

## **No Company is an Island – Beyond the Enterprise**

By Timothy Leonard

Companies are corporate citizens. They interact with their customers and the general public every day. Those interactions are not one way. The events are not solely recorded by the company. Customers and other affected parties also record the events and give their opinions electronically. Evolving from blogs, with sizable but still small readership, into giant social media aggregators such as Facebook and Twitter, everyday people can quickly and easily distribute their experiences and opinions on their interactions with your company to millions of people. Regardless if the feedback is negative or positive, it is extremely valuable information that is key to decision makers in marketing, customer support, manufacturing, fulfillment, etc. Other important “conventional” sources of data, which are now in digital format, are important to your decision makers. News websites with text, audio and video are all important sources of information.

This information, which is created beyond the confines of the enterprise, is literally a fire hose. How do we get a handle on this deluge of information? We must consider that these are all sources of data that must be integrated and transformed into a product we can use for analytics and reporting.

We are going to take a look at three new example concepts that will truly add business value – media stream data warehouse, textual integration and voice data warehouse. As we take a look at these three examples, I realize we may be far off from any of them, but the need for innovative ideas to support the business needs to keep evolving. We need to contribute to the success of the business by being creative with approaches and delivery. Are we, as a community of professionals, exploring these things? Are we up to the challenge?

### **Media Stream Data Warehouse**

How do we harness the new creative social concepts such as Facebook and Twitter into business data sources that we can use to make valuable decisions? How do we gather news as it gets reported on the TV and put it into an analytics engine that enables decision makers to react as it's reported? In the past, people would literally tape and then transcribe the news, press meetings, etc., and the scripts would be sold, on a subscriptions basis, to interested parties. They might focus on keywords such as your company name, a product name, etc.

Today, with technology as it is and will soon be, this can be automated and the coverage will expand exponentially. Again, the challenge is not the amount of available data; it is the ability to integrate a myriad of outside, media-rich data into the corporate warehouse. One way is textual integration.

### **Textual Integration**

Simply put, textual integration is processing unstructured and semi-structured data sources so they become another data source for your data warehouse and can be used in analytics and reporting. There are different approaches and techniques to textual integration. One, natural

language processing (NLP), uses grammatical rules to locate and understand words in a category. While NLP addresses certain needs in an organization, it can also introduce a level of complexity and inefficiency that can impact the primary objectives of leveraging unstructured data. NLP is often a proprietary, closed architecture, which is limiting by its very nature. You may have seen free NLP tools available to your company; but if they inhibit the ability to integrate your unstructured data then you will not realize the benefits of your unstructured data business case that are based on lower costs, revenue uplift and a competitive advantage.

Textual integration needs to be able to gather data from sources that are in more than one language, which requires special treatment to be properly organized. The current best-known practice is to utilize a multi-lingual taxonomy that organizes items by “core” concepts that are maintained universally but translated locally. By using universal core nodes, the system allows the data to be shared across the different languages without losing organization or context. I would prefer a single translated taxonomy but recognize there are other techniques, which include translating all documents to a single language and creating separate taxonomies per language.

Textual integration is here to stay. Unstructured data combined with traditional structured data provide a myriad of benefits that include business process improvement, higher revenue generation, increased profit margins by delivering cost benefits, improved CRM by a better understanding of the customer – the 360-degree view. Finally, textual integration can help provide a competitive advantage and differentiation in the marketplace.

## Voice Data Warehouse

I’m not talking about voice activation commands for the data warehouse. I’m talking about storing voice information transformed into textual information that is then loaded into the data warehouse. There have been a lot of advances in technology catering to people with disabilities, who might not otherwise be able to access the Internet and computers in the way that they were traditionally designed. These advances to provide for those who might need some assistance have produced all manner of features and software. Among these is software that reads text from the screen. It then translates it into speech to be perfectly understood by those who might not be able to read or see the monitor. It literally reads the screen and then speaks it.

We are moving into a world where communications and storing voice into the data warehouse are going to become bigger and bigger. Voice is quickly broken down into patterns and automatically transcribed into text that is then run through a textual engine for integration in the data warehouse making itself immediately available for analytics.

## Conclusion

Information about our companies is all around us. Our decision-making data is not just what we generate within the confines of our campuses. Being able to integrate information generated from social media aggregators, news sites, etc. is critical to decision makers having the whole picture. While it is important that we don’t lose sight of the basic features that got us here, we can’t lose sight of new ideas that will take us into the future. Budgets, timetables and conservative clients limit many architects. Some innovators, however, are breaking convention and collaborating to

push data warehousing into the future. The data warehouse always needs to be out in the front versus being the back end of the processing.