Semantic Technology: From Sentiment to Applications

In the last few months, sentiment has become the next big thing in enterprise content processing. Manning & Napier, an investment firm, funded a number of projects for its search and content processing system that could determine what the computer scientists call "polarity" and I call the positive and negative aspect of a document.

A human can read a document and make a comment like "This customer is really annoyed at our warranty program" or "We need to get this letter over to marketing because our customer is raving about our new product." Computers, as it turns out, can do a reasonably good job of determining the sentiment of a document or processing a large number of documents and providing a report that says, "63 percent of the messages about our service are positive."

Sentiment analysis is one of the facets of text analytics that can discern the softer or intentional components of a report, an email, or other communication. The interest in tapping into the sentiment of a document captured the attention of the New York Times. The newspaper reported its own interest in sentiment analysis and highlighted several vendors active in this field.

What has boosted sentiment analysis to the rarified heights of a 2,000 word essay in the New York Times? My view is, "Social media and a potent tool for competitive advantage." The rise of Facebook and the proliferation of information from Twitter users have created awareness of social media across consumer and business markets. The competitive advantage angle has become increasingly important as examples of sentiment-centric content processing have diffused at conferences and in professional journals.

My own research into this facet of content processing identified the Southwest Airline incident as one pivot point. You may remember that Kevin Smith, a Hollywood notable, was asked to get off a Southwest flight because he was too large for the seat. Mr. Smith used social media to call attention to the Southwest's action. In a matter of minutes of Mr. Smith's exiting the air craft, Southwest found itself behind the curve. Over the span of a day, the incident went viral. Not surprisingly, Southwest Airline apologized and set up a social media program. The story flitted across the national media, and the message was not lost on other organizations' managers.

Vendors of sentiment analysis technology and related monitoring and reporting services found that after that fateful day in February 2010, their products and services were of considerable interest to many companies.

One company has become a touchstone for me in measuring the interest in sentiment analysis. The firm is Attensity, originally positioned as a specialist firm with "deep extraction" technology. The founder of the company focused on processing text iteratively. With each cycle of his novel approach, the Attensity system would generate additional data about the text. In today's lingo, Attensity was generating metatags, connections, and entities. The system would then perform a range of analytic functions. Attensity's approach to the

analysis of unstructured content was of interest to the US government and certain agencies which performed intelligence and analytic functions for law enforcement and other Federal activities. Attensity received an injection of investment money from In-Q-Tel in 2002.

Attensity, like other firms with technology of interest to In-Q-Tel, continued to win contracts in the government markets. However, Attensity wanted to grow more rapidly, so the company decided to expand its products and services for the commercial sector. Attensity's most significant decision came in 2009 when the company teamed with two German companies to deliver customer analytics.

Last year, Attensity entered into what I have described as a friendly merger. Attensity merged with Empolis (itself a conflation of several other German software companies involved in knowledge management) and Living-e (a developer of an intelligent interaction manager) to form Attensity Group and Attensity Europe GmbH. (Attensity also has a unit that focuses on the government market.) The primary investor was Aeris Capital AG, an investment firm based in Switzerland and owned by one of the founders of the software giant, SAP.

Attensity is a privately held company, but the firm has been growing rapidly. Unlike some content processing companies, Attensity's jump from a government-centric vendor to the commercial sector has been successful. As early as 2009, Attensity was trumpeting the payoff from the nascent interest in social media. The firm's blog ran a story with the prescient headline "Why the Economy Is Fueling Social Media's Growth." At a time when many companies were mired in the financial doldrums, Attensity was revving its engines.

But sentiment may not be enough. According to Riza C. Berkan:

> *I think the next phase of the search will have credibility rankings. For example, for medical searches, first you will see government results – FDA, National Institutes of Health, National Science Foundation. – then commercial – WebMD – then some doctor in Southern California – and then user contributed content. You give users such results with every search; for example, searching for Madonna, you first get her site, then her official fan site, and eventually fan Web logs.*

The firm he founded, Hakia (www.hakia.com) has rolled out what its president described as a "remarkable" service. In early 2010, the company introduced Sensenews. The service uses some of the methods of "sentiment" and applies them in a way that has struck a nerve with the financial services market. Priyank Mohan in "Hakia's SenseNews: Can It Really Tell You When to Buy and Sell Stocks? describes its semantic application this way:

> *[Sensenews is]a service to make buy and sell recommendations for any stock. Sensenews reads news sources (over 30,000 news sources), blogs (over 1 million), Twitter and performs an advanced computation to make buy/sell recommendations to you.*

Hakia's engineers know that news about a company is not mathematical in nature. Hakia has looked at precursors like Monitor 110 (no longer in business), Relegence (acquired in AOL and recycled via "the AOL way"), Need to Know (acquired by Deutsche Borse in November 2009), and the Reuters' NewsScope (now part of Thomson Reuters' Eikon http://thomsonreuters.com/products_services/financial/).

Hakia uses semantics in a way that makes sense to enterprise professionals and individual investors. In my opinion, Sensenews is an application that makes "sense" to the user. Who does not want an edge when making an investment in a publicly-traded stock?

Hakia's innovation moves beyond sentiment analysis, which is useful in customer support and advertising. Hakia's approach, according to Semantic Technology Blog (http://priyankmohan.blogspot.com/2011/01/hakias-sensenews-can-they-they-really.html) pushes ahead by

> *building a recommendation model on top of semantic filtering of the content….Their delivery model is subscription/consumer based which is different from most of the ones who have tried this before or trying now as they were/are more enterprise focused with the exception of Stocktwits.*

There are some considerations one should weigh before getting rid of one's personal financial advisor. For example, the service requires a subscription. Details about the inner workings of the Sensenews system is sparse. The latency between the availability of information and the updating of the Sensenews index is not evident.

As I look over the uptake of semantic technology in the enterprise and in consumer markets, three changes are taking place. First, there is a growing awareness that key word search is not appropriate for certain types of information retrieval tasks. A few years ago, the notion that a search box would unlock the treasure chest of high value information was exhausted. Users grew tired, then annoyed, and understandably more vocal in their criticism of search systems that generated a need for endless opening, scanning, and processing of "hits". When most of the items in the search results list were irrelevant to the user's information need, a semantic window opened. Attensity and Hakia are two firms trying to squeeze through that aperture.

Second, the semantic technology is not exposed. The systems and methods are manifested in an application that solves a problem or does something the user can look at, evaluate, and decide if the output has value. Equations, canned demonstrations, and lectures about latent semantic indexing lose sales in my experience. What is encouraging about Attensity's push into Madison Avenue and Hakia's foray into stock picking is that the outputs are useful and concrete. My view is that semantic technology is plumbing was unpopular when I voiced it five or six years ago. Today unless "there's an app for that", semantic technology will be just another content processing complexity. Made fungible, semantic applications can be magnets which pull customers.

Third, the semantic application market is just now taking shape. Are Attensity and Hakia going to remain static? No. Both firms will continue to evolve. In addition, the attention both

attract will create the now-familiar cycle of what I call the sheep approach. Innovators with semantic technology will follow Attensity, Hakia, and others into the market for semantic applications. This is exciting on one hand and a repeat of the all-too-familiar pattern of competitor proliferation. For those who are stakeholders in Attensity and Hakia or other companies working in this space, a buy out may be in the future.

I am not ready to say "a semantic revolution" has occurred. But if I draw a historical precedent, it looks to me as if there are signals that traditional enterprise search, content processing, and knowledge management solutions will face a choice: Go semantic or go away.

Stephen E Arnold, March 17, 2011

Mr. Arnold is a consultant. More information about his practice is available at www.arnoldit.com and in his Web log at www.arnoldit.com/wordpress.