

# **Evaluation of a Natural Language Dialog Based Web Navigation System--A Case Study**

Joyce Chai<sup>\*</sup>, Jimmy Lin<sup>+</sup>, Wlodek Zadrozny<sup>\*</sup>, Yiming Ye<sup>\*</sup>  
Margo Budzikowska<sup>\*</sup>, Veronika Horvath<sup>\*</sup>, Nanda Kambhatla<sup>\*</sup>, Catherine Wolf<sup>\*</sup>

IBM T. J. Watson Research Center<sup>\*</sup>  
30 Saw Mill River Rd.  
Hawthorne, NY 10523

MIT Artificial Intelligence Lab<sup>+</sup>  
545 Technology Square  
Cambridge, MA 02139

## **Introduction**

With the emergence of e-commerce, websites must accommodate both customer needs and business requirements. Menu driven navigation and keyword search provided by most commercial sites have tremendous limitations. There is no way to balance the current needs and intentions of a user with the business requirements of the site. Often, as a result, users are overwhelmed and frustrated by the lengthy interaction, because it's hard to precisely describe their intentions, e.g. buying "dark pants without cuffs". The solution lies, in our opinion, in centering electronic commerce websites around natural language and multimodal dialog. This claim is supported by results of a recent study we performed, and which we present in this paper.

## **Natural Language Dialog Based Information Access System**

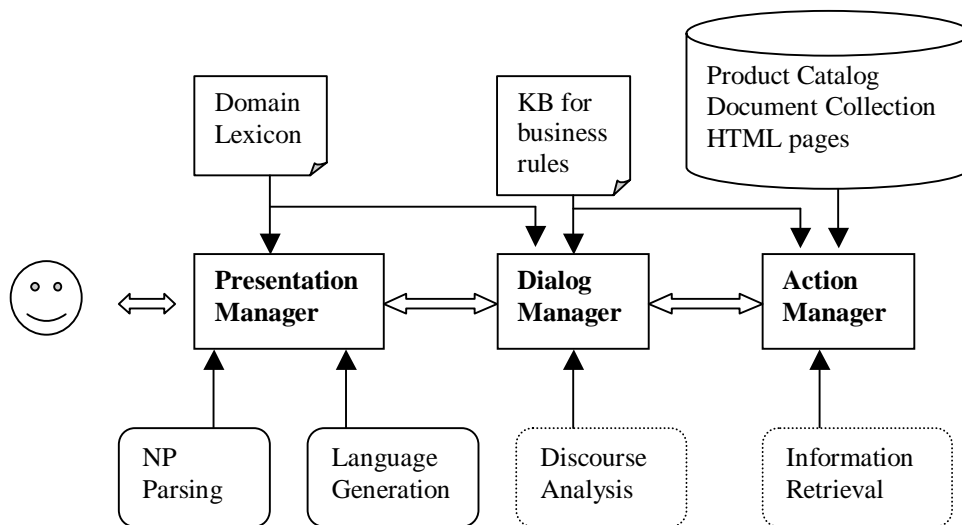
We have built and evaluated a NL dialog prototype system for guiding users towards computer products: PCs, notebooks, servers and services, currently sold by IBM. The system allows customers to make requests in natural language and be directed towards appropriate web pages that sell the product. Users can directly type what they are looking for in natural language. The system identifies and understands key concepts from the user's input. If there is a direct match from the user's concepts to the business rules, the system will show the web page that satisfies user's requests. Otherwise, the system will arrange dialog with the user to either ask for additional information or clarify the request. (Some screenshots will be shown in the full paper).

Our approach is novel in many aspects. First of all, it applies natural language dialog in an e-commerce environment to understand users request/intention and provide help/advice/recommendations to the user. It differs from websites such as [www.askjeeves.com](http://www.askjeeves.com), [www.neuromedia.com](http://www.neuromedia.com), which give a feeling of a "natural language interface" without using dialog with multiple turns. Often, when searching e-commerce sites, users do not know where to find or how to specify the desired information, although they have targets in their minds. Sometimes, users only have vague targets or no target at the beginning and need to formulate their interests/targets based on additional information. Those observations indicate that a dialog with the user is required in order to understand the user's cognitive state, intentions and interests, and to allow them to continually revise their search strategies.

Secondly, business rules are incorporated in the dialog management. This is an important point, since the system must balance business requirements with understanding of users request. We use XML to represent and manage Domain Lexicon (concepts) and Business Rules. The Domain Lexicon maps keywords onto concepts. A list of questions is associated with the concept, and used to elicit information for that particular concept during the dialog. Concepts can be as concrete as a category of computer, e.g. desktop-computer-concept, or as abstract as an idea, e.g. performance-and-value. A Business Rule consists of a list of concepts together with the weight that reflects how strong the business wants to push this rule. The dialog is arranged by matching concepts from the user's query to Business Rules. If a match is found, then a web page associated with that rule will be prompted to the user. Otherwise, based on the weights of partially matched Business Rules, the system finds the most important missing concept and asks a question to elicit descriptive responses from the user.

Furthermore, the architecture of the system supports multimodal dialog. For this case study, the system was implemented for the textual input. However, it can be extended to include inputs from different channels and modalities. The system consists of three major modules: Presentation Manager, Dialog Manager and Action Manager. The presentation manager is responsible for separating content from the presentation. In the case study, the presentation manager employs a natural language parser to transform user's natural language query into a logical form, and sends the logical form to the dialog manager. It is also responsible for obtaining the system's response from the dialog manager and presenting it to the user. The dialog manager is responsible for determining the specific action(s) requested by the user and filling the parameters (e.g. the attributes of the computers users are interested) of the identified action by way of a dialog with the user. The knowledge base for business rules specifies the translation from user requests to action plans for the action manager to satisfy the requests, for example, retrieving information about particular computer models from the product catalog.

The overall architecture of the the system is the following:



## Comparative Evaluation:

A user study has been conducted to give the comparative evaluation of the natural language dialog based prototype system and a fully developed menu driven system. The menu driven is based on a question-answer paradigm where the user has to choose an answer from a list of answers to a particular question. One answer will lead to another question with a set of answers. After a sequence of choices, the user will be directed to a web page that describes the product information.

An independent testing agency recruited a total of 17 people. They tested the natural language dialog based system and the menu driven system. A screener was used to recruit the participants. Among those participants, four of them considered themselves with advanced computer skills, eight with intermediate level of proficiency and five with limited experiences with internet.

A testing room was set up with a division to allow for the respondent and the moderator to work at one monitor while we were positioned behind the divider at another monitor. The systems were linked in such a way that we could manipulate the Happy Assistant prototype if necessary and observe the testing as it occurred without interfering with the interview. Such manipulation was intended only for some fatal errors like infinite loops or unexpected exceptions, and were rarely used. In the testing, it turned out less than 5% of interactions were intervened.

Each interview began with an introduction by the moderator explaining the purpose of the interview. It was explained that they would be using two prototype web sites. In addition, they were informed of the moderator's independent and objective position and encouraged to be open with their opinions of the prototype. After the introduction, the participants were given various scenarios. These scenarios were designed to let them experience critical parts and navigation of each web site in order to form an opinion of the tool's concept. They were then asked to rank the scenarios on a 1 to 10 scale (where 10 is easy) with regards to the ease of navigation and the series of events leading up to the result. The moderator probed the participant throughout the scenario, and the participants were asked for their overall reaction to the concept of each prototype upon completion. In total, six scenarios were used for the testing. For each scenario, there were two similar versions presented: one for the natural language system and the other for the menu driven system. Each participant was randomly assigned three scenarios.<sup>1</sup>

The user study shows that users prefer the natural language based navigation two to one over the menu driven navigation. Comparing NL dialog based navigation with the menu driven navigation, the number of clicks were significantly reduced by 63.2% (indicated by the T-test,  $P < 0.0005$ ), the time was significantly reduced by 33.3% (indicated by the T-test,  $P < 0.025$ ). Users liked the fact that they could express their needs in their "own jargon" instead of the foreign "computer jargon". There was also the perception that with the natural language dialog based model, the computer did all the work for them instead of them doing all the work for the computer (as in the menu-driven model). Reactions from users with different online experiences varied. The less experienced user prefers the NL enabled navigation much stronger than the experienced user. After each task, we asked participants to rate two systems as to how easy they felt in using it. For the ease of use rating (1-10, 10 is the easiest, 1 is the most difficult), from subjects with limited internet experience, the average rating for the natural language based system is 9.4, for the menu driven

---

<sup>1</sup> An example of a scenario is: one version: "Your daughter's birthday is coming soon. Currently she is a high school student and you would like to get her a computer for birthday. You would also like to have internet access at home and occasionally play some computer games with your kids," and the other version: "You have just moved into a new house and would like to get a computer for your home. You are interested in internet access to check email, browse websites, and trade online from home. Occasionally, you would also need to write letters."

system is 6.3; from the users with intermediate experience, the average rating for the natural language based system is 8.5, for the menu driven navigation is 8.1; from the users with advanced experience, the average rating for the natural language based system is 8.3., for the menu driven navigation is 8.9. We have learned that in order to accommodate users with different experiences, natural language and menu driven should be intelligently combined. More discussions about findings and observations will be given in the full paper.

## **Conclusion and Future Work**

The goal of NL processing is full understanding of users' intentions. The choice of appropriate technology depends on both its availability and reliability, and on suitability for a particular application. In this study we have shown that natural language dialog is something customers want, and that sophistication in dialog management is more important than the ability to handle complex natural language sentences. In particular, to provide an easy access to information on e-commerce sites, natural language dialog based navigation and menu driven navigation should be intelligently combined to satisfy user's different needs.

Our current work, extending the results of this paper, includes enhancement (which are also be rigorously tested) in language analysis and dialog management, automatic learning business rules and automated ontological mapping between customer terms and business terms. More details will be described in the full paper.

We leave it to another paper to discuss how business processes and catalog design would have to be changed to accommodate natural language/multimodal dialog. We will provide some illustration of these business issues in the full version of the paper.