

# Finding alpha needles in unstructured data haystacks

Capitalizing on unstructured data



*By Barry L. Star, CEO, Wall Street Horizon*

The challenges of working with unstructured data came into sharp focus last year when a self-driving car was involved in a fatal crash on a Florida highway. The National Transportation and Safety Board ultimately found that the driver-assistance system was not at fault; federal regulators warned that they can only rely on these systems to handle *some* situations that occur on the roads.

The Wall Street equivalent changes only a few words of warning. Traders can only rely on these systems to handle some of the situations that occur in the markets. While not as catastrophic as a collision with an 18-wheeler, a bad trade based on a poor assessment of unstructured data could cost millions. There lies untold value and potential alpha in unstructured data. The key questions are how best to identify, extract and leverage that data to improve trading.

## The problem

Unstructured data does not follow a defined data model. Unlike the tidy rows and columns format of structured data, unstructured data is free-form content that people generate. It's data from email and text messages, blogs, videos, podcasts and chats. It's what does not neatly fit into spreadsheets.

Gartner and other market research firms agree that unstructured data comprises the lion's share of most organizations' informational assets – somewhere in the 80% range. Yet we're still doing a poor job of mining the value out of this data.

Wall Street firms want to leverage unstructured data to generate profits and avoid losses. They seek to wring out the value to gain a clearer picture of their markets, spot patterns and anticipate developments more effectively – taking faster action to seize opportunities while sidestepping risk. Firms ultimately have three choices. They can build the software themselves, try a shortcut that skips the hard translation and analytics by using keywords to attempt to gauge sentiment, or buy/subscribe to a third-party.

The 'build' option works best with a narrow focus of unstructured data. A good in-house development

team can create the taxonomy, text parsing functionality, analytics and data management functions required for a focused application. The problem is that you need to have a specialized focus or you leave too much value on the table.

The other build option leverages lighter-weight programs that detect and score the presence of certain keywords in unstructured data to discern positive or negative sentiment. Like the narrowly-focused ones, these systems generally lack the precision required to move the investment performance needle.

The 'buy' option has its pros and cons. Buy offerings invariably have been built by specialists and experts, and reflect the years of focus and experience they've poured into their products.

While buy options come with costs, working with specialists is usually a smart play. As the adage goes, 'You get what you pay for.'

## Choosing a path

At Wall Street Horizon, we've been dealing with unstructured data for years. While there are no hard and fast rules, certain patterns tend to pop up with unstructured data. Large trading firms have strategies that call for pulling specific, proprietary signals out of unstructured data streams. With plenty of resources at their disposal, they tend toward building solutions. It's an expensive, but highly customizable approach. The sentiment option often works for latency-sensitive momentum traders. The buy option is usually the preferred choice for market makers and traders, who need accurate, machine readable data fast and on a reasonable budget.

The best approach for any firm is dictated by three factors: specific goals or requirements, operating schedule, and budget. Despite great strides made in capturing and leveraging business intelligence, much remains locked in files and datasets. And it's growing. It is no longer a viable strategy for investment firms to continue to marginally use this data.

Whether it's building an in-house system, or teaming up with outside experts, many companies are moving ahead with unstructured data initiatives. ●