The Evolution of Search:

Moving from Information Retrieval to Dialogue-driven Advisory Systems

A White Paper by

AnswerChase™

White Paper

The Evolution of Search:

Moving from Information Retrieval to Dialogue-driven Advisory Systems

Overview

Online searching has permeated the Internet culture. After e-mail, searching typically ranks as the second most popular activity among Internet users. No wonder. The amount of information available on the Internet continues to grow at a staggering rate. The Web alone is estimated to be more than 1 billion pages in size.

Since the early days of electronic information processing in the 1960s, keyword-based searching has been the dominant approach to text retrieval. Most Web-based search tools, or search engines, continue to use the keyword approach. It's easy to use, relatively straightforward to implement, and works well for specific applications. But despite their prevalence and a variety of refinements, keyword search methods often fall short of the ideal: A given search session may return many links, but usually only a relative few (sometimes none) are relevant.

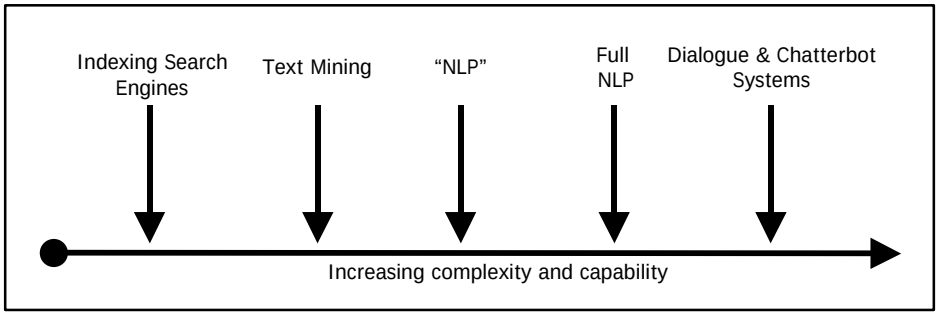

Figure 1. The overall complexity and capability scale for search tools.

Newer approaches to information retrieval seek not just to overcome the shortcomings of well-established methods, but to deliver a dramatic change in the way information is discovered. Text mining is the next step after keyword-based searching (see Figure 1), both in terms of complexity and capability (i.e., relevancy of results). Related to data mining, an approach to searching structured data bases, text mining employs a variety of analysis tools to better extract relevant text from unstructured documents.

The next leap in capability is Natural Language Processing (NLP), a technology for computerized language understanding. Unfortunately, many applications offer only *partial* NLP and are therefore only partially successful.

Full NLP, on the other hand, delivers both in terms of query understanding and information location capabilities. Full NLP and related chatterbot technologies are the most promising and capable information retrieval approaches. These

are intelligent, next-generation, and often interactive tools that understand ordinary language and are capable of inference, contextual analysis, and full-text search and response.

Poised to transform Internet-based information searching, full NLP and dialogue systems are also especially promising beyond the realm of personal computers. As electronic information content expands to Internet appliances, cell phones, and other devices, full NLP technologies offer the greatest performance, flexibility, and device independence.

This paper reviews prevalent methods for Web-based information searching and examines new technologies that expand searches beyond the Web. In particular, special emphasis is given to next generation technologies that combine full NLP with interaction capabilities.

Search Engines

The term "search engine" has come to refer to any tool for finding information online. Strictly speaking, search engines are automated Web page indexing tools. Such engines employ Web "crawlers" or "spiders" to continuously wander the Web in search of new sites and create and update searchable lists of candidate pages.

Another online Web-based information resource often called a search engine is actually a directory. Directories depend on humans to find, compile, and categorize searchable information. Directories allow for more accurate classification and better authentication of candidate pages than is possible with search engines. Because it requires human intervention, however, directory content often lags behind that of automated search methods.

A third approach, the "hybrid" search engine, seeks to blend the best of both worlds, by combining human- and crawler-compiled information to generate its page indices.

Nearly all popular search "engines" employ user-provided keywords to search their lists for matches. Keyword searching alone introduces semantic ambiguities that drastically reduce the reliability of a search, however. If searching for "bill," for example, a search is equally likely to return, among other things, information about paper money, hats, invoices, ducks, or President Clinton.

To address this, one of the early improvements made to keyword searching was the use of Boolean operators (AND, OR, etc.) logically organize search terms. Using with a variety of modifying symbols, Boolean operators allow searches to be expanded or restricted, and they usually increase the usefulness of search results.

Relevancy scoring, a more recent enhancement, sorts results based on internal search engine rules for statistical matching between the user query and the search engine's indexed document list. Some search engines go so far as to qualify the information source, making a distinction between sites considered to be authoritative and casual pages such as individuals' personal Web pages.

But even with these and other improvements, when a user enters keywords for the topic of interest, the end result is always the same: a list of links, some more relevant than others. The searcher still has to spend time exploring each link to see if it leads to the desired information. If none of the links apply, the search process has to be repeated.

Moreover, while search engines, directories, and hybrid engines are continuously updated, most manage to cover only a fraction of the Web. A recent industry review found that more than 90% of searches failed to find relevant information or to arrange the results in a meaningful way. To some degree, this low success number is the result of the deliberate elimination of duplicate or clearly irrelevant information, but the major cause behind missed pages is directly related to the inherent shortcomings of keyword-based indexed search methods.

Another significant contributor to this figure is that indexed engines, automatic or otherwise, cannot keep up with the Web's rapid and constant rate of change. As a result, their indices do not always reflect currently available content and may leave the user unable to track down information, particularly if it is very recent.

Text Mining

Text mining is a variant of data mining, a comparatively mature technology for data search within well-formed schemes such as relational data bases. Whereas the typical search engine compares keywords to a previously compiled index of Web pages, text mining has a wider application as a general text search tool.

Text mining seeks to apply some of the same types of analysis, such as knowledge discovery, or trend analysis, to unstructured textual data, that data mining applies to structured data. Its aim is to discover patterns in large textual sources in a way that delivers greater relevance to user queries. Text mining also incorporates association rules and takes synonyms and phrases into account, enabling it to link "PC" with "personal computer," for example.

Another advantage of text mining is its applicability to a variety of textual sources. Searches are not limited to the Web. The searchable domain is easily expanded to include word processing and other text-containing file types residing in a variety of locations.

Advances in computational resources and new statistical algorithms for text analysis have helped develop text mining as a field. Text mining has significant promise in dedicated use, and its techniques are important foundational components for more advanced information searching technologies such as natural language processing and dialogue systems.

Natural Language Processing

The main goal of natural language processing, or computational linguistics, is to provide a means for computerized natural language understanding. NLP can be used to analyze, classify, detect, and retrieve specific knowledge by means of computerized algorithms. This makes NLP an ideal tool for information search, as it allows users to interact more naturally, without first having to transform ideas into context-free keywords.

Keyword search engines return links based on word matching and/or the frequency of word occurrence within an indexed list of candidate documents. On the other hand, a well-programmed natural language search can infer meaning from both queries and a variety of information sources, thereby enabling the search to retrieve only truly relevant items.

The effectiveness of NLP is directly related to the degree to which it is employed. While some search methods either feature NLP or the ability to accept questions, their partial use of the technology usually leads to far less than optimal results.

Partial NLP

A number of Web-based search tools feature a natural language interface that allows users to begin a search by asking a question. This approach can work reasonably well for establishing the user's topic of interest but typically functions only on the front end.

Such partial, or "front end," NLP systems parse the user's question solely for the purpose of extracting keywords. The keywords are then used to search for information on the site's indexed search list or knowledge base. The remainder of the query is discarded; little or no context information is retained, and no query refinement is done.

Partial NLP is, by definition, only partially effective. It depends on the quality of pre-indexing and assumes its pre-defined link list covers all possible user questions. And because it works only at the front end, partial NLP is not applied to searches. The net result is output that is not substantially better than that produced by traditional keyword-driven search engines.

Full NLP

Full NLP provides superior performance because it applies the technology to both the user query (front end) and information location (back end) segments of a search process and because it can analyze individual document *content* not just a document *index*.

Advanced techniques are required to successfully parse text at both ends of the process, including syntactic match analysis, ontology, and semantics. A full NLP implementation also employs fuzzy logic to interpret both the question and the relevancy of candidate answers. A set of mathematical principles, fuzzy logic (also known as fuzzy set theory) provides a structured means for dealing with ambiguity or relative degrees of "truth." Natural language is full of "fuzziness," thus fuzzy logic is ideal for computer-based analysis of natural language.

To round out its capabilities, full NLP incorporates artificial intelligence (AI) to make effective use of the search paradigms natural to every human being. The main objective of AI is to embed human-like intelligence into computer programs to make them think, learn, communicate, advise, and decide. Both fuzzy logic and AI systems are used extensively in engineering applications, in everything from missile guidance to aircraft automation and landing control.

These components combine effectively to set the stage for searches based on similarity of content rather than strict vocabulary. Full NLP analysis at the front end first extracts both keywords and contextual information from the query.

Back end NLP then applies the same text analysis process to the search. Its semantic analysis connects the question with categories and documents that are related to the concepts the user is interested in—directing the search to information sources that would have been missed in a traditional keyword-based search.

By contrast, keyword- or frequency of occurrence-based searches do not read candidate documents on the fly. Instead, these search methods are limited to scanning pre-defined document categories and topic indices for information that may match the question. The method assumes the indices cover a sufficiently large number of questions, but there is no guarantee that the links provided have been indexed for the correct content, even with exact keyword matches.

Another advantage of full NLP is the breadth of documents that can be searched. Whereas search engines and directories are limited to scanning indices of Web pages, full NLP scans a variety of files for embedded information. This includes hypertext, word processing, rich text format, and other common document types.

A full NLP implementation can thus be a powerful information search and discovery tool for a variety of environments. Hosted on a Web site, it can serve to answer frequently asked questions or provide visitors with product information and selection guidance. Appropriate segments of a corporate domain could be set up to be searchable with full NLP, giving employees a new way to tap into company-wide knowledge.

In a controlled domain, such as a corporate network, automated concept indexing delivers extremely fast search results. This is especially useful, for example, on a domain that has a substantial number of documents. Periodic full NLP-based concept indexing tames even very large data resources to deliver exceptionally quick and accurate results.

Dialogue Systems

Dialogue systems, sometimes called chatterbots or intelligent agents, enable human-like dialogue with computers. With a full NLP foundation, dialogue systems function by asking a series of questions in response to a user inquiry, thereby refining the topic until a very accurate answer can be provided. This is especially useful if the initial concept is vague.

Consider the following example in which the goal is to book a seat on a commercial airline flight:

User: I need a flight to Portland.
Bot: Portland, Maine or Portland, Oregon?
User: Oregon
Bot: When would you like to go?
User: In October.
Bot: During which week in October?
User: The first week.
Bot: Which airline(s) schedules would you prefer I search?

… and so on.

Figure 2. A hypothetical conversation with an airline reservation desk's chatterbot.

This conversation is very similar to one that might take place with a human travel agent. Here, the user interacts with an automated chatterbot system capable of understanding the questions and, very importantly, determining when additional information refinement steps are required. In this case, the interaction can continue until the chatterbot application gathers sufficient information to deliver a list of flights that match the user's schedule, price, and route preferences.

Chatterbots are never expected to replace humans beings, of course, but they can be very valuable in applications where the general objective of the dialogue is well defined. For example, many calls to reservation or help desks can easily be handled by automated dialogue systems. This is especially true for help desk representatives, who spend much of their time answering the same sets of questions.

Chatterbots deliver good information-finding results even when the objective dialogue is restricted or poorly defined. Their ability to process user input and generate associated questions allows them to narrow the topic sufficiently, even in a general purpose application.

Their conversational nature also means users do not have to spend time figuring out how chatterbots work. The user asks an initial question, and the chatterbot takes it from there, refining the topic step-by-step. This is a significant advantage over wrestling about how to best combine keywords and Boolean operators, endlessly wandering Web pages looking for needed

information, or trudging through a list of links only to find that few, if any, are relevant.

Dialogue systems have exceptional versatility. They can be implemented in a variety of environments, including Web servers, corporate networks, and in computer help systems, to name just a few. In addition to their use in text-driven software applications, the chatterbot's power and flexibility also makes them ideal for use in speech recognition and voice response systems. Embedding this functionality in a cell phone, for instance, extends the users' reach and delivers information with greater ease and flexibility.

Advisory Systems

One application for the combination of dialogue systems and full NLP is advisory systems. These systems are capable of proactive data and information extraction and can play a critical role in the operation of any organization.

Advisory systems serve a very important role in coporate environments, where they can generate advice based on embedded fuzzy rules that capture the expertise and experience of the enterprise.

For example, if a financial services firm offers on-line advice, an advisory system can perform rules-based capture of selected in-house financial forecasting expertise for automatic dissemination to clients. Using the knowledge gathered, an advisory system is capable of synthesizing an appropriate answer and is especially valuable for cases in which user queries cannot be satisfied by standard responses.

Conclusion

[TBD: Conclusion to be completed after first round of comments. The goal of the conclusion is to tie together all topics and conclude that full NLP and dialogue systems are the latest, greatest, and coolest tools for information search. These technologies: provide greater results relevance; eliminate ambiguity (and thereby go directly to the topic of interest); build upon previous queries and associated results; are flexible in terms of domains searched and operating environment; and in general offer intelligence (i.e., processing capability) not available in standard text search tools. The technologies are therefore poised to transform the information search model to the next-generation level, expand the searchable information base, and extend search to a wide variety of applications. The conclusion should also allude to the next white paper, on NLP and dialogue systems.]

About AnswerChase

AnswerChase is driving the evolution of search by creating highly intelligent, dialogue-driven advice solutions that deliver a new level of interaction between people and computers. AnswerChase develops intelligent search and advisory systems for accessing relevant information across intranets, the Internet and local, wide-area, and wireless networks.

AnswerChase's enabling infrastructure technologies provide a platform to help organizations harness the power of information content by automating search processes, providing advisory systems and creating human-like dialogue with computers. AnswerChase's search and advisory technologies facilitate electronic commerce, customer service and technical support. Other business applications include intelligence monitoring and knowledge management applications.

Created by a team of mathematicians, linguists, and engineers, AnswerChase solutions are the most comprehensive, advanced line of search and advisory systems technologies on the market. Its core technologies include intelligent navigation, sequence indexing (Syndexing), fuzzy inference, artificial intelligence, patent-pending complete natural language processing (NLP), and computational linguistics.

For more information: AnswerChase is located at AnswerChase, Inc., 34 Defense Street, Annapolis, MD 21401, and can be reached at +1 (410) 266-3440, via e-mail at sales@answerchase.com, or on the Web at www.answerchase.com.

A trial version of AnswerChase's full NLP searchbot (AnswerChase Desktop to Internet) is available on the Free Downloads section of the Web site.

Rev A, [date [TBD]]