



May 17, 2010
Page 1 of 3

The Electronic Marketplace:

Machine-Readable Data Feeds Give Algorithms A Helping Hand With Millisecond News Filters

NEW YORK — One of the biggest problems with the information superhighway is overload. It can take an inordinate amount of time to wade through all the junk to get to useful data. The problem is even more acute on Wall Street and at investment firms where seconds can mean the difference between making and losing millions of dollars.

A new technology, called machine-readable news, seeks to solve the problem by filtering information in such a way that clients only receive relevant news they can use. These systems aggregate thousands of news feeds with easy-to-identify tags and filter them to find data relevant to traders. The systems then prioritize the filtered news to give traders actionable buy and sell signals which can be integrated into trading software, strategies and algorithms.

Machine-readable news refers to the end user, typically a computer at an institutional investor, such as a hedge fund, money management firm, or the trading desk at a broker-dealer. The most likely customer would be one that participates in high-frequency trading, ultra-fast electronic trading that happens within microseconds. To take full advantage of these computer-driven trading strategies, traders require the lightning-fast delivery of data and news. According to **The Tabb Group**, a New York-based advisory group, high-frequency trades now account for 61 percent of all the shares traded in the United States.

Bloomberg and **Dow Jones** are currently working on machine-readable news systems. Bloomberg's offering tracks 20,000 economic indicators culled from government sources, press releases, Web sites, company announcements, rating changes and unpredictable events. In April, **Deutsche Börse** launched **AlphaFlash**, a machine-readable high-speed news and data feed.

Thomson Reuters appears to have moved the concept forward the most with its **NewsScope** service, a direct result of its 2007 purchase of **ClearForest**, an Israeli-U.S. company that had created software to search large numbers of archives for relevant information, as well as its close collaboration with **Infonic**, a sentiment software company in the United Kingdom.

Global Investment Technology®

Page 2 of 3

THE ELECTRONIC MARKETPLACE

Bloomberg and Dow Jones are currently working on machine-readable news systems. Bloomberg's offering tracks 20,000 economic indicators.

Thomson Reuters's NewsScope Direct offers millisecond delivery of highly structured news and economic data to support trade execution before the market moves.

"NewsScope allows clients to easily incorporate news and events into a variety of trading strategies," says Thomson Reuters, by offering "milliseconds delivery of highly structured news and economic data which enables clients to buy and sell financial instruments before the information moves the market."

In the context of machine readable news, information often originates in three forms: structured data, semi-structured data and unstructured data. Structured information consists entirely of predefined data, typically numerical in nature. For semi-structured data, which is typically free text where certain patterns of data can be extracted, techniques can be used to extract key items in the text, such as facts or characteristics, which are turned into structured data for the computers to interpret. Finally, unstructured information electronically captures free-form text, basically anything written, such as news articles, news alerts or blogs, and turns the information into a form the computer can read.

"The most exciting opportunity exists with the unstructured information," says Richard Brown, Global Business Manager for machine-readable news, at Thomson Reuters. "That article might have only meta data that would tell what companies are in the article, but you can analyze it for other characteristics, such as the tone or sentiment of the item, its relevance, or how repetitious it is."

The NewsScope Analytics system, a subset of the NewsScope product line, analyzes articles using a technique called natural language processing, or NLP. It breaks down unstructured text by looking at the parts of speech, such as nouns, verbs and adjectives, compares how close together these words are, what the surrounding words are and other linguistic cues. The system assigns sentiment scores which measure how positive, neutral or negative an item is for a given company. In scoring for relevance, the system analyzes the item to determine if the story was about the company specifically, or just a passing mention.

"Its novelty or uniqueness detector works by looking at a vocabulary fingerprint in the article, essentially, the words and phrases contained in each article," says Brown. "When they look substantially similar the system marks the item as repetitive. These combined metrics along with others make it easier for the computers to understand what they're looking at and what to do in response. Humans can also better understand patterns that may be emerging across thousands of news items."

The stories may also contain predetermined codes to highlight the topic, or subject of an article. For example, it may note whether it's a story about a merger, a dividend announcement or a research report about the company. "The larger trend is making sense of unstructured content," says E. Paul Rowady Jr., Senior Analyst, at **The Tabb Group**, a consultancy.

The service not only supports trading by making news easier for computers to read and understand, it can also be presented in a form for humans. In addition to streams of news data, the program helps the human traders understand patterns that can only be accumulated from thousands of news items. It not only creates trading synergies, but yields important data that humans can interpret more efficiently.

While the applications can be rather simple, such as sentiment measurements,

Global Investment Technology®

Page 3 of 3

Machine readable news and information originates in three forms: structured data, semi-structured data and unstructured data.

Alacra's PulsePro presents on a single screen four specific filters: Street Pulse, Chief Pulse, Deal Pulse and Weak Pulse.

stock screening and back-testing systems for trading algorithms, it can also be used to analyze to identify much broader trends among macroeconomic data to determine correlations and market signals on which way the market is moving. "You can slice and dice the data for interpretation on individual articles or on companies over time. You can also expand that to understand trends in a sector, a specific market, or even entire countries," says Brown. "With unstructured data, the timeframe for acting on this information can be much longer, even measured in months rather than the sub-second reactions typical of structured economic data."

The machines deliver the information in a number of formats and work across a number of operating systems, platforms and databases. The output can also be displayed in ways that humans can read such as a scrolling news feed or advanced business intelligence and visualization tools such as **SpotFire** or **Panopticon**. The **NewsScope Analytics** works on Windows servers, works with SQL, and is published on Reuters market feed format. Other products work in a range of formats, such as XML and are database agnostic. Of course for the individual investor, these services can be prohibitively expensive.

Alacra, an aggregator and packager of custom information for more than 400 global financial institutions, in March released **PulsePro**, a Web application that presents this filtered information in an easy-to-access, easy-to-read form for wealth- and asset managers, financial-services professionals and individual investors. In a user-friendly Web browser format, **PulsePro** lets the portfolio manager input the company he or she is interested in researching. The program then filters through the deluge of Web-based information to extract pertinent business events, including analyst comments, CEO remarks, M&A rumors and debt restructurings.

"Sometimes the most valuable business information comes from alternative or nontraditional sources," says Steve Goldstein, Chief Executive Officer, **Alacra**. "These have various levels of credibility and investors have trouble keeping track of which ones deserve attention. We picked 3,000 Web-based sources. Half are from the mainstream media, such as *The Wall Street Journal* and the *Financial Times*, and half come from bloggers, such as those under the umbrella of the **Gartner Group** and **Forrester Group**."

PulsePro presents on a single screen four specific filters: **Street Pulse**, **Chief Pulse**, **Deal Pulse** and **Weak Pulse**. **Street Pulse** spotlights company-specific comments by 25,000 analysts, including those from the sell-side, ratings agencies, the industry and influential bloggers. **Chief Pulse** filters for remarks by 'C-level' corporate officers. **Deal Pulse** looks for news and commentary on rumored, announced, and completed M&A transactions. **Weak Pulse**, meanwhile, reports corporate distress signals such as debt restructurings, reorganizations, layoffs, bankruptcies and "seeking strategic alternatives."

Alacra says the feeds and XML-based APIs support easy integration within portals, CRM systems or other applications, including algorithmic trading models. The proprietary software is all home-grown. The price depends on how many users use the service and the number of pulses a customer wants to receive. It can range from a couple thousand dollars a year to more than \$100,000. Currently, 7,000 clients use the service. □