Website Interaction with Text-based Natural Language Dialog Systems

Markus Berg and Antje Düsterhöft

Hochschule Wismar
University of Applied Sciences: Technology, Business and Design
Department of Electrical Engineering and Computer Science
{markus.berg,antje.duesterhoeft}@hs-wismar.de

Abstract

In this paper we describe the extension of an existing web-based booking interface in the tourism domain with a natural language interface. This allows the user to interact with the system in form of a written natural-language-based dialog. The main focus lies in a user-centered, intuitive dialog design, which allows the system to guide the user effectively. To accomplish this, a dialog manager is designed and implemented as a prototypical text-based application, which gives valuable information for further research. The results of this work show that it is possible to improve today's systems, yet revealing linguistic challenges, which will be addressed in future research.

1 Introduction

Voice User Interfaces (Cohen 2004) today have been evolved to a major alternative to Graphical User Interfaces and thus are an important variant of Human Computer Interfaces (Fischer 2001). Many companies take advantage of using Voice Portals as a cheap and effective way of supporting customers. Besides spoken language also written language becomes popular when interacting with the computer. So called *chatbots* or *agents* appear on many websites. But most of those systems don't support natural interaction, i.e. users are not able to formulate their concerns freely. In this paper we are going to extend an existing tourism web-portal by a linguistic input possibility. In contrast to common systems, we are going to provide a mixed-initiative approach, which allows the user to control the dialog and allows the system to guide the user.

After describing related work, key characteristics of tourism portals are analyzed. The fourth chapter gives a short description of dialogs with focus on psychological influences and linguistic challenges. In the realization part the extraction of relevant information from user utterances, its interpretation and the dialog flow are described. Finally the results are summarized and an outlook on future work is given.

2 Related Work

The theoretical foundation of this work is chatbots: interactive, text-based dialog systems. ELIZA (Weizenbaum 1966) is the most famous chatbot and was developed by Joseph Weizenbaum. A similar program which bases on ELIZA is the A.L.I.C.E.-Bot. It is developed in the XML language AIML (Artificial Intelligence Markup Language), which includes decomposition templates and reassembly rules as Weizenbaum defined more than 40 years ago. As those systems are only reacting (user-initiative), they don't help to reach a goal. That's why such systems cannot be used for supporting users. Today many so called agents emerge on the market. You can ask them questions concerning the products of the company. Mostly they extract the keywords of the sentence and redirect you to the respective webpage. Instead of formulating the answers in form of a sentence, they are displayed on the website. These systems are not able to process complex information like *"Show me all white shelves!"* or *"How much is 'Ekby Järpen'?"*. One of the state-of-the-art-systems is *IKEA Anna*, which helps you to get to the right product category. In comparison, this is relatively simple, as there is only one semantic value: the product category. Other approaches include machine-learning-algorithms (Xu et al. 2009) or concentrate on open domain question answering features (Artstein & Vieu 2007).

3 Tourism Portals

Today's most common way of planning holidays is – apart from going to the travel agency – the gathering of information from the internet. Many forms with many questions have to be filled in before you are able to get more information. Often users forget to really input all information, which results in error messages. Especially elder people are not used to work with computers. But exactly this part of society spends money in the field of tourism. That's why it makes sense to develop systems which are easy to control by this target group. Language is the most effective natural form of exchanging information. Everyone is used to it and you don't have to explain how it works.

When comparing common touristic web portals one can identify many similarities. The following attributes exist on nearly every website:

- Earliest possible begin of journey & latest possible end of journey
- Trip length
- Age of children (implicitly number) & number of adults
- Destination

These attributes are at least necessary to create a query to retrieve information from the database. Furthermore there are facultative attributes like hotel-class, type of accommodation, name of hotel and fare. These statements help to reduce the result set if too many hits have been found. Due to the number of required information and the lack of

linguistic capabilities and thus the non-existence of natural dialogs, current chatbots are seldom used to query complex information.

4 Dialogs

A Dialog is the exchange of information between two partners. All dialog forms have the following commonalities: the existence of a sender, a receiver and a channel. The transmission channel (seeing, hearing, feeling) is not defined. This leads to various ways of communication, e.g. GUI-based-dialogs and language-based dialogs. The latter will be topic in this paper. Here we comprise spoken and written language by regarding written language as a part of the transformation process from voice into an abstract form.

Dialogs can be differentiated and described by several aspects. The most common form of categorizing dialogs is the *initiative*. It can be described as *the dialog partner who has influence on the control of the dialog*. Depending on who owns this control, one speaks of system-initiative, mixed-initiative and user-initiative dialogs (Carstensen 2009; Hauser 2000). The *system-initiative* dialog today is the most often used form of communication in Interactive Voice Response systems, so called Voice Portals. The user has no ability to speak freely or to take influence on the dialog flow. It's just a question-answer-sequence which is unnatural but robust. *Mixed-Initiative* is state of the art and is used by more and more companies. The user has the ability to change the dialog flow through his own activities. The system adapts to the user and not vice versa. A *user-initiative* dialog doesn't lead the user to his goal at all. It's just a reactive system like ELIZA and can therefore only be used for entertainment tasks, but not for business applications.

Mixed initiative consequently is the most appropriate way for designing dialogs. As already said, the user has the ability to influence the dialog. But there are other attributes which are strongly connected with this term. An open question (*"How may I help you?"*) is one of the most important features to really allow the user to influence the dialog and to speak freely, i.e. naturally. Depending on what the user says, the next dialog step is chosen dynamically. This requires the processing of more than single words but whole sentences. This can be realized with multi-slot-filling. It allows the user to give more input than is asked for:

> *"Where do you want to start from?"*
> *"From Berlin at 9:00 pm"*

In this case, a question for the departure time will be skipped, which as well modifies the dialog flow. Features like Barge-In, Random Prompting, Caller Adaptive Pronunciation or Implicit Correction (Voice Objects 2010; Herda et. Al. 2010) are necessary to create a NLU-system, but the mere existence of some of the above techniques is not enough. A NLU-system comprises the whole process of being able to interpret natural utterances and to answer in an appropriate way. By using natural dialog techniques you still don't have a NLU-system as this includes more than dialog design but linguistic methods of processing syntax, semantics and pragmatics.

4.1 Psychological Influences

Speech Processing is a multi-disciplinary field of research. Besides Theoretical Computer Science, Computer Linguistics and Signal Processing also Psychology is an important part as it helps to understand the user and facilitates his manipulation (Watzlawick 2007; Whorf et al. 1984). With an intelligent way of formulating a question, you can lead the user to react in a certain way (Pinker 1998). When a user's reaction is predictive, it is much easier to process it. Thus dialog design influences speech understanding and should not be underestimated (Meggle 1993). Imagine a dialog which aims at retrieving a city name:

"*Where do you want to travel to?"*
"*To France"*
vs.
"*To which city do you want to travel?"*
"*Paris"*

As you can see, the form of the question has direct influence on the answer. It's a matter of fact that complexity increases along with naturalness. Along with that, natural systems animate the user to forget that he is talking with a machine. This results in a more colloquial – more human one might say – form of speaking, which influences processing a lot. When I ask clear and short questions I will get crisp answers. When I ask colloquial or unsharp questions I will get answers of the same style. Men are influenced psychologically. Some dialog designers have the opinion to make use of this and claim the best way would be to ask simple questions to get easy processable answers.

"*How many adults?"*
"*Two!"*
vs.
"*How may I help you?"*
"I want to travel to Lisbon with my wife."

The less liberties a user has, the easier processing is. Nevertheless it is worth to develop natural language dialog systems and to address the challenges described in the following section.

4.2 Linguistic Challenges

In the field of NLU the most demanding part is the interpretation of the user's aim. A question like "*Can you open the window please"* should not be answered with yes or no. It's an indirect appeal. The interpretation of the user's intention is called pragmatics. Men are able to distinguish between ambiguous phrases because of their experience. Machines have to weigh which interpretation is more likely. This is only possible with statistics and machine learning methods. Another interesting question is the sense of using syntactic analyses. On the one hand it helps to structure the sentence and to refer words to each other, which allows the resolutions of references (anaphora resolution):

"*I want to book a room$_i$. It$_i$ should have…"*

On the other hand users don't speak grammatically correct: They might use colloquial language or they might be non-native speakers. But a system should be available for many persons and not only for those who speak perfect English. That's why a syntax-analysis is only useful in a limited way. A misused article or a subject-verb-disagreement don't prevent understanding and should not result in an error message. The following sentence is syntactical wrong but it is understandable by men without any problem:

"I want to book a hotels with two bed."

Negations are another important aspect. Consider the sentence:

"I want to go to Spain, except of Mallorca."

If the system is not able to understand the meaning of "except" the system probably would only search for hotels in Mallorca, as this location replaces Spain because it is said later in the sentence, so the system could interpret it as a correction like:

"I want to go to Spain, um, I mean Mallorca."

If the system is not able to resolve negations, the sentence

"I don't want to have a smoker's room."

would result in smoker rooms only, although the user wanted the opposite. As you can see, syntax is important, but not every word of the sentence has to be categorized. Chunk parsing and heuristic methods can help to reduce the workload.

5 Realization

The heart of the system is the Dialog Manager, which is responsible for the whole workflow. This is the part, where the dialog flow is modeled by the Dialog Designer, who takes care of a user-friendly interface and uses a combination of psychological and technical knowledge. Beforehand it is important to understand the user input. Different parsers analyze the user input in a three-level-hierarchy. If the sentence has been analyzed and transformed into a machine-processable form, a query is generated which is send to the booking system. Depending on the number of results, the system asks for more wishes or produces an error message. The results are being displayed on the website.

The base for an intelligent dialog management is the extraction of the relevant parts of the user input. It is necessary to interpret and transform them into a machine processable form. The user input "*I want to travel with my wife*" has to be transformed to the information "2 adults". This process includes the following steps: Removal of stop words, numerical normalization, keyword spotting, context analysis and a multi-level parsing system. The different notation forms of dates have also to be considered. Intelligent adding of information transforms 3rd July to 3rd July 2010 as this is the next 'occurrence' of July. Another point is linguistic descriptions of dates like "tomorrow" and "eastern" or fuzzy information like "next

week" and "in August". They have to be resolved to 'real' dates. Furthermore there are answers which don't exactly answer a question:

"How many children do you have?"
"I have no children" vs. "0"

You can see that it is not enough to only process numbers (although one might expect only numerical answers). Often a linguistic expression represents a number. This is the same with "both of my brothers" which represents "two". To extract relevant information, regular expressions are used. They are more powerful than a simple keyword-matching-procedure as they focus also on the context and allow an efficient heuristic syntax analysis. With the help of the knowledge base, which is realized as an OWL-based ontology, the extracted parts are being referred to semantic categories like accommodation type or number of adults. This process is called frame-filling and represents the syntactic and semantic analysis on a heuristic base. Furthermore the ontology holds the language model containing synonyms and hierarchical structures. This allows the system to realize that "car parking" is more general than "garage" or to detect that "New York" and "New York City" refer to the same place. Also transitive inferences are possible: $P(x,y) \cap P(y,z) \rightarrow P(x,z)$. If pets are allowed and a dog is a pet, then dogs are allowed, too.

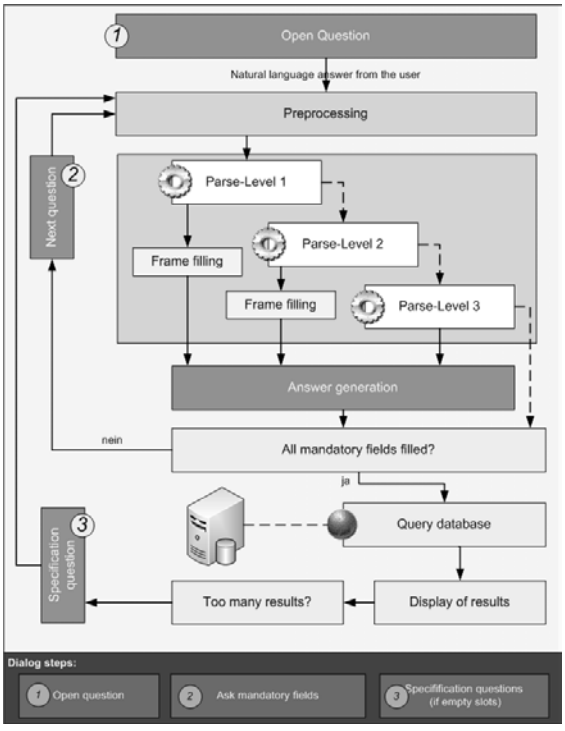

Figure 1: Dialog flow and processing steps

The dialog consists of three main parts: At the beginning a global parser extracts all relevant information needed for the processing of the *open question*. In the second step it is checked if all *mandatory information* is given. This includes: Begin of journey, end of journey, number of adults and age of the children (if any). Otherwise a specific question is generated. The answer is then analyzed in a three-level-parser-mechanism which consists of the following steps, as you can see in figure 1:

- Specific Parser
- Global Parser
- Small talk parser

The specific parser searches for answers on the posed question. This is necessary because often an answer is only distinct in context of the question, as the following example demonstrates:

> *"How many persons take part in your voyage?"*
> *"Two."*

Without knowing the question, the answer could not be interpreted. That's why the system firstly assumes that the user answers the question. If the parsing fails, the global parser analyzes the user input. This parser searches for answers to any question. This is necessary to allow the user to have control on the dialog flow by not answering the question and instead giving different information. Because this analysis has no relation to the question (no context available) the answer from the last example could not be interpreted. Another important feature is that the system insists on unanswered questions but also processes answers which don't relate to the question, giving the user control over the dialog flow. This is called user-assistance because the system acts as a guide helping the user to reach his aim. This is also useful when it is necessary to correct some information:

> *"How many nights do you want to book?"*
> *"Two"*
> *"How many children take part in your travel?"*
> *"Oh, I rather want to stay for three nights."*
> *"Ok. And with how many children do you want to make your journey?"*

If the global parser doesn't find any relevant information, the user input is passed to the small talk parser. This parser extracts no relevant information and only exists for psychological/user-friendly reasons. Many people use polite conversation and greet the computer or ask questions like *"How are you?"*. Others just try to 'play' with the system and ask for the age or the name. In this situation the system answers the user's question and afterwards asks a relevant one to guide through the booking process. The system reacts on the user's utterance although it has no relevant input. This is very important for the naturalness. One of the most valuable aspects in marketing is that the user feels pleasant. Only if he feels comfortable he will use the system again.

These steps are repeated until all information is given. Subsequently a query is generated and sent to a web service, which returns the appropriate URL and the number of results. Now the

system presents the website with the results and gives the user the possibility to enter further information. This is the third and last step called *specification phase*. Because many people don't know what to say spontaneously, the system actively generates questions depending on the information already given. If an adult books a voyage for one person, the system might propose special single offers. Or it might ask for family offers if traveling with children. Additional information might be: Target group, themes, destination, accommodation features or room type. These attributes allow the user to also give special wishes like the need for an oven, lactose-free food or a waterbed. Some of the compulsory attributes can also be inferred automatically as a person who wants to have a tennis court will be connected with the tennis theme.

The following dialog shows a real working example of how this natural language based online travel agency can be used. It addresses indirect answering and correction facilities as well as small talk features:

> *"Hello, how are you? I'd like to go to Rio with my girlfriend."*
> *"Hi, I'm fine. How many children take part?"*
> *"I have no children."*
> *"When do you want to begin your travel?"*
> *"Oh, before I forget… I need a wellness area for my girlfriend"*
> *"Alright. And when do you want to start?"*
> *"We want to start on Mai 4th"*
> *"And when do you want to return?"*
> *"On Mai 14th"*
> *"Ok, I'm searching appropriate offers for you. Wow, so many results… Can you tell me the type of your accommodation?"*
> *"Oh, I just see I'd rather like to come back on Mai 15th."*
> *"Ok. And what form of housing do you wish?"*
> *"We would like to have an apartment."*

Compared to IKEA Anna and other systems, this approach is more complex as it retrieves information in many dialog steps. This requires an intelligent method of matching the input to semantic categories. This system has been implemented as a prototype and is now beta-tested in order to get real user input.

6 Conclusion and Future Work

In this paper an existing web information system of the tourism domain has been extended by a natural language interface. In contrast to common systems, the focus was set on a mixed-initiative dialog, which guides the user through the booking process. Other features, which make it more extensive in comparison to other electronic assistants are: Reaction on irrelevant utterances and personal allusions, randomized prompts to prevent monotony, high naturalness by an open question, resistance against user uncooperativity, correction of previously entered information, supporting/advising system, small talk capabilities, avatar

visually simulates dialog partner, reduces fear of contact and the system is able to process grammatical incorrect sentences. These characteristics underline the naturalness of this dialog system. The emphasis of this project is that the user has total control over the dialog. Nevertheless the system accompanies the user and helps him to reach his aim. The results of the first steps of this project are very promising and are also supported by a positive industry resonance, as this system will decrease support costs and increase sales by the introduction of a new sales channel.

Of course there are sentences, which are not yet interpretable. When it is asked for the return date, it is impossible to correct the departure date, as the specific parser detects a valid date and relates it to the question which means that it is interpreted as return date:

"When do you want to begin your travel?"
"We want to start on Mai 4th"
"And when do you want to return?"
"Ah, I'd rather want to start on Mai 5th"

Begin: 5/4/2010
End: 5/5/2010

The global parser comes only into action if the specific parser fails. Because the specific parser doesn't fail, as it recognizes a date, a misinterpretation is the result. That's why the parser has to be extended by a plausibility check. A further aspect we are currently working on is the interpretation of the return date in context of the departure date:

"When do you want to begin your travel?"
"We want to start on Mai 4th"
"And when do you want to return?"
"Two weeks later"

As you can see, the last answer relates to the first given answer and should be automatically converted into a date. Negations and suitable reactions are also a matter of research. The statement *"But I 'don't want to go to Rio"* should be answered with *"Where do you want to go then?"* instead of just ignoring the mistakenly as 'unimportant' classified information. There are more examples, which can't be processed yet. It is obvious that natural language is very extensive and often ambiguous. In the next steps these concerns are addressed. As semantic interpretation still is a manual process, it requires time and knowledge about user behavior, grammar and the domain. Another emphasis is the correction of (respectively the robustness against) typos.

Besides the extension of the chatbot-based system it is also planned to add a telephony module, which of course requires the processing of spoken language. This leads to new problems, e.g. the recognition of unexpected words. It is often the case, that out-of-grammar words are falsely recognized as in-grammar (nearest match). Furthermore garbage-rules decrease performance. That's why we are examining statistic speech recognition with domain specific corpora. Furthermore we are evaluating IVR hard- and software to realize the goal of adding a telephony platform. The existing prototype is very helpful for the future development because it allows the evaluation of user behavior and gives us important hints

concerning language style, grammar, usability and user acceptance. This is a prerequisite for building a corpus, which can later be used for the speech recognition in the IVR-system. The results of this analysis build the base for forthcoming advances.

Acknowledgements

We would like to thank our industry partner MANET Marketing GmbH for providing user statistics and the booking interface. This research project is supported by the European Funds for Regional Development (EFRE).

References

Artstein, R. & Vieu, L. (2007). *A Chatbot-based Interactive Question Answering System*. Decalog, Trento, Italy.

Carstensen, K.U. et al. (2009). *Computerlinguistik und Sprachtechnologie: Eine Einführung*. Heidelberg: Spektrum Akademischer Verlag.

Cohen, M. et al. (2004). *Voice User Interface Design*, Addison-Wesley.

Fischer, G. (2001). User Modeling in Human-Computer Interaction. *User Modeling and User-Adapted Interaction*, Volume11, Issue 1-2, 65 - 86

Hauser, R. (2000). *Grundlagen der Computerlinguistik – Mensch-Maschine-Kommunikation in natürlicher Sprache.* Heidelberg: Springer Verlag.

Herda, N. et al. (2010) *Natural Dialog Management: Wie Sprachportale mit möglichst natürlichen Dialogen gestaltet werden*, White Paper.

Meggle, G. (1993). *Handlung, Kommunikation, Bedeutung.*, Berlin: Suhrkamp.

Pinker, S. (1998). *Der Sprachinstinkt. Wie der Geist die Sprache bildet*., München: Knaur

Voice Objects (2001). *Interact Naturally With Your Callers*. White Paper.

Watzlawick, P. (2007). *Menschliche Kommunikation*, Bern: Verlag Hans Huber

Weizenbaum, J. (1966). *ELIZA - a computer program for the study of natural language communication between man and machine*, Communications of the ACM, 36-45.

Whorf, B. L. et al. (1984). *Sprache, Denken und Wirklichkeit*, Reinbek: Rowohlt Tb.

Xu, F. et al. (2009). *Gossip Galore: A Conversational Web Agent for Collecting and Sharing Pop Trivia*, Proceedings of ICAART 2009, Portugal, INSTICC Press.