

A Conversational Interface for Online Shopping

Joyce Chai, Veronika Horvath, Nanda Kambhatla, Nicolas Nicolov & Margo Stys-Budzikowska

Conversational Dialog Systems
IBM T. J. Watson Research Center
30 Saw Mill River Rd, Hawthorne, NY 10532, USA
{jchai, veronika, nanda, nicolas, sm1}@us.ibm.com

ABSTRACT

We present a deployed, conversational dialog system that assists users in finding computers based on their usage patterns and constraints on specifications. We discuss findings from a market survey and two user studies. We compared our system to a directed dialog system and a menu driven navigation system. We found that the conversational interface reduced the average number of clicks by 63% and the average interaction time by 33% over a menu driven search system. The focus of our continuing work includes developing a dynamic, adaptive dialog management strategy, robustly handling user input and improving the user interface.

1. INTRODUCTION

Conversational interfaces allow users to interact with automated systems using speech or typed in text via "conversational dialog". For the purposes of this paper, a conversational dialog consists of a sequence of interactions between a user and a system. The user input is interpreted in the context of previous user inputs in the current session and from previous sessions.

Conversational interfaces offer greater flexibility to users than menu-driven (i.e., directed-dialog) interfaces, where users navigate menus that have a rigid structure [5,4]. Conversational interfaces permit users to ask queries directly in their own words. Thus, users do not have to understand the terminology used by system designers to label hyperlinks on a website or internalize the hierarchical menus of a telephone system [3] or websites.

Recently, conversational interfaces for executing simple transactions and for finding information are proliferating [7,6]. In this paper, we present a conversational dialog system, Natural Language Assistant (or **NLA**), that helps users shop for notebook computers and discuss the results of user studies that we conducted with this system.

2. NATURAL LANGUAGE ASSISTANT

NLA assists users in finding notebooks that satisfy their needs by engaging them in a dialog. At each turn of the dialog, **NLA** provides incremental feedback about its understanding of the user's constraints and shows products that match these constraints. By encouraging iterative refinement of the user's query, the system finds more user constraints and, ultimately, recommends a product that best matches the user's criteria.

The system consists of three major modules (cf. Figure 1): Presentation Manager, Dialog Manager, and Action Manager. The Presentation Manager interprets user input and generates system responses. It embodies the user interface and contains a shallow semantic parser and a response generator. The semantic parser

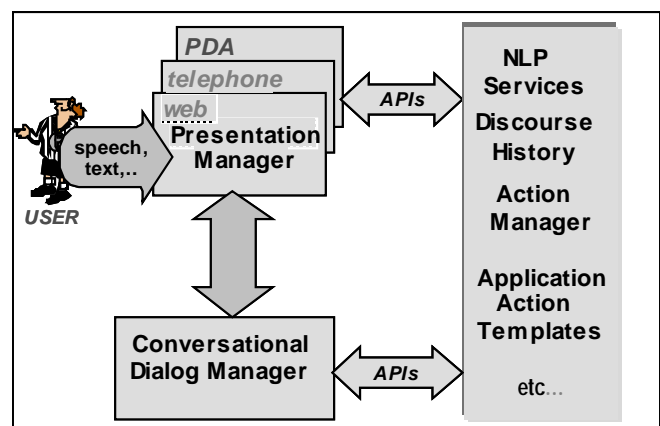


Figure 1. Architecture of the NLA conversational system.

identifies concepts (e.g., MULTIMEDIA) and constraints on product attributes (e.g., *hard disk size more than 20GB*) from the textual user input. The concepts mediate the mapping between user input and available products through product specifications. They implement the business logic.

The Dialog Manager uses the current requirements and formulates action plans for the Action Manager to perform back-end operations (e.g., database access¹). The Dialog Manager constructs a response to the user based on the results from the Action Manager and the discourse history and sends the system response to the Presentation Manager that displays it to the user. The system prompts for features relevant in the current context. In our mixed initiative dialog system, the user can always answer the specific question put to him/her or provide any constraints.

The system has been recently deployed on an external website. Figure 2 shows the start of a dialog.²

¹ See [1] for a survey of natural language interfaces to databases.

² We are demonstrating the system at HLT'2001 [2].

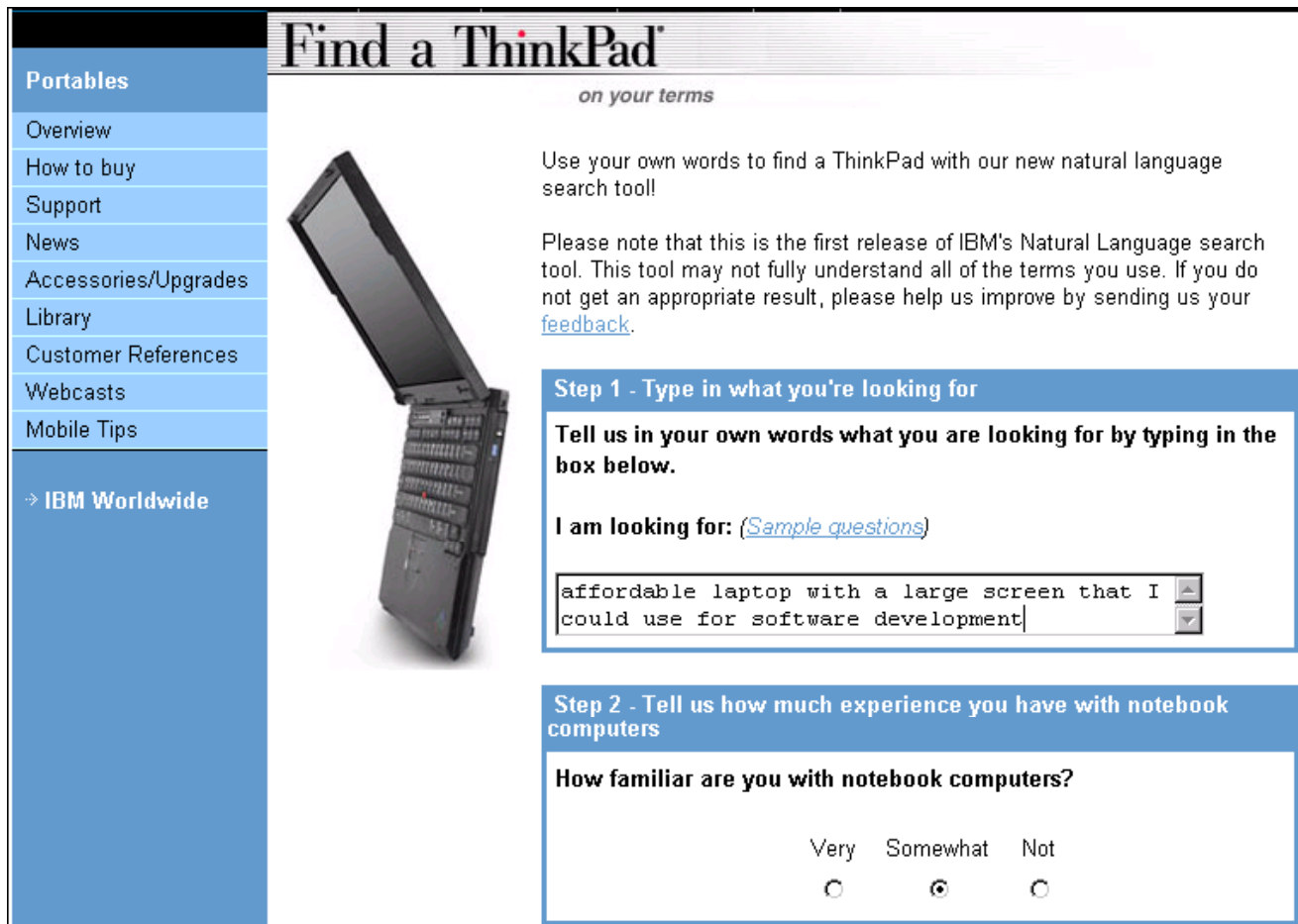


Figure 2. The start of the dialog.

3. USER STUDIES

We conducted a preliminary market survey and two user studies described in subsections 3.1 and 3.2 respectively.

3.1 Market Survey

For understanding specific user needs and user vocabulary, we conducted a user survey. Users were given three sets of questions. The first set, in turn, contained three questions: "What kind of notebook computer are you looking for?", "What features are important to you?", and "What do you plan to use this notebook computer for?". By applying statistical n-gram models and a shallow noun phrase grammar to the user responses, we extracted keywords and phrases expressing user's needs and interests. In the second set of questions, users were asked to rank 10 randomly selected terms from 90 notebook related terms in order of familiarity to them. The third set of questions asked for demographical information about users such as their gender, years of experience with notebook computers, native language, etc. We computed correlations between vocabulary/terms and user demographic information. Over a 30-day period, we received 705 survey responses. From these responses, we learned 195 keywords and phrases that were included in NLA.

3.2 Usability Testing

3.2.1 Experimental Setup

We conducted two user studies to evaluate usability of the system, focusing on: dialog flow, ease of use, system responses, and user vocabulary. The first user study focused on the functionality of NLA and the second user study compared the functionality of NLA with that of a directed dialog system and a menu driven navigation system.

The moderators interviewed 52 users in the user studies: 18 and 34 in the two studies, respectively. All participants were consumers or small business users with "beginner" or "intermediate" computer skills. Each participant was asked to find laptops for a variety of scenarios using three different systems (the NLA, a directed dialog system and a menu driven navigation system). Participants were asked to rate each system for each task on a 1 to 10 scale (10 – easiest) with respect to the *ease of navigation*, *clarity of terminology* and their *confidence in the system responses*. The test subjects were also asked whether the system had found relevant products and were prompted to share their impressions as to how well the system understood them and responded to their requests.

3.2.2 Results

In both studies, participants were very receptive to using natural language dialog-based search. The users clearly preferred dialog-based searches to non-dialog based searches³ (79% to 21% users). Furthermore, they liked the narrowing down of a product list based on identified constraints as the interaction proceeded. In the first user study, comparing NLA with a menu driven system, we found that using NLA reduced the average number of clicks by 63% and the average interaction time by 33%.

In the second user study, we compared NLA with a directed dialog system and a menu driven search system for finding computers. One goal of the comparative study was to find out if there were any statistical differences in *confidence*, *terminology* and *navigation* ratings across the three systems and whether they were correlated with different categories of users. The ANOVA analysis reveals statistical differences in terminology ratings among the three systems for the category of beginner users only. There were no statistical differences found in the other ratings of *navigation* and *confidence* across the three sites for different categories of users. Sandler's A test confirmed that the terminology rating was significantly different for the categories of *consumers*, *small business owners*, *beginners* and *intermediates*. These comparative results suggest that asking questions relative to the right level of end user experience is crucial. Asking users questions about their lifestyle and how they were going to use a computer accounted for a slight preference of the directed dialog system over the NLA that uses questions presented on the basis of understanding features and functions of computer terms.

3.2.3 Lessons from the user studies

Both user studies revealed several dimensions along which NLA can be improved. The first user study highlighted a definite need for system *acknowledgement* and *feedback*. The users wanted to know whether the system had understood them. User comments also revealed that a *comparison* of features across the whole pool of products was important for them.

The focus of the second study, incorporating 34 subjects, was to compare systems of similar functionality and to draw conclusions about the functionality of NLA. Both the ANOVA and the Sandler's test point out that terminology was a statistically significant factor differentiating among the systems. We believe that using terminology that is not overly technical would contribute to the success of the dialog search. While the questions asked by NLA were based on features and functionality of notebook computers, the users preferred describing usage patterns and life style issues rather than technical details of computers.

We also found that users' confidence in NLA decreased when the system responses were inconsistent i.e., were not relevant to their input. Lack of consistent visual focus on the dialog box was also a serious drawback since it forced users to scroll in search of the dialog box on each interaction page.

³ We define a dialog-based search as one comprising of a sequence of interactions with a system where the system keeps track of contextual (discourse) information.

3.2.4 Future work

Based on the results of the user studies, we are currently focused on: developing a dynamic and adaptive dialog management strategy, improving the robustness of the natural language processing (NLP), and improving the user interface. Some of issues mentioned here have been implemented in the next version of NLA.

We are currently re-designing the questions that NLA asks users to be simpler, and to focus on usage patterns rather than technical features. We are also implementing a new dialog management strategy in NLA that is more adaptive to the user's input, and implements a mapping from high-level usage patterns to constraints on low-level technical features.

We are integrating a statistical parser with NLA to more robustly handle varied user input. The statistical parser should enable NLA to scale to multiple languages and multiple domains in a more robust and reliable fashion. We are aiming at an architecture that separates the NLP processing from the business logic that will make maintenance of the system easier.⁴

Improvements to the GUI include better acknowledgement and feedback mechanisms as well as graphical UI issues. We now reiterate the user's last query at the beginning of each interaction page and also convey to the user an explanation of features incrementally accumulated in the course of the interaction. We have designed a more uniform, more compact and consistent UI. In the welcome page, we have abandoned a three-step initiation (typed input, experience level and preferences for major specifications) keeping the emphasis on the dialog box. The user preferences contributed to creating confusion as to the main means of interaction (many users just clicked on the radial buttons and did not use the full dialog functionality). We now infer the technical specifications based the user's stated needs and usage patterns. Our UI now has a no scrolling policy and we allow for larger matching set of products to be visualized over a number of pages.

4. DISCUSSION

In this paper, we have presented a conversational dialog system for helping users shop for notebook computers. User studies comparing our conversational dialog system with a menu driven system have found that the conversational interface reduced the average number of clicks by 63% and the average interaction time by 33%. Based on our findings, it appears that for conversational systems like ours, the sophistication of dialog management and the actual human computer interface are more important than the complexity of the natural language processing technique used. This is especially true for web-based systems where user queries are often brief and shallow linguistic processing seems to be adequate. For web-based systems, integrating the conversational interface with other interfaces (like menu-driven and search-driven interfaces) for providing a complete and consistent user experience assumes greater importance.

⁴ Many systems' fate has been decided not because they cannot handle complex linguistic constructions but because of the difficulties in porting such systems out of the research environments.

The user studies we conducted have highlighted several directions for further improvements for our system. We plan to modify our interface to integrate different styles of interaction (e.g., menus, search, browsing, etc.). We also intend to dynamically classify each user as belonging to one or more categories of computer shoppers (e.g., gamers, student users, home business users, etc.) based on all the user interactions so far. We can then tailor the whole interface to the perceived category including but not limited to the actual questions asked, the technical knowledge assumed by the system and the whole style of interaction.

Another area of potential improvement for the NLA is its inability to handle any meta-level queries about itself or any deeper questions about its domain (e.g., NLA currently can not properly handle the queries, "How can I add memory to this model?" or "What is DVD?"). Our long-term goal is to integrate different sources of back-end information (databases, text documents, etc.) and present users with an integrated, consistent conversational interface to it.

We believe that conversational interfaces offer the ultimate kind of personalization. Personalization can be defined as the process of presenting each user of an automated system with an interface uniquely tailored to his/her preference of content and style of interaction. Thus, mixed initiative conversational interfaces are highly personalized since they allow users to interact with systems using the words they want, to fetch the content they want in the style they want. Users can converse with such systems by phrasing their initial queries at a right level of comfort to them (e.g., "I am looking for a gift for my wife" or "I am looking for a fast computer with DVD under 1500 dollars").

5. CONCLUSIONS

Based on our results, we conclude that conversational natural language dialog interfaces offer powerful personalized alternatives to traditional menu-driven or search-based interfaces to websites. For such systems, it is especially important to present users with a consistent interface integrating different styles of interaction and to have robust dialog management strategies. The system feedback and the follow up questions should strike a delicate balance

between exposing the system limitations to users, and making users aware of the flexibility of the system. In current work we are focusing on developing dynamic, adaptive dialog management, robust multi-lingual NLP and improving the user interface.

6. REFERENCES

- [1] Androutsopoulos, Ion, and Ritchie, Graeme. Natural Language Interfaces to Databases – An Introduction, *Natural Language Engineering* 1.1:29-81, 1995.
- [2] Budzikowska, M., Chai, J., Govindappa, S., Horvath, V., Kambhatla, N., Nicolov, N., and Zadrozny, W. Conversational Sales Assistant for Online Shopping, Demonstration at *Human Language Technologies Conference (HLT'2001)*, San Diego, Calif., 2001.
- [3] Carpenter, Bob, and Chu-Carroll, J. Natural Language Call Routing: A Robust, Self-organizing Approach, *Proceedings of the 5th Int. Conf. on Spoken Language Processing*. 1998
- [4] Chai, J., Lin, J., Zadrozny, W., Ye, Y., Budzikowska, M., Horvath, V., Kambhatla, N., and Wolf, C. Comparative Evaluation of a Natural Language Dialog Based System and a Menu-Driven System for Information Access: A Case Study, *Proceedings of RIAO 2000*, Paris.
- [5] Saito, M., and Ohmura, K. A Cognitive Model for Searching for Ill-defined Targets on the Web - The Relationship between Search Strategies and User Satisfaction, *21st Int. Conference on Research and Development in Information Retrieval*, Australia, 1998.
- [6] Walker, M., Fromer, J., and Narayanan, S. Learning Optimal Dialogue Strategies: A Case Study of a Spoken Dialogue Agent for Email, *36th Annual Meeting of the ACL*, Montreal, Canada, 1998.
- [7] Zadrozny, W., Wolf, C., Kambhatla, N. & Ye, Y. Conversation Machines for Transaction Processing, *Proceedings of AAAI / IAAI - 1998*, Madison, Wisconsin, U.S.A. 1998.