

Industry Watch: Is 2007 the Year of Question-Answering?

Robert Dale

*Centre for Language Technology
Macquarie University
Sydney, Australia*

(*Received 17 December 2006*)

Back at the beginning of 2000, I was a member of a working group tasked to come up with some guidelines for revamping my University's website. During one of our meetings, someone made the suggestion that decisions about how to structure and present information on the website should be driven by the kinds of questions that users come to the site with. Suddenly a light went on, and there appeared an idea for data gathering that might provide us with some useful information. To find out what people really wanted to know when they visited the website, we would replace the University's search engine by a page that invited the user to type in his or her query as a full natural language question. Appropriately chosen examples would be given to demonstrate that using real questions delivered better pages as a result. The data gathered would tell us what people were really looking for, more than could be gleaned from conventional search queries, and would therefore help us to better structure the information available on the website.

Thanks to a very cooperative University webmaster and a supportive administration, our JustAsk! web page was up and running within weeks. Of course, we didn't actually implement a question-answering system. Instead, behind the page was a script which, given any submitted query, would strip out any stop words it contained (after logging the full query, of course), and then simply pass the remaining words in the query on to the same search engine that had been used previously. Within a few days of the deception being put in place, a number of people independently commented on how much better the University's search engine was, now that we had added a natural language capability. I came clean and denied all responsibility for any perceived improvements, of course; but it's entirely possible that users did get better results simply because, even after stripping out stop words, their queries tended to become longer, and contained more content words, when expressed as questions.

Every Sunday at midnight since then, I have been sent a log file containing the previous week's queries. The web page in question has now moved and become

⁰ Industry Watch is a semi-regular column that looks at commercial applications of natural language technology. The author can be contacted at rdale@acm.org.

inaccessible, at least to humans: these days, the logs I receive contain nothing but the URLs of porn sites, online casinos, and Viagra ads. But over the period from 2000 to 2005, before the spambots kicked into action, we accumulated around two million queries. That's not a lot for a search engine, but still enough, one would think, to be instructive about user interests in a limited domain.

Now, there are all sorts of methodological flaws with this experiment, but the single most important result of analysing the data is hard to deny: despite our best attempts at exhorting the user to provide full natural language questions, only around 8% of the queries submitted via the page were in fact questions. The habits of googling 2.4 words at a time are well-entrenched and hard to break. It has always seemed to me that this is a significant problem that any attempts to commercialise natural language question answering as the next step beyond conventional web search must somehow address.

Which is not to say that many have not tried. Most people will be familiar with Ask Jeeves, for example (now simply www.ask.com), which burst onto the scene in 1998. In an excellent example of web democracy at work, popular questions were rewarded with the attention of human editors to ensure that they returned quality answers; if you had a minority interest, you'd be just as well to use a conventional search engine. But as the underlying technical issues to be faced in real automatic question answering became better understood, Ask Jeeves was followed by a slew of question-answering systems that really did attempt to take the human out of the loop. Many of these no longer exist, or have morphed into other products. The Electric Monk (1998) claimed to find answers to real English questions by reformulating questions to produce Alta Vista queries. Albert (2000) was branded as a multilingual, natural-language-capable, intelligent search engine, subsequently purchased by FAST, passed on to Overture and then to Yahoo!. iPhrase (2000) appeared for a time as a natural language search and navigation solution at Yahoo! Finance, and was bought by IBM in 2005. All of these have more or less disappeared from the face of the web, but some more recent offerings that attempt to utilise some kind of natural language processing are still around. BrainBoost (2003; www.brainboost.com), like the Electric Monk, finds answers to natural language questions by translating them into multiple search engine queries and then extracting relevant answers from the returned pages. BrainBoost still exists, but is now owned by Answers.com (www.answers.com). MeaningMaster, created by Kathleen Dahlgren, appeared in 2004, claiming that it 'delivers results three times more accurate than Google', using a lexicon that provides contextual meanings for 350,000 words; MeaningMaster has since been rebranded as CognitionSearch (www.cognition.com).

There are no doubt many other attempts at natural language search. But the point is this: none of these have brought QA onto the mainstream internet user's horizon. As noted in this column a while back, even Google's own limited question answering capability remains effectively unadvertised (and in fact seems to have lost functionality it once had).

But maybe this is all set to change. If you hang out at Buck's Restaurant in Wood-

side in the Valley,¹ you will already have heard of Powerset (www.powerset.com). But if eavesdropping on venture capitalists over pancakes is not your thing, or you're not otherwise part of a local in-the-know circle, it's more than likely that Powerset has been below your radar. As they indicate on their web site, they have been in semi-stealth mode for over a year, only recently being outed to the general public by a story in the UK's *Sunday Times*² and the announcement of a US\$12.5 funding deal.³ Powerset, it appears, is the company that will knock Google off its pedestal, by bringing real question answering to the great unwashed web. Barney Pell, the company's founder, talks up the possibilities on his blog:⁴ 'Search is in its early days, and natural language is the future of search.'

Great stuff. All of Pell's arguments for QA as an improvement on search will be familiar to readers of this journal. A widely-used and effective natural-language based question answering interface to the web would do wonders for the visibility of natural language processing as a research field. But, given the lacklustre appeal of earlier attempts, why should we expect Powerset to do any better? Danny Sullivan's rant against question answering⁵ presents the sceptics' position. Sullivan shares the concern I indicated above about getting people to stop typing 2.4-word queries: 'People search however they want—and right now, they use only a few words . . . Getting inside the minds and whispering 'type longer' isn't going to be fun.' Google's response to challenges from companies like Powerset was summed up by Peter Norvig, Director of Research, thus: 'They have maybe one small lever that they suspect is huge. They don't realise that [all] they have [is] a better door latch on a [Boeing] 747. Now all they have to do is build a 747.' Nonetheless, Powerset is backed by some big names—people who you'd assume have a good sense of judgement—so maybe this time it will be different.

Powerset has been getting the media attention, but over on the East Coast of the US there is QA activity too. Hakia claims to be building the Web's new 'meaning-based' search engine; check out the beta version at <http://www.hakia.com>. Hakia's Riza Berkan may not have Esther Dyson and an army of angel investors from the search world on board, but he is being advised by Yorick Wilks. Also on the East Coast, Teragram is quietly plugging away at improving search by adding linguistic sensitivity. The company's Direct Answers product, with an enterprise version announced in June, was chosen by *KMWorld Magazine* as the Trend-Setting Product of 2006.⁶ In use at AOL, this technology delivers short, specific answers by 'grasping the essential question'. Teragram has also been named to the 2006 EContent

¹ www.valleywag.com/tech/buck's/wanted-bucks-restaurant-vc-spotter-175471.php.

² See www.timesonline.co.uk/article/0,,2095-2459650.html.

³ Remember, this is being written some three months before you would have had any chance to read it. All sorts of things could have happened in the interim. Powerset may already be your home page.

⁴ See www.barneypell.com.

⁵ *Hello Natural Language Search, My Old Over-Hyped Search Friend*, <http://blog.searchenginewatch.com/blog/061005-095006>.

⁶ See <http://public-issues.com/2006/12/teragrams-direct-answers-chosen-by.html>.

100, the premier list of ‘100 Companies That Matter Most in the Digital Content Industry’ as judged by *EContent Magazine*.⁷

Microsoft looks to be gearing up for QA too. The company has bought Colloquis (www.colloquis.com), whose offering claims to use natural language processing to give companies a way to do automated customer-service online without the need for a human customer-service agent. You can now buy the Colloquis technology as a hosted service, rebranded Windows Live Service Agents.

Which almost brings us back to where we started with Ask Jeeves: perhaps inspired by a deviant reparsing of ‘Windows Live Service Agents’, ChaCha (www.chacha.com) is a ‘human-powered search engine’ that uses an army of humans to provide answers to questions: ‘ChaCha only provides quality, human approved results’. I typed into ChaCha: ‘Is 2007 the year for question answering?’, and BrettaF, my guide, went off looking for an answer, only to pass me a minute or two later to jamieT, ‘another guide who can help you search even better’; jamieT in turn passed me on to KarenB, ‘another guide who can help you search even better’ But I’m in an airport lounge with my plane about to depart, so I reverted to feeling lucky with Google. I got a job ad for a position in Ruslan Mitkov’s group, but I was no closer to an answer. Maybe by the time you read this, things will be different.

If 2.4-word queries are the major roadblock to question answering on the web, then the analogous challenge for voice recognition on mobiles must be the widespread use of SMS. It’s pretty hard to beat those thumbs, especially when they are on the hands of a teenager. But that may be set to change too: Nuance’s mobile speech platform now provides a ‘dictate-anywhere’ function on the Windows Mobile and Symbian operating systems. To showcase the technology, Nuance challenged Ben Cook, the world champion texter, to compete with its software to determine what would be the fastest and most accurate way to send text messages. Cook holds the record for the fastest entry of a 160-character message on a mobile device, at 42.22 seconds. In the bake-off with Nuance, he finished in 48 seconds; using the Nuance software, it took 16.32 seconds to compose the message. Watch it on YouTube at <http://www.youtube.com/watch?v=-L4Jk6GDud0>.

So it looks like the peaceful, SMS-induced, silence on public transport over the last two or three years was only temporary. I see a bleak future where everyone sends text messages by talking, just like they used to a few years back, into their phones in public places. Just one more excuse to buy those neat Bose noise-cancelling headphones for your iPod.

⁷ See <http://www.econtentmag.com>.