Intelligent search engines

The increased use of Web resources has created a need for more efficient and useful search methods. The current mechanisms for assisting the search and retrieval process are quite limited, mainly because they lack access to documents’ semantics and because of the underlying difficulties in providing suitable search patterns.

Recent advances in intelligent search suggest that these limitations can be partially overcome by providing search engines with more intelligence and with the user’s underlying knowledge. In this sense, intelligence is seen as the ability of systems to interact with users by natural language dialog so that the engine can learn user profiles and likes. User behavior suggests that feedback in terms of natural dialog interactions can play a key role in decreasing information overload and getting accurate search results.

Smarter search engines. Intelligent searching agents have been developed to assist information retrieval systems. Agents can utilize spider technology used by Web search engines, but in new ways. Usually these tools are robots that are trained by the user to search the Web for specific information. The agent can be personalized so that it can build individual profiles or precise information needs. An intelligent agent can also be autonomous, so that it is capable of making judgments about the likely relevance of the material on its own.

To guide the Web search process, one promising method is to discover user preferences and needs by either extracting deep knowledge from what users are looking for or interactively generating explanatory requests to focus users on their interests. Although some research has been done using natural language processing (NLP) technology to capture users’ profiles, it has only been in very restricted domains that use general-purpose electronic linguistic resources to act on their requirements. In particular, using techniques for automatically generating natural language (NL) sentences allows the system to produce a useful dialog with the user and guide her or his preferences.

Designers of natural language generation (NLG) systems have strongly focused on generating natural language text and its contents at the discourse level, where complex tasks such as discourse planning play a key role in generating effective texts. When dialog processing involves managing dialog interactions (user-system), NLG systems are capable of capturing underlying knowledge, such as conversational turns (interactions), to provide replies according to the user’s knowledge and goals, to react to mistakes, and to deal with unexpected reactions from the user.

Natural language feedback. In recent years, a few approaches to intelligent Web search using natural language processing technology have emerged, mainly designed as question-answering systems. These address the problem of using linguistic processing on different levels to retrieve documents containing specific paragraphs in which target natural language queries are answered. So far, there is no dialog, and the effort is centered on obtaining accurate paragraphs (within documents), instead of capturing a user’s preferences.

As a part of a major interactive searching system, the model is based on task-dependent discourse and dialog analysis capabilities.

Figure 1 shows a model for intelligent searching and filtering using natural language feedback. The operation starts with natural language queries provided by a user (that is, general queries, general responses, feedback, and confirmation) and then passes them on to the discourse-processing phase, which generates the corresponding interaction turns (natural language output), arriving at a specific search request. As the dialog continues, the system generates a refined query that is sent to a search agent.

Natural language generation for dialogs. The dialog generator is based on a number of stages that state the context, participants’ knowledge (user and system), and goal of the interaction. It also consists of a set of modules for which input and output is delimited according to different stages of linguistic and nonlinguistic information extracted from the dialog. This dialog-processing component is based on state-of-the-art linguistic models for discourse processing.

The natural language generation component in Fig. 2 is capable of generating discourse outputs (that is, natural language utterances) from the results of a bibliographic Web search. This starts with the user’s input (natural language query) and produces either an output consisting of a natural language conversation exchange to guide the dialog and focus the user, or a search request that is passed to the search agent. In order to understand its underlying
Based on several samples obtained from experimental studies of users searching on the Web, basic initial criteria are extracted to restrict the natural language generation process, such as language, type of homepages, type of documents, and so on.

The natural language dialog generator can then produce two kinds of answers to explain the results of the search. One kind is for obtaining a more detailed specification of the user’s query, for example, “Your query is too general, could you be more specific?” The other kind requires the user to state some feature of the topic, for example, “Which language do you prefer?” The discourse analyzer again performs the analysis of the user’s specific answer in order for the search agent to perform a refined search. The search agent repeats the task, searching for the specific information on the topic in question.

At this point, the dialog analyzer processes the user’s response in order for the generator to produce an output confirming or expressing the action being performed. Whenever a user’s response is positive, the system will generate a sentence to give the user the opportunity to choose a new search topic. Otherwise, the dialog goes on. The overall process starts by establishing a top goal to build up the full structure in the sentence level.

**Adaptive search agents.** Unlike traditional search engines, the model for intelligent searching and filtering does not deliver all the information from Web search results to the user. Instead, the agent waits until it has sufficient knowledge about the user’s feedback and goals, which has a positive effect in terms of information overloading. As the interaction continues, the agent refines the requests and filters the initial information obtained from the user’s feedback, and then the search proceeds until a proper amount of information can be displayed (for example, 30 retrieved documents).

Using the knowledge obtained from user feedback, dialog samples, and current context information, the model detects the most frequent search patterns in vectorlike criteria, some of which involve the Web address of the page being selected, the author, language do you prefer?", the discourse analyzer again performs the analysis of the user’s specific answer in order for the search agent to perform a refined search. The search agent repeats the task, searching for the specific information on the topic in question.

At this point, the dialog analyzer processes the user’s response in order for the generator to produce an output confirming or expressing the action being performed. Whenever a user’s response is positive, the system will generate a sentence to give the user the opportunity to choose a new search topic. Otherwise, the dialog goes on. The overall process starts by establishing a top goal to build up the full structure in the sentence level.

**Adaptive search agents.** Unlike traditional search engines, the model for intelligent searching and filtering does not deliver all the information from Web search results to the user. Instead, the agent waits until it has sufficient knowledge about the user’s feedback and goals, which has a positive effect in terms of information overloading. As the interaction continues, the agent refines the requests and filters the initial information obtained from the user’s feedback, and then the search proceeds until a proper amount of information can be displayed (for example, 30 retrieved documents).

Using the knowledge obtained from user feedback, dialog samples, and current context information, the model detects the most frequent search patterns in vectorlike criteria, some of which involve the Web address of the page being selected, the author, language do you prefer?", the discourse analyzer again performs the analysis of the user’s specific answer in order for the search agent to perform a refined search. The search agent repeats the task, searching for the specific information on the topic in question.

At this point, the dialog analyzer processes the user’s response in order for the generator to produce an output confirming or expressing the action being performed. Whenever a user’s response is positive, the system will generate a sentence to give the user the opportunity to choose a new search topic. Otherwise, the dialog goes on. The overall process starts by establishing a top goal to build up the full structure in the sentence level.

**Adaptive search agents.** Unlike traditional search engines, the model for intelligent searching and filtering does not deliver all the information from Web search results to the user. Instead, the agent waits until it has sufficient knowledge about the user’s feedback and goals, which has a positive effect in terms of information overloading. As the interaction continues, the agent refines the requests and filters the initial information obtained from the user’s feedback, and then the search proceeds until a proper amount of information can be displayed (for example, 30 retrieved documents).

Using the knowledge obtained from user feedback, dialog samples, and current context information, the model detects the most frequent search patterns in vectorlike criteria, some of which involve the Web address of the page being selected, the author, language do you prefer?", the discourse analyzer again performs the analysis of the user’s specific answer in order for the search agent to perform a refined search. The search agent repeats the task, searching for the specific information on the topic in question.

At this point, the dialog analyzer processes the user’s response in order for the generator to produce an output confirming or expressing the action being performed. Whenever a user’s response is positive, the system will generate a sentence to give the user the opportunity to choose a new search topic. Otherwise, the dialog goes on. The overall process starts by establishing a top goal to build up the full structure in the sentence level.

**Adaptive search agents.** Unlike traditional search engines, the model for intelligent searching and filtering does not deliver all the information from Web search results to the user. Instead, the agent waits until it has sufficient knowledge about the user’s feedback and goals, which has a positive effect in terms of information overloading. As the interaction continues, the agent refines the requests and filters the initial information obtained from the user’s feedback, and then the search proceeds until a proper amount of information can be displayed (for example, 30 retrieved documents).

Using the knowledge obtained from user feedback, dialog samples, and current context information, the model detects the most frequent search patterns in vectorlike criteria, some of which involve the Web address of the page being selected, the author, language do you prefer?", the discourse analyzer again performs the analysis of the user’s specific answer in order for the search agent to perform a refined search. The search agent repeats the task, searching for the specific information on the topic in question.
this kind of interaction to decrease information overloading as well as the time spent by users searching the Web. The following is an example of a dialog between the system (S) and a user (U):

S: What are you interested in?  
U: I need basic information on linguistics.  
S: Your query is too broad; could you please be more specific?  
U: Fine.  
S: There are twenty references about that topic. Do you want to check all of them?  
U: Yes.  
S: Do you need further information about that topic?  
U: Yes.  
S: I found commercial pages. What kind do you prefer?  
U: I’d like to look at financial services.

The model was then assessed in terms of the average number of conversational turns in the dialog necessary to get a precise requirement, and information was filtered against the number of references/documents that matched these requirements. Initially, the set of possible candidate documents was more than 30,000, but the scope was reduced to 1000 or less.

Several experiments were done involving themes ranging from Java to Animaniacs (Fig. 3). In order to understand the analysis, each interaction is defined by one or more dialogs (exchanges) between a user and the system. Interactions for the experiment in Fig. 3 showed an increase in the number of documents matched as more than three turns are exchanged—this result does not come up by chance. For the same number of interactions (five), different results are shown mainly due to the adaptive way the dialog goes. That is, the context and kind of questions made by the agent are changing, depending on the situation and the document’s contents.

Different results were obtained for the same number of interactions because the kind of document searched for was changed as other features were restricted. A similar situation occurs as the dialog states a constraint regarding the language, in which case most of the original document references were not matched.

Experiments showed important drops in the results with a minimum of conversation turns due to constraints on the nature of the information finally delivered. The prototype search agent took into account previous issues, so there are some classes of high-level requests that are more likely to happen than others, depending on the context.

Overall, the current model, based on dialog interactions, shows promise as a novel and interesting work strategy to deal with specific information searching requirements. In addition, designing and implementing a natural language generation system easily can be adapted to tailored communicating situations.

For background information see ARTIFICIAL INTELLIGENCE; INFORMATION MANAGEMENT; INTERNET; LINGUISTICS; NATURAL LANGUAGE PROCESSING; WORLD WIDE WEB in the McGraw-Hill Encyclopedia of Science & Technology.

Anita Ferreira; John Atkinson