l e x a n a \ n e t

September 2008

Market Disappointments: Natural Language Interfaces to Relational DBMSs

By Drew Hamre

email

Roughly 30 years ago, Yale computer scientist Drew McDermott wrote an essay - "Artificial Intelligence Meets Natural Stupidity" - that gleefully punctured the claims of researchers in artificial intelligence (AI), including those of McDermott himself. McDermott believed that the central research problem in AI[i] was natural language ("the most fascinating and important research goal"). However even back in 1976, advancements in natural language processing (NLP) seemed frustratingly slow.

"Everyone takes natural language interfaces for granted, though none has been written", McDermott observed. "It is hard to say where they have gone [more wrong], in underestimating language or overestimating computer programs".

McDermott foresaw disappointment especially for those creating information retrieval systems that responded to *natural language queries* (questions posed in everyday conversational language, rather than in formal retrieval statements). "We all trick ourselves into thinking that the statement of a problem in natural language is natural (but) real discourse is almost never of the literal-minded information-retrieval variety."

Three decades later, NLP researchers can point to many successes: usable (if not always accurate) language translations are freely available on the web, grammar and spelling checkers are integrated into workaday office software, and utilities for disambiguating word senses and determining parts of speech are embedded in search engines, document managers, and text miners.

However, natural language query tools have not been accepted, much as McDermott predicted. Despite some expensively-promoted product launches, it remains rare today for anyone to use natural language tools to a) query a database or b) conduct question-answer dialogs with unstructured document stores, including the web. The brass ring for NLP remains out of reach. The "natural stupidity" of human language (its ambiguities and rule-bending irregularities) remains a challenge.

In the following commentary, we'll review recent attempts to commercialize natural language interfaces (NLIs) for two different technologies:

1. **Natural language interfaces to relational databases** including tools such as Microsoft's "English Query", ELF, and Progress Software's EasyAsk.

Software of this type allows users to retrieve information from structured databases using everyday language. For example, a question like ***"How many audio CDs by Sonata Arctica were sold last month in Boston?"*** would be sufficient to pull the answer from a sales database.

2. **Question-answering search interfaces to document stores including the web itself** from vendors such as AskJeeves (now Ask.com) and Powerset.

   Software of this type allows users to pose questions to document collections (including HTML pages) and receive an answer in response. A question like ***"How many times did Bjorn Borg win Wimbledon?"*** would return the answer (5) rather than a list of key-word matched web pages or documents.

Background: Why NLIs for databases differ from NLIs for the Web

Computer information can generally be characterized as either structured or unstructured. Structured information is 'schema-first' and is commonly held in the tables and columns of an RDBMS; unstructured information is 'schema never' and commonly resides in byte streams on a file system.

Modern relational databases use formal query languages (like SQL and XQuery) to access optimized, proprietary storage structures; easy-to-use administrative, query, and reporting tools have evolved for these systems. Unstructured document stores are commonly accessed via metadata indices and search engines; sophisticated crawling, indexing, link-ranking, and text analytic tools have evolved for these systems.

These structural differences explain why – although it might seem that natural language interfaces for databases and for the web should be similar – the NLI implementations are quite different (summarized recently by Jagadish and others):

- We expect database queries to return precise, complete, and reliable (repeatable) results. Conversely, we tolerate (and sometimes enjoy) the imprecise results returned by unstructured search engines.
- We expect database queries to support more sophisticated questions than unstructured data stores (for example, reservation systems allow hotel selection filtered by data, availability, location, and a range of amenities).
- We expect that database query results will themselves have structure (for example, cross-tabbed summaries and statistics). We carry no such expectations for unstructured sources.

NLI vendors who enter the market face formidable competitive challenges because the current interface technologies are so strong:

- For structured databases, visual or forms-based query tools are extremely easy to use (undercutting the primary advantage of NLIs). In addition, current tools support more complicated queries than NLIs and generate reliable results.
- For unstructured data stores, the speed, transparency, predictability, usability, and comprehensive coverage of current search technologies (such as Google's) provide rich and reliable result sets and unrivalled ease of use.

NLI vendors today face fundamental competitive questions: Does an NLI truly offer additional value (enhanced ease-of-use), or are users' needs being met by existing technology? Have we all been trained to "think in Google"?

As researchers have summarized (emphasis added): "In the time of Google and graphical user interfaces, where people are used to formulating their information needs with keywords … or (are accustomed to) clicking through menus and graphically displayed functions, ***full-fledged NLIs may be redundant***."

NLIs for Databases: Microsoft English Query, ELF, and EasyAsk

Commercial NLIs for databases (NLIDBs) build upon research implementations from the 1970s (e.g., LUNAR and LIFER/LADDER). Ever since relational databases became dominant in the 1980s, an NLIDB will typically translate users' queries from informal, everyday English into SQL:

- The user enters their question ("Which store didn't re-order Pepsi in May?")
- The NLIDB may interact with the user to correct errors in spelling or grammar
- The NLIDB may then display its interpretation of the question (paraphrase) and perhaps its SQL translation before passing the translated query to the DBMS.
- The NLIDB receives the result set from the DBMS and displays it for the user.

Rationale and market history

NLIDB development is intended to support people who interact with databases but who a) don't know a formal query language and b) don't know the target databases' schema (potentially all of us, in other words). Another incentive for development is the proliferation of mobile devices (e.g., spoken cell phone queries transcribed by speech recognition software, and then answered by NLIDB/text-to-speech systems).

However, most NLIDB products have fallen by the wayside. For example, Symantec once offered a natural language tool called *Q&A*, NRI offered a tool called *Natural Language*, Unisys tried to commercialize their *NLU-Natural Language Understanding* project, Software AG offered a product called *Esperant*, BBN offered a tool called *Parlance*, and IBM offered a tool called *LanguageAccess*. These and other NLIDB products have failed commercially for a variety of reasons (summarized by Androutsopoulos and others and by Laukaitis and others):

- Natural language is often inappropriate (ambiguous) for expressing directives that must ultimately be formalized. Today's NLP software can't yet conduct the complicated 'conversations' with the user needed to resolve ambiguities (analogous to the process of clarifying reporting requirements).
- Other interfaces (graphical, menu-driven, query-by-example) are now extremely powerful, yet remain easy-to-use and are being adapted to a variety of portable devices.
- An NLIDB's linguistic 'coverage' (the range of knowledge it represents) isn't obvious to the user, causing frustration and confusion.
- NLIDB operation isn't transparent, and users may be uncertain whether a miscommunication occurred and whether its cause was linguistic or conceptual

- If an initial dialog is successful, users may assume intelligence on the part of the NLIDB leading to later frustration
- NLIDB tools require expensive tailoring of database interfaces, and these are the responsibility of expensive specialists

Despite these hurdles, it remains possible today to purchase and deploy commercial NLIDBs. Following is a summary of three 'surviving' products (though this may exaggerate their vitality somewhat): ***Microsoft English Query***, ***ELF***, and ***EasyAsk***.

Microsoft English Query

Microsoft English Query (EQ) was based on technology acquired from Natural Language Inc., a firm in which Microsoft had first invested in 1987. (To place this in historical context, the news story reporting the NLI investment also noted Microsoft had acquired the firm that made a little-known presentation package: 'PowerPoint'.)

Microsoft first made EQ available with SQL Server 6.5 Enterprise in 1997, and then bundled EQ with two subsequent SQL Server releases (7.0 and 2000). Microsoft promoted the tool heavily, and Bill Gates highlighted EQ in his 1998 COMDEX keynote.

English Query found some initial success (and even plays a supporting role in US Patent 7,376,645). However, the marketplace didn't respond. EQ was last updated in 2001 via an optional hotfix in SQL2000/SP1. The tool was quietly dropped from SQL Server 2005, presumably due to lack of customer interest.

While EQ *can* be used with current Microsoft database products, its supportability is clearly at risk:

- **EQ Model Editor** – The final version of EQ's model editor is integrated into Visual Studio 6.0, the 10-year-old development environment that pre-dates .NET. Extended support for this environment expired in April, 2008.
- **EQ Runtime Engine** – The EQ runtime is implemented as a handful of redistributable DLLs. These files didn't ship with SQL Server 2005, but they can be a) installed from a properly licensed copy of SQL2000 and b) used against a SQL2005 instance. It's unclear whether this work-around is viable with SQL Server 2008.

Natural Language Inc.'s founder, John Manferdelli, remains with Microsoft, though he researches software security rather than natural language. The product manager of EQ, Adam Blum, left Microsoft in 2000 and hasn't worked with NLP since.

Although EQ has essentially disappeared without a trace, Microsoft remains active in NLP research and retains hopes for this technology. As CEO Steve Ballmer recently told Gartner analysts: *"People laugh and say that's futuristic, but natural language is a big thing".* In one of Ballmer's first strategic moves after Bill Gates' retirement in 2008, he this belief with a $100-million bet on NLI vendor Powerset (see below).

ELF

ELF Software (English Language Frontend) is a small company founded by computational linguist Jon Greenblatt. Greenblatt dates ELF's development to the mid-1980's, and cites the influence of Noam Chomsky's transformational grammars in its design. Although ELF software remains available, ELF's products haven't been updated in several years.

ELF is marketed in two forms: Access ELF (an interface for Microsoft Access), and VB ELF (a set of ActiveX controls that connect to any ODBC/OLEDB compliant database, including SQL Server and Oracle). Access ELF was apparently last updated coincident with Office XP (circa 2002); as for VB ELF, it pre-dates .NET and thus faces the same portability challenges as English Query.

Even though ELF appears all but dormant, it remains prominent among surviving NLIDBs and is featured in academic reviews of the technology. ELF positions itself against the competitors discussed here (EQ and EasyAsk), and claims multiple instances where the ELF product answers more questions, more accurately than the others. (This publication is interesting because it includes all test questions and answers, illustrating the strengths and weaknesses of today's NLIDB products.)

ELF includes (as does EQ) an automated customization feature. Such tools automate the initial configuration of the NLIDB system, analyzing the schema of a target database and mapping between linguistic concepts and the tables/columns of the implementation. For both EQ and ELF, the best automated results are obtained for target databases that a) have a normalized relational design, b) have tables and columns that are descriptively labeled, and c) are well-constrained by foreign key relationships. Of course, robust applications require additional NLIDB customization (and this effort is often substantial).

EasyAsk

Unlike the above two moribund products (EQ and ELF), EasyAsk is actively marketed by Progress Software. Its lineage stretches back 20-years to the *Intellect* natural language tool from Artificial Intelligence Corporation – a company founded by former Dartmouth professor Larry Harris. Harris later developed the *English Wizard* NLIDB tool, forerunner of *EasyAsk*.

Revenue figures aren't available, though EasyAsk is self-described as a "small business". The bulk of its installations support enterprise query, though it's seeing new activity in e-commerce applications. EasyAsk claims some large corporate clients, including *The Gap* and *Sony*.

Within the last year or so, EasyAsk's focus has expanded from transactional databases and online product dictionaries to include operational BI. This strategy shift echoes Microsoft, which added BI support (MDX emission) in the final release of its NLIDB product.

Unlike EQ and ELF, EasyAsk offers a blended strategy that goes beyond natural-language-to-SQL translations. Harris notes (emphasis added), "First, we generate an ad hoc, SQL-based query, *and second, we do a search on a repository of standard reports*." Harris argues that reports may be a better information source than ad hoc queries. "A structured report might include additional data that wasn't explicitly mentioned in the question but that might be useful in the analysis. Reports also offer formatting that might better highlight key results."

EasyAsk's novel query interface supports *Guided Navigation* (likened to a 'bread crumb trail') that shows users the attributes and categories associated with the result set, and allows them to refine the search though a point-and-click UI. (See EasyAsk's white paper, "Simplified Data Access Through Everyday Language").

EasyAsk is the most successful of the three NLIDBs discussed, and its 'blended' strategy may show the way for future products. Its combination of NLI and GUI (the

Guided Navigation tool), and its combination of database queries and textual search both help address the shortcomings of pure NLIDB products.

EasyAsk's document search facility (their product literature implies coverage of both textual RDBMS fields and file-system-based documents and web pages) indicates evolution toward the second class of NLI discussed here – natural language interfaces for unstructured document stores, including the world wide web.

Web NLIs: Ask.com (formerly AskJeeves), Powerset, and Google

The business value of unstructured information has been increasingly recognized over the past decade, just as its volume has exploded. Tools such as Microsoft Word and Apache web servers today manage far more terabytes of critical information than do RDBMSs like SQL Server or Oracle, and this gap is widening. The software market for managing unstructured data (formerly dominated by specialized information retrieval (IR) packages such as WestLaw and LexisNexis) has expanded to include search engines, crawlers, document portals, and text miners. The market has grown to spawn some of the most valuable corporations on earth (Happy 10th Birthday, Google).

Because of this potential, huge investments are being made to improve search tools and other software for unstructured data. NLP plays a prominent role in this quest.

Rationale and market history

Natural language technologies are pervasive in the software that manages documents, but usually the technology is hidden. Question/Answer (Q/A) interfaces are a visible exception. Whereas classic search/IR software provides a list of documents in response to keyword queries, Q/A systems provide answers in response to natural language questions.

Q/A NLIs became prominent around 10-years ago. DARPA/NIST's annual Text REtrieval Conference (TREC) has run throughout this span, and the software's steadily improving performance during this time is impressive (e.g., compare 2002 to 2007).

Q/A vendors see a huge potential market. Studies show one-third of keyword-based searches aren't answered by the first query/click. Some evidence shows that people *expect* search portals to answer questions: multi-token queries are the norm (average query length reached 4.8 words in one study, and in 2006 averaged 3.3 words overall), and as many as 37% of English queries began with a 'wh'-token.

However (as we saw with NLIDBs), these findings don't automatically translate to consumer acceptance. Industry observer Danny Sullivan cites commercial disappointments for web NLIs ranging from Excite (which in 1995 ballyhooed 'intelligent concept extraction'), Electric Monk (1998, natural language query), FAST/Albert (2000, natural language interpretation), iPhrase (2001), BrainBoost (2003), MeaningMaster (2004), and Stochasto, Kozoru, and Accoona (all from 2004).

John Lowe (formerly Chief Linguist of AskJeeves) drew from a base of several hundred million user queries in summarizing the challenges faced by such software. He notes the varied nature of queries, ranging from keyword lists, keywords with Boolean operators, phrases with linguistic coherence, interrogative or imperative sentences, short discourses concluded with a question, or some combination of the above. Peter Norvig[ii], Google's Director of Research, recently said that the primary

challenge faced by deeper search strategies will be the 'incompetence' of the general user. (Just how many alternative spellings of 'Britney Spears' should a search engine accommodate?)

Lowe notes that 'stop words' (usually eliminated in search string parses) play a central role in deeper search, for example in distinguishing 'books by kids', 'books for kids', and 'books about kids'. Lowe also discusses the sophistication required to answer particular questions (e.g., tech support dialogs or this striking example: "What did Tom Hanks say to Private Ryan as he was dying?" - a question that confuses the universes of actors and fictional characters, along with other difficulties).

We humans interpret these complications intuitively, but any software that tries to 'understand' such questions faces enormous challenges. Ironically as Lowe notes, a standard keyword web search for the 'Private Ryan' question would likely bring us within striking distance of the answer (and in fact, Google returns a page with the answer - "Earn it" – near its top-ranked link).

Ask.com (formerly AskJeeves)

During its successful IPO in 1999, the AskJeeves search engine described itself as a "provider of natural-language question answering services for consumers". In March 2005, media conglomerate IAC acquired AskJeeves in a stock swap ultimately valued at $2.3 billion (US). The chairman of IAC believed AskJeeves had "the potential to become one of the great brands on the Internet" and compete directly against Google and Yahoo. Toward this end, AskJeeves was re-branded as Ask.com.

In March 2008 in a "dramatic about face", IAC announced they were laying-off 8% of the Ask.com workforce and abandoning the competition against Google and other general-purpose search engines (articles aggregated here). "No matter what (Ask) did, it just wasn't enough to get people to leave Google," said one analyst. "This looks like they are raising the white flag."

Ask.com remains the US's fourth most popular search engine. Comscore ratings for July 2008 give Ask sites a 4.5% share … well behind Google [61.9%] and Yahoo! [20.5%], and trailing Microsoft/MSN/Live [8.9%].

Ask.com's underlying technology remains strong (see this overview) including their ExpertRank algorithm (based on acquisition Teoma's technology, which ranks links only within collections of search-relevant 'expert' documents), scalability supported by 'Neptune' load-balanced clustering, and a blended search UI that (for certain queries) combines structured data with web search.

Powerset (now part of Microsoft)

Powerset's technology originated at PARC (Xerox's famed Palo Alto Research Center). In 2007, PARC and Powerset agreed to commercialize this patented technology and a public demo was released to the web earlier this year. Microsoft purchased PowerSet for a reported $100 million (US) in early July 2008, shortly after Microsoft's takeover of Yahoo! was aborted.

One analyst said, "As Microsoft (attempts) to make inroads against Google, it hopes that the semantic or natural language approach will yield fruit". However, it's unclear how Microsoft will integrate Powerset's technology. Microsoft's says Powerset was acquired "first and foremost because we're impressed with the people there". Powerset's linguists and engineers have joined Microsoft's search relevance team.

Powerset claims to create and use *semantic web* information ('semantic web' is a loaded term nowadays and this claim is discussed below), combining 'deep' NLP with scalable search technology. Powerset a) interprets source content, b) indexes the digest, c) interprets the query, and d) produces matched results.

In a speech to the International Semantic Web Conference (2007), Powerset CEO Barney Pell noted the industry's progress in processing documents a) first, as "bags-of-keywords", then as b) "vectors-of-keywords" (where proximity is coded), and next for c) adding contextual richness via "off-page text" via anchor tags. The next step, Pell said, will be to exploit the text's linguistic structure. Powerset does this with a semantic indexer that cracks each document's linguistic structure to extract meaning.

Powerset explicitly markets its technology in the context of the semantic web (in the more generic sense of linked documents embellished with computer-readable semantics). Pell believes the roadblock to semantic web acceptance (the cost of marking-up content) can be overcome by NLP systems that automatically create annotations from unstructured text and generate ontologies. Powerset claims it does this, and then leverages these semantic markups to improve search performance.

Powerset's public demo offers a Q/A frontend to encyclopedias Wikipedia and Freebase, *not* to the entire web. This may portend scalability challenges for the technology. The co-founder of Wikipedia, Jimmy Wales, remains unimpressed: "It doesn't present much of a challenge. Wikipedia isn't a very large data set, and it's a pretty simple thing to do, to index Wikipedia. So whether (Powerset's) approach is going to be useful on a bigger data set [is hard to tell]."

Powerset - in fact, all search engines - ultimately may need user guidance in understanding the intent of a query. This means richer, longer queries are required. Industry trends are hopeful: Yahoo! research shows the average search query length was 1.2 words in 1998, 2.5 words in 2004, and reached 3.3 words in 2006.

In this more complex linguistic environment, Powerset can point to some impressive results, such as returning the same answer to linguistic variants of a question (for example: What disease did Sir Edward Heath die from? What killed Sir Edward Heath? From what did Sir Edward Heath die? What was Sir Edward Heath killed by?)

Unfortunately for Powerset and other challengers, Google also returns the answer to these questions (pneumonia) on its initial results page, often as the top link.

Google and NLP

Marketers and the media often characterize the NLP-driven search engines as brash upstarts seeking to supplant keyword-driven Google, but this is misleading. Google is extremely active in natural language research. The resulting enhancements have been implemented behind the scenes (rather than in the UI) to incrementally improve search quality (for example, PageRank changes, mid-page query refinement, dictionary lookups and other special searches, site navigation query results, search tips (Google Suggest, released this summer), and so on).

When Google's Research Director Peter Norvig was asked about Q/A NLIs, he replied: "We don't think it's a big advance to be able to type something as a question as opposed to keywords. Typing 'What is the capital of France?' won't get you better results than typing 'Capital of France.' But *understanding* how words go together is important … Most of what we do is at the word and phrase level; we're not

concentrating on the sentence. We think it's important to get the right results rather than change the interface."

Beyond search, another less-obvious Google NLP implementation is their targeted advertising software. The stakes are very high: Google has earned $48 billion from Internet ads since 2001. Many aspects of the Google ad system are unique, including auction-based placement and pricing (see an overview here). However, it's Google AdSense that makes the most visible use of NLP (Figure 1).

AdSense is based on patented CIRCA technology from Applied Semantics (acquired by Google in 2003). CIRCA's custom ontology includes millions of words, meanings and relationships. In addition to disambiguating word senses (as with "Java" in Figure 1), Google claims CIRCA technology 'understands' web pages, enabling more effective information retrieval and delivering more relevant ads. These claims (from Google's 2003 press release) may seem similar to Powerset's recent announcements, and perhaps they are. For the industry, *understanding* remains the common, but still unattained, goal.

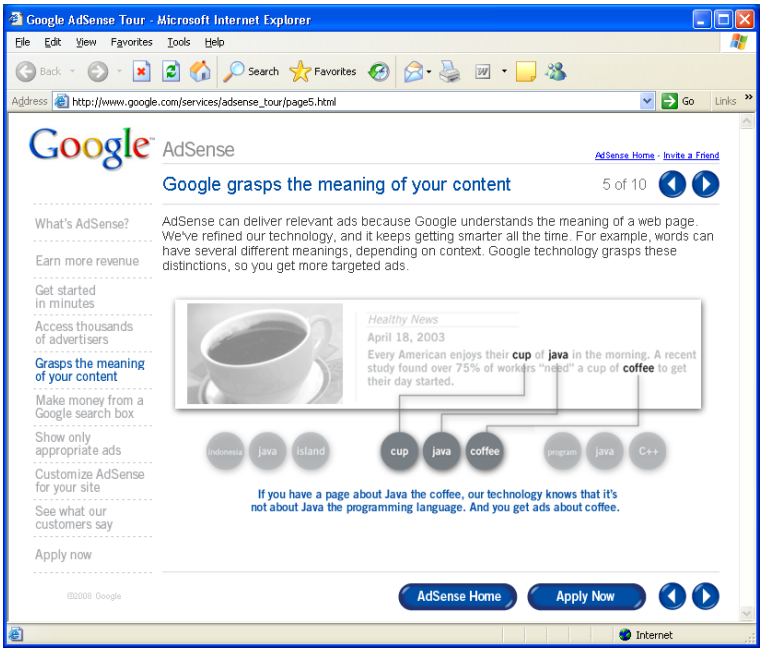

Figure 1: Google promotes NLP for AdSense and targeted advertising

Marissa Mayer (Google VP for search and user experience) sums it up: "Natural language is really hard. I don't think it will happen in the next five years." On the other hand, Powerset's Barney Pell believes, "I think we are going to look back in 10 years and say, remember when we used to search using keywords."

As Pell says, there are two camps in the industry: "*search is great, nobody can compete with Google"* versus "*search is broken, go for it*". For Pell, Norvig, Mayer, Ballmer and others – it's time to place their bets.

Acknowledgement

Thanks to reviewers for their time and suggestions; any inaccuracies remain mine.

References and Other Resources

Artificial Intelligence Meets Natural Stupidity by Drew McDermott
SIGART Newsletter, (57), 4-9 (April 1976). Later re-published in
Mind Design by John Haugeland (Ed.); Cambridge, MA: MIT Press, 1981.
(A copy can be found at http://rakaposhi.eas.asu.edu/mcdermott.pdf)

Economics of Natural Language Interfaces: Natural language processing as a scarce resource by Sumali J. Conlon, John R. Conlon, and Tabitha L. James;
Decision Support Systems, Volume 38 , Issue 1 (October 2004) Pages: 141 - 159

What's in store for question-answering? by John B. Lowe (Chief Linguist, AskJeeves)
http://www-csli.stanford.edu/~hinrich/sigdatlowe.ppt

Large Scale Internet Search at Ask by Tao Yang (Chief Scientist, Ask)
http://www.cs.hku.hk/infoscale06/presentation/keynote.ppt

Natural Language and the Semantic Web by Barney Pell (Founder & CTO, PowerSet)
http://velblod.videolectures.net/2007/sekt/iswc07_busan/pell_barney/iswc07_pell_nlpsw_01.ppt

Summary

For more than 20-years, software companies have tried to commercialize natural language interfaces (NLIs) for databases and the web. They have not yet succeeded. The "natural stupidity" of human language (Drew McDermott's catch phrase for language's ambiguities and irregularities) remains a challenge.

This paper reviewed several recent disappointments in the marketplace, including Microsoft's *English Query* (a heavily-promoted NLI for SQL Server, now discontinued) and *AskJeeves* (purchased for $2-billion as an NLI for web search, but now re-positioned in a niche role).

Some linguists warn that NLI efforts may be fruitless, but investments continue including Microsoft's recent purchase of Powerset's web NLI technology. Google invests heavily in natural language research, and the technology is pervasive in their systems for search and targeted advertising.

Drew Hamre is a principal of lexana\net and lives in Golden Valley, Minnesota.

---

[i] McDermott's essay was a reaction to the self-congratulation of the time (e.g., Nobel-winner Herbert Simon's 1965 pronouncement: "machines will be capable, within twenty years, of doing any work that a man can do"). AI is a *wide*-ranging discipline that includes sub-fields such as NLP (our focus here), expert systems, speech recognition, robotics, computer vision, and neural computing. In many of these areas, the progress has been astounding. As George Heilmeier (one of AI's strongest critics in the 1970s), said recently: "AI has been a great success ... But you may not know that because the successes aren't called AI. They're called data mining, design optimization, expert systems, dynamic factory scheduling, and lots of other names. But AI is the cornerstone."

[ii] For some, Peter Norvig will also be familiar as the author of The Gettysburg PowerPoint Presentation and as the co-author of a standard college text on AI.