

3.1. Corpus Based Analysis

As the system has to deal with a well limited sublanguage, a study of domain dialogues corpus was carried out. This corpus was artificially built for this task, as no real corpus was avalibale, and was made of 202 user

queries to the system and 10 complete dialogues, a short size, but enough to work with and extract useful conclusions. As a first step, the domain specific terminology was extracted, together with the relations among them.

From that data and and with an abstraction step from the concrete terminology to higher level concepts, a conceptual model of the domain was obtained, which included the relations among concepts and the terminology associated to each concept, that is, a domain ontology.
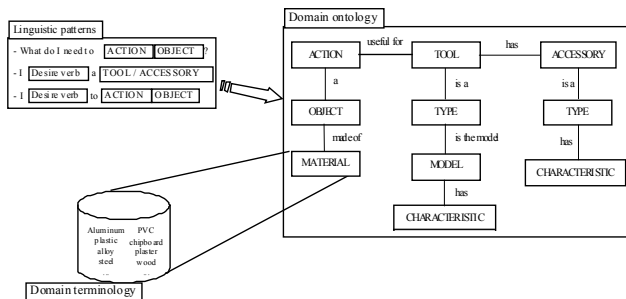

Figure 1. Sentence pattern extraction based on the domain ontology

Based on this ontology, a new corpus study phase was carried out. In this phase, the terminology of the sentences was substituted by the general concepts that they were representing, aiding in the detection of generic sentence patterns that could drive the development of a semantic grammar to perform the linguistic analysis (see figure 1).

3.2. Semantic Grammars

The object of semantic grammars is to interpret the user utterances in order to obtain a feature-typed semantic structure containing the relevant information of the sentences (Dale et al., 2000). These techniques are useful in specific domains that have a specific terminology that helps to drive the linguistic analysis, but have two serious drawbacks. First, the analysis with a semantic grammar tends to offer low coverage, no matter how wide is the domain corpus studied, as the language inherent richness causes that we will always find sentences that do not match any of the patterns we have built. To alleviate this problem, two additional analysis mechanisms have been added to the system. The second main drawback comes from the fact that when a grammar has been designed to cope as well as possible with the peculiarities of a certain domain, it is difficult to adapt to any other domain, even if it is a similar one. To lighten this problem, we have organised both the rules and the lexicon of the grammar in three different levels, depending if the sentence (or vocabulary) belongs to a general domain, to the commerce domain or to the specific shop domain. This kind of organisation can be useful on order to reuse as much as possible the work done when porting the system to a new domain.

At this moment, the grammar is made of 37 different sentence patterns and the lexicon contains 549 specific entries.

3.3. Extending Coverage

As was aforementioned, two extra analysis approaches were added to the system to complement the analysis made by the semantic grammar, in order to improve the coverage of the interpreter.

The first of them is a grammar rules relaxing technique, so that it is not necessary that the user sentences perfectly matches any of the patterns in the grammar, but it allows some parts of the sentence to be skipped if that allows the matching of the rest of the input. This allows the recognition of a bigger number of sentences, however if we skip some important words we might be losing relevant information and so precision in the result.

The other mechanism stems from the domain ontology extracted in the corpus study phase. This ontology puts together the basic concepts of the domain with the terminology that lexicalizes each of them. In this kind of systems it is essential that the user trusts the systems ability to understand him, as if this confidence is lost, the user will not use the system any more. Therefore, the set of concepts in the ontology should always be understood by the system, they can be considered as the base line of the analysis. An analysis has been implemented that everytime that a word representing a key concept of the domain appears, is able to retrieve the associated information, but is not able to retrieve anything that is not related to this words. The coverage of this analysis is very high, because most sentences contain any of these words, but the precision is quite low, as the sentence structure is being put aside. This analysis can be seen as a small evolution from the traditional keyword based analysis to a higher level concept based analysis.

Finally, the system has to choose one from the three, possibly different, answers that the analysis mechanisms produce. First of all, if the analysis using the original semantic grammar parses the user sentence, its result is chosen as the final result of the system, because the analysis with this grammar has a very high precision. If the grammar fails to analyze the sentence, the strategy chosen has been to combine the results obtained by the
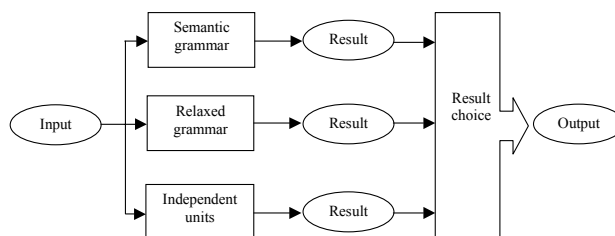

Figure 2. Natural language interpreter structure

two additional analysis methods, as the first of them focuses on the sentence structure and may skip some

important words and the second focuses on the important words, disregarding the sentence structure.

The global architecture of the module can be seen in figure 2.

4. Interaction Agent

Once the input information has been interpreted by the interface agent, it has to be classified, processed and, if necessary, stored. Then, proper actions will be performed by the system, which finally should be able to compose a sketch of the next intervention.

The interaction agent is going to observe three different sorts of information out of the semantic structures provided by the interface agent. First, it has to extract the data that shapes the circumstance, the so called static information. On second place, it should get the underlying intentions of the dialogue, this is, the dynamic information. Finally, it has to attend to the structure of the interaction, for attaining a valid dialogue.
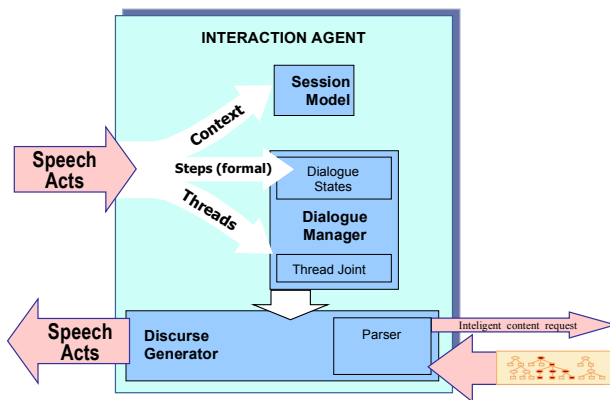

Figure 3. Overview of the dialogue manger structure

The main components of the interaction agent (figure 3) are lined up with this assortment. Hence, the 'Session Model' will deal with the context and the details of the interaction, while the 'Dialogue Manager' pay attention to the state of the interaction while checking the coherence with valid dialogue-patterns and thread management.

On a second division, the dialogue manager has two components. Firstly, it has to handle the steps taken by both interlocutors through the interaction. For this, it will heed the interaction as a game in which both players make their moves by turns (Levin et al., 1977; Poesio et al., 1998). The dialogue state component role is to 'shape' a valid dialogue that will help to understand upcoming user movements as well as it provides a set of adequate steps for the system to take.

The thread joint component stands for the intentional processing. In this matter, it will consider two different points of view: the user thread, and the system thread. Furthermore, it has a third point of view, the thread joint, that brings those conceptions together into a commitment ground for keeping the coherence of the interaction (Cohen et al., 1991).

The interaction agent will compose its own intervention given a state and a thread, the system discourse. For this aim, a third module is necessary: the discourse generator. It has to select a set of communicative acts that fits the requirement of the intervention and then to fill it up with the context that is provided by the session model. Eventually, it might need 'intelligent content'. In this case, it will construct a request to the Intelligent Agent, who will analyse the context and then provide a knowledge structure. This structure will be inspected by a module inside the discourse generator: the parser. Analysing it, the parser will acquire the new necessary context information. In this process, some events may occur, and could even originate new threads. This would force to restart the discourse generation.

5. Natural Language Generator

Finally, the system has to answer the user with an output according the information, created by the interaction agent and represented by a stream of communicative acts, which is presented to the user by the interface agent, using the available possibilities (avatar, GUI and natural language)

Concerning the generation of answers in natural language, a domain-specific template-based approach is proposed. The templates used to generate natural language answers to the user can require arguments to fill the slots in or not, in the case of agreements, rejections or topic movements.

In the available prototype, sentences do not contain issues concerning a possible user model, although the templates are prepared to cover them. For example: some templates include a special argument (ExpertiseLevel) that causes different levels of explanation in the system answers displayed to the user. These explanations are in a glossary containing each domain term (tool class, accessory and so on) together with its different explanations. Moreover, each template includes several possibilities of answer. Thus, the system does not generate always the same answer under the same conditions in order to achieve natural dialogues.

The design of this module makes impossible an evaluation of its processing. The sentence templates are built ad-hoc for the necessities of the interaction agent, therefore, no intelligent processing is made, it is an automatic process in which one of the appropriate templates is filled with the required information and directly presented to the user.

6. Evaluation

To carry out the evaluation of the system, five people of the research team who had not been involved in the system development were selected. Each of them was given two scenarios, one in which the user had to buy a certain product, and another one in which the user has to acquire any accessory valid for the product that was previously bought and in both cases they freely played the role of the buyer. The behavior of the system was traced and later analyzed.

6.1. Natural Language Interpreter Evaluation

Along the system development, two separate evaluation processes were performed, the first one after the development of the semantic grammar and the second after the addition of the extra analysis mechanisms. The results obtained confirm the statements made about the low coverage of the semantic grammar and its improvement with the additional methods developed. In the aforementioned ten dialogues, there were a total of 68 user utterances, having one or more sentences per utterance. The longest dialogue has 14 interactions and the shortest has 3.

The results of the module parsing each sentence were classified, depending on its correctness into four different categories:

- P: the result was perfect.
- Par: the result was correct, but not complete, it was only partial.
- N: the module is not able to parse the sentence and, therefore, no result is given.
- W: the module returns a wrong answer.

Having an "W" a a result must be considered as a serious mistake. It points out a fault in the module design. However, obtaining "N" or "Par" as result does not announce a mistake, but a lack of linguistic or domain knowledge to parse the sentence. As the module has been designed keeping in mind the ease of enlargement, adding the knowledge to turn this "Par" and "N" results into "P" results is an affordable task. Finally, the "P" results mean that all the relevant information in the user utterance has been correctly extracted, and is the aim to reach with every sentence.

As can be seen in table 1, the module has evolved from being able to recover correct information (addition of the "P" and "Par" results) 60.29% of the times in the first evaluation to 77.93% in the second, so the coverage of the module has clearly grown. Moreover, as the number of correct and complete answers ("P" results) has also improved, it can be stated that the precision has also been positively affected by the additional analysis mechanisms.

Result	Eval. 1	Eval. 2	Difference
P	30.88 % (21)	41.17 % (28)	+ 10.29 %
Par	29.41 % (20)	36.76 % (25)	+ 7.35 %
N	38.23 % (26)	19.11 % (13)	- 19.12 %
W	4.41 % (3)	2.94 % (2)	- 1.47 %

Table 1. Figures of the evaluation

In general terms, the behaviour of the module can considered satisfactory, as nearly 80% of the times the module is able to extract useful information, so that the dialogue can go on without requiring too many confirmation or repetition steps from the user. Even though there was no real customers corpus available, the figures of the evaluation with project-independent people were quite positive, what leads us to think that the results could have been also very possitive if a large domain dialogue corpus had been available.

On the contrary, the techniques used in this module are highly domain dependent, and a change of domain would result in intensive work to adapt the grammar to the new domain, although some parts of the grammar may be reusable, depending on how similar is the new domain to the old one.

6.2. Dialogue Manager Evaluation

Our approach was to design an evaluation that could point out the strengths and weaknesses of the system, while being generic enough to be used in other dialogue systems.

First, the key points in the dialogue managing were identified and how this points could be assigned a measure was decided. The different functionalities in the dialogue manager were considered separately, that is, the static information processing, the dynamic information, the need of intelligent information and, finally, the global performance.

Regarding the performance, an interesting value to focus on is the amount of information, which can uncover bottlenecks in the dialogue processing: the number of atomic pieces of static information relevant for each utterance, the number of formal steps that each user utterance causes, and the number of active dialogue threads in each intervention.

Item	Value
Pieces of static information active per utterance	2.98
Simultaneous active threads	1.08
Formal steps per utterance	1.78
Interactions needing intelligent assistance	36 %
Time spent per system utterance	0.278
Number of interactions for achieving a goal	9.67
Unsuccessful discourses	6.3 %
Buggy discourses	0.14%

Table 2. Average values of the evaluation

The processing of static information seemed to be slightly heavier than the other two, yet efficient enough. However, there is another important matter to focus on: the need of intelligent information.

The need of knowledge based content proved to be a critical point since it slows down the response. Hence, the number of interactions in a dialogue that need this kind of processing is also observed, revealing the need of improvement.

On the other hand, some figures are needed to disclose the quality of the interaction and the overall behavior. Apart from the number of interactions needed for achieving the goal and the average answering time, there should be observed the interventions that do not make progress in the dialogue. Even more, there should be differentiated between those interventions without progress (unsuccessful discourses) and those that get the dialogue backwards (buggy discourses). These results might open the need of addition of dialogue corpus for the specific domain.

Figure 4. General view of the system interface

The final average values of the evaluation with the set of ten dialogues can be seen in table 2.

7. Conclusions

The paper included the work done in the development of the prototype in the ADVICE project (IST-1999-11305).

One of the key points in the virtual assistant was to generate trust in the user by understanding the user sentences and being able to answer in a useful, coherent way. The values obtained in the evaluations prove that a minimum achievement of this objective has been reached.

Moreover, the uniform management of the information (Martínez et el, 2000) based on communicative acts allows a multimodal interaction, as the input does not need to come from the natural language interpreter, but may come from any of the other interface modules.

In any case, the proposed method for the natural language understanding is limited by its design, as a semantic grammar cannot be generic enough to parse an unrestricted language. Therefore, in the future, it will be necessaaary to design a more robust kind of analysis, including the morphological and syntactic levels, and based on adaptation of existing linguistic resources. (García-Serrano et al., 2001; Miller et al., 1990).

An example of this integration can be seen in figure 4, in which natural language, graphic user interface and avatar coordinate to answer the user.

However the system can be impoved in several ways when a real, big sized corpus, that makes the conclusions extracted from its study as realistic as possible. This would have direct consequences both in the dialogue interpreter and the dialogue manager.

8. References

R. Dale, H. Moisi, H. Somers (editors). Handbook of Natural Language Processing, Part I: Symbolic approaches to NLP. Marcel Dekker Inc. 2000.

Martínez, P. and García-Serrano, A., The role of knowledge-based technology in language applications development. Expert Systems with Applications 19 , 31-44, 2000

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., Miller, K., Introduction to WordNet: An On-line Lexical Database. (Revised August 1993). Princeton University, New Jersey, 1990.

J.R. Searle. Speech Acts: an essay in the philosophy of language. Cambridge Univ. Press., 1969.

J.A. Levin and J. Moore. Dialogue games: Metacommunication structures for natural language interaction. Cognitive Sc. 1(4), 395-420. 1977.

J.C. Kowtko, S.D. Isard, and G.M. Doherty. Conversational Games Within Dialogue, Rsrch P. 31, Human Comm. Research Centre, Edinburgh. 1993.

M. Poesio and A. Mikheev. The Predictive Power of Game Structure in Dialogue Act Recognition: Experimental Results Using Maximum Entropy Estimation. Proceedings of ICSLP-98, Nov 1998.

P.R. Cohen and H.J. Levesque. Confirmation and Joint Action, Proceedings of International Joint Conf. on Artificial Intelligence, 1991

Cole, R.A. ed. (1997) Survey of the state of the art in Human Language Technology.

Young, R. M., Moore, J. D., Pollack, M. E., (1994). Towards a Principled Representation of Discourse Plans. Proceedings of the Sixteenth Conference of the Cognitive Science Society, Atlanta, GA, 1994.