Dr.Dobb's Portal
The World of Software Development

March 19, 2008
Natural Language and Database Technology
Jonathan Erickson

Combining semantics and relational database theory

Joining us today is Marvin Elder, founder and vice president of R&D at Semantra, a company that focuses on Natural Language and Semantics applied in a database access context.

Marvin Elder

DDJ: What is "conversational analytics"?

ME: Conversational analytics is a methodology that allows non-technical end users to get facts and information from databases by issuing requests in their own familiar business terms.

A more technical answer requires an understanding of where conversational analytics is positioned within the realm of Natural Language Processing (NLP). For this perspective, let's unpack NLP into constituent segments leading to the conversational analytics niche. In its broadest sense, NLP historically has been associated with text queries going against unstructured data sources: documents, emails, RSS, etc. "Semantics" is an emerging discipline of NLP that marries computational linguistics and conceptual ontologies.

Natural Language Database Query (NLDQ) is a subset of NLP that deals with NL inquiries against structured databases. The essential specialization of NLDQ is that it transforms NL requests for information into SQL or some other database query language. So, semantics and relational database theory are combined to parse requests for contextual meaning, transforming the recognized concepts into a well-formed database query that returns precise facts to the user.

Many analysts are tempted to equate NLDQ products with "ad hoc BI tools," but we don't believe a tool is very "ad hoc" if it requires database-savvy analysts to operate. For ad hoc BI tools to be useful, truly non-technical end users must be able to obtain their own reports and graphs without depending on IT resources. Sadly, the lack of inferencing power prevents such tools from producing analytics for business users.

Conversational Analytics goes beyond ad hoc BI by delivering "actionable information" to users who want or need to make business decisions based on accurate facts. Consider a "fact-rich" question such as: "Which wholesale distributor accounts in Houston have sales opportunities with revenue over $100,000 and an estimated close before 3/31/08?" Add to this the ability to redefine concepts with business jargon and abbreviations and you get real-world conversation between non-technical users and the enterprise data. This is exactly the capability Semantra has developed.

DDJ: So your product is really for natural-language database query, rather than NLP for unstructured sources. Since this has been tried many times over the years, to what do you attribute your success over past attempts? Better algorithms?

ME: Absolutely! In the quest for a breakthrough in conversational analytics, Semantra was founded on the belief that there **are** better algorithms, and that a successful melding of semantics, relational navigation and user interaction would result in technology that removed the complexities normally associated with ad hoc query products from the user experience.

As Dr. Dobb's readers are well aware, Natural Language database query systems have been attempted since the 1970s. There were a few "prototype systems" coming out of Artificial Intelligence research labs that could gather facts from databases (one was PLANES, developed at the University of Illinois at Urbana in 1975).

more

In the 1980s, my company, Software Automation, developed a 4GL for end users called Salvo, which had a built-in Natural Language feature that generated the Salvo 4GL code straight from NL requests. PC Magazine featured Salvo as one of the "Nine Best Database Products of the Year" in its September 1984 issue.

In the late 1980s, English Wizard was introduced as a shrink-wrap product, but its NL query algorithms only worked with very small databases, and it proved to be non-scalable to real-world enterprise databases.

In the early 1990s, Microsoft introduced English Query, which utilized a type of NL called "guided navigation" that was also tried in other commercialized NL products. During the installation phase, guided navigation systems required an organization's analysts or IT personnel to name the relationships between database tables: e.g., "SalesReps **place** Orders, Orders **consist of** Order Details, Products **are characterized as** Product Categories."

Using this guided navigation approach, users were required to trace the correct "navigation path" through the relationships between entities (tables). So if a user wanted to ask "which SalesReps sold Retail Products?", he/she really couldn't ask the question in conversational form. Rather, the user would enter (or be prompted to enter): "list SalesReps who place Orders having Order Items associated with Products characterized as Product Category having category name 'Retail'."

Guided navigation NL query algorithms do retrieve correct results, and Semantra uses this method for "inquiry validation." Users generally balk at having to learn a prescribed, non-conversational method of inquiry, which is why Semantra chose to adopt a more intuitive, interactive approach. Not only can Semantra's users see how their inquiry was interpreted, they can also see other possibilities in the context of their original inquiry, and can even progressively increase or decrease the scope of their request.

DDJ: It would seem that a natural-language database query system is only as good as the database it's connected to. Does the database have to have unique features, or do "off-the-shelf" databases (such as Oracle, MySQL, and the like) get the job done?

ME: Semantra's search technology is designed to work with any relational db system. We support all of the commercially viable relational databases including Oracle, Microsoft SQL Server, IBM DB2, and even open-source RDBMs like MySql. Large databases are no problem since most subject areas can be isolated to a fraction of the total database for any given query.

DDJ: Since your query product is adaptable to a wide variety of databases, can you share with our readers your selection strategy (from a technical sense) of which markets to address first?

ME: We put a lot of serious thought into which market to target with our initial product offering. From a user requirement standpoint, our research clearly led us to "horizontal applications" such as CRM and ERP. These applications usually impose a particular definition of terms on users who might define those terms differently within their vertical industry. For example, while users in a travel consultancy may refer to "agency" and "agent," the CRM application consists of more generic entities like "account" and "salesperson."

Enter Semantics, which allows users to express their inquiries in familiar business terms that are then automatically related to corresponding concepts within the CRM or ERP application. A conversational analytics product like Semantra's allows users within the enterprise to access data without the long learning curve associated with most of today's "ad hoc query" tools.

DDJ: Is there a web site that readers can go to for more information?

ME: Yes, they can visit www.semantra.com for an in-depth look at Conversational Analytics (just click on the Flash demo link).