

DOI:10.1145/1378727.1378743

**Beyond Google, emerging question-answering systems respond to natural-language queries.**

**BY DMITRI ROUSSINOV, WEIGUO FAN, AND JOSÉ ROBLES-FLORES**

# Beyond Keywords: Automated Question Answering on the Web

SINCE THE TYPICAL COMPUTER USER spends half an hour a day searching the Web through Google and other search portals, it is not surprising that Google and other sellers of online advertising have surpassed the revenue of their non-online competitors, including radio and TV networks. The success of Google stock, as well as the stock of other search-portal companies, has prompted investors and IT practitioners alike to want to know what's next in the search world.

The July 2005 acquisition of AskJeeves (now known as Ask.com) by InterActiveCorp for a surprisingly high price of \$2.3 billion may point to some possible

answers. Ask.com not only wanted a share of the online-search market, it also wanted the market's most prized possession: completely automated open-domain question answering (QA) on the Web, the holy grail of information access. The QA goal is to locate, extract, and provide specific answers to user questions expressed in natural language. A QA system takes input (such as "How many Kurds live in Turkey?") and provides output (such as "About 15 million Kurds live in Turkey," or simply "15 million").

Search engines have significantly improved their ability to find the most popular and lexically related pages to a given query by performing link analysis and counting the number of query words. However, search engines are not designed to deal with natural-language questions, treating most of them as "bags," or unordered sets, of words. When a user types a question (such as "Who is the largest producer of software?"), Google treats it as if the user typed "software producer largest," leading to unexpected and often not-useful results. It displays pages about the largest producers of dairy products, trucks, and "catholic software," but not the answer the user might expect or need (such as "Microsoft"). Even if the correct answer is among the search results, it still takes time to sift through all the returned results and locate the most promising answer among them.

It is more natural for people to type a question (such as "Who wrote King Lear?") than to formulate queries using Boolean logic (such as "wrote OR written OR author AND King Lear"). Precise, timely, and factual answers are especially important when dealing with a limited communication channel. A growing number of Internet users have mobile devices with small screens (such as Internet-enabled cell phones). Military, first-responder, and security systems frequently put their users under such time constraints that each additional second spent browsing search results could put human lives at risk. Finally, visually impaired computer us-