





With contributions from:



Mark Pitts Vice President, Enterprise Informatics, Data & Analytics Highmark Health

HIGHMARK.



Farouk Ferchichi Executive Director Toyota Financial Services

TOYOTA



Matthew P.T. Ruttley Manager of Data Science Mozilla Corporation*

mozilla



Ramkumar Ravichandran Director, Analytics Visa*



*Please note that contributions made by Matthew P.T. Ruttley and Ramkumar Ravichandran solely represent the thoughts and opinions of these experienced practitioners, and are not intended to represent opinions or positions of their respective companies.

This topic is being discussed in more depth by Mark, Farouk, Matthew, Ramkumar, and other text analytics practitioners from companies like Yahoo!, 3M, GE, eBay, LinkedIn, Blue Shield, Nordstrom, and CISCO at the 13th Annual Text Analytics Summit West 2014 in San Francisco, November 4-5. For more information on the event visit the website at:

http://textanalyticsnews.com/west/

The whitepaper is prepared by: Geoff Whiting, Principal, GWhiting.com Alesia Siuchykava, Project Director, Data Driven Business



Overview

Unstructured data may be the most prevalent form of information on the planet. It exists in our e-mails, surveys, social media accounts, call center logs and even reports like this. Not only is between 60% and 80% of corporate data unstructured, but the way people communicate is predominantly unstructured.

The human brain is an immensely powerful processer that can make a sense of this information, and text analytics is helping our computers follow suite.

Improvements in data science knowledge, machine learning and text analytics itself are propelling its adoption by an ever-growing number of companies seeking a competitive advantage. However, many of these deployments struggle because the space is so new that best practices are not always readily available and many existing structures are tailored to very specific use cases. This report hopes to establish some of the key barriers that prevent successful commercial deployments while providing real-world assistance so obstacles can be overcome.

It will focus on the different needs of an initial text analytics adoption, including what our contributors all cited as the top company need: strong high-level executive support to help ensure necessary long-term funding.

Text analytics can be applied in almost every business case and multiple units within the same organization can benefit from a centralized analytics division. The market's future is still a concern because of the shortage in text analytics professionals, and this reality is a guiding force for today's successful pilots and programs.



Current Benefits of Text Analytics

Text analytics can provide a wide variety of benefits in enterprise deployments because of its ability to support existing business cases. This type of analytics uses common business metrics to establish its value, so specific benefits are delivered by bolstering a company mission or vision that is established prior to the analytics program.

"Text analytics uses a lot of the traditional business key performance indicators (KPIs) because it is really there to augment existing business processes instead of being a new business process itself," **Matthew P.T. Ruttley,** Manager of Data Science, **Mozilla Corporation**

With a strong analytics unit in place, companies can achieve three main benefits:

- Simplified data: Text analytics involves taking unstructured data and finding ways to quantify it, or even convert its lessons into structured formats. This can speed up insights provided by team members and make the data easier to search and sort through a simpler dashboard.
- Improved analytics: Text analytics can provide a significant benefit to other analytics and business development efforts because it provides additional context. Many companies have found useful lessons in unusual places such as call center conversations over billing where customers mention product features they would enjoy.
- Multi-unit analytics: Text analytics focuses on a general processing and doesn't have silo issues that other data processing runs up against. Since text data is captured throughout the business, product and consumer lifecycles, analytics can be applied to various situations at any touch point. To achieve this, companies need to develop an initial framework and work with a core group of data scientists to match infrastructure with needs.



According to our expert commentators, some of the top areas where text analytics adds complementary value are in:

- Attrition rate management and reduction
- Customer satisfaction
- Customer management effectiveness
- Enhancing predictive modeling and other data processes
- Maintaining or increasing long-term revenues
- Managing brand perception
- Marketing campaign evaluation
- Product design insights

These represent a small set of areas where benefits are already being seen. However, even a complete list of areas where value has already been found can't accurately capture the potential of text analytics.

"One thing that's interesting about this type of analytics is that the data and capabilities are fairly new, so people are inventing new ways of dealing with this type of data every day. There are many applications that you and I haven't thought of yet and people are looking for them all the time,"

Mark Pitts, Vice President, Enterprise Informatics, Data & Analytics, **Highmark Health**

However, that type of innovation requires a forward looking strategy to ensure that new processes and applications facilitate adoption instead of creating impediments.

Barriers in Today's Text Analytics

Chief among the challenges faced by text analytics deployment is the wanderlust and confusion that come from not having a narrowly defined strategy. All big data initiatives, and especially initial text analytics, need a specific strategy that covers all aspects, from team structure to data governance, or they struggle.

ROI Concerns

Lacking a strategy makes it incredibly difficult to demonstrate the value of text analytics programs. This demonstrable return is important because cost is also a barrier for many.

"For small companies, hiring a data scientist is often simply financially out of the question," says Ruttley.

To combat this, companies should target low-hanging fruit through simple business problems and use cases where text analytics can provide a small but fast ROI.

Initial Costs

Costs just for pursuing these "easy wins" have sometimes mounted because of the data in question.

"The data landscape today is very heterogeneous; there are many, many different systems and practices. In the medical community, some practices are still using paper for their record keeping," says Pitts.

Limiting data sources is a much-needed step for initial deployments.



Talent Shortages

That low-hanging fruit can be hard to identify depending on the data being used, so companies often need an existing data scientist team to start any text analytics operations. This need is also a barrier for many organizations because there is a current lack of available skills on the market.

"The bulk of analysts today are either business or data analysts, who have significant expertise with non-text analytics, while text analytics is restricted only to data scientists and machine learners,"

Ramkumar Ravichandran, Director, Analytics, Visa

Legacy Infrastructure

Historically, internal system architecture, design and policies have not been developed with sharing, management and multi-department use in mind. When this type of siloed environment is present, data scientists and other technology professionals are forced to work back through a system and influence how it handles this data. These updates are sometimes entirely new platforms with unique governance challenges around how systems make and share logs, store voice recordings and allow for various access and analysis.

Understanding Bias

Much of the data used by text analytics has the potential for bias.

"Most surveys or social media comments are provided by aggrieved consumers – some of whom only comment because they're prompted by CRM execs during their call. Hence there is also some self-selection bias in the data," says Ravichandran.

Text analytics is faced with a double duty of determining the overall sentiment while at the same time attempting to couch it with estimates of how many customers are happy but nonresponsive to surveys.

Niche Challenges

There are also some industry-specific challenges where text data is more difficult to use than structured data. Pitts pointed specifically to the medical industry's struggles to remove identifying factors from records because the facts of each case and its paperwork are unique. The inability to scrub this identity can limit its sharing and keep insights restricted to a small number of business units.

For businesses that predominantly look to social media or online customer reviews (like those found on Amazon.com pages) text may not always provide enough information for useful processing.

"Although there are many challenges in processing Big Data, these are due to hardware constraints. At the other end of the spectrum the problem lies in retrieving context surrounding small text fields such as tweets that can be just a few words" Matthew P.T. Ruttley, Manager of Data Science, Mozilla Corporation

Organizational Challenges

Some companies will take significant convincing to make the analytics investment. The only way to consistently overcome this is to firmly plant a vision for company improvements that focuses on real-world, substantive criteria. Unfortunately, data is often initially abstract, making the post-conception conversation hard to initiate.

"Management needs proof of concept around a key business metric," says Ravichandran.

"When you have at least one senior executive's support and the right expertise available, organizational dynamics become less of a barrier and you're able to focus on building and maturing the capability. If you lose this, that's fundamentally a tactical roadblock," **Farouk Ferchichi,** Executive Director, **Toyota Financial Services**



Executive Support

Perhaps the most critical piece to deploying a text analytics program within an organization is for the analytics unit to have a high-ranking bankroller. Text initiatives need someone at the top decision-making table to justify their initial spend and then future investments that will allow the department to flourish and spread insights throughout the enterprise.

"We've found that a senior executive – not just an executive, but a senior executive – is very, very mportant. His or her support is essential to mplementing and then utilizing text analytics to their fullest," said Ferchichi. "When Management supports and trusts the value of using text analytics, they allow data scientists and others to fully get on board."

It's very unlikely that an executive will come to the table with a supportive vision or extensive knowledge of text analytics, so it's up to practitioners to educate and demonstrate value. The simplest way is to create quick, yet compelling uses cases that tie to existing areas of data.

"People are often too conditioned to think about text analytics solely in terms of social media. That they don't really think about the many other cool things it can do for existing business. You really have to focus on an existing business problem so you can get to value when you solve that problem,"

Mark Pitts, Vice President, Enterprise Informatics, Data & Analytics, Highmark Health

Part of this overall management vision requires belief in the importance of text analytics. The executive lead and the data team must not only believe that text analytics can help solve business problems but they must also believe that the current team has the capability to deliver on those promises. Vital to that belief is having the right person on point for analytics operations.

Starting Points

While many other data units operate in their own silo, text analytics divisions operate best when they send scientists to work within other business units to achieve improvements. The complexity of text analytics makes the overall business model simpler if a data scientist maintains their analytics duties and works with a point person from each individual business unit, instead of trying to train those other unit leaders.

Data scientists need someone who understands the potential but not necessarily the mechanics of the analytics. This guides the data scientists to the right business case and established the proper KPI, but doesn't prevent the overall analytics from having too narrow of a focus.

"Having someone available to explain to the data scientist what this language means is absolutely critical. You've got to have people that understand the language and the business process development," **Mark Pitts,** Vice President, Enterprise Informatics, Data & Analytics, **Highmark Health**

This understanding keeps the analytics on track while also showing the data scientist areas where more analysis can be applied once the initial use case is complete. This type of partnership first answers the set of predetermined questions from the business unit and then provides the opportunity to say:

"We answered 'X'; now did you know that 'Y' can help too?" says Ravichandran.

Many text analytics strategies work equally well across customer management, marketing and maintaining revenue, so the data scientists should be neutral in order to let overall business needs guide where their operations go. This also means being willing to adjust metrics to fit each situation, even if overall analysis stays the same.



Standard KPIs

Adopting new services such as text analytics often comes with research into best practices and ways to mirror the successes of others, especially when it comes to KPIs and business metrics.

However, text analytics doesn't have a set of industry-specific KPIs that new organizations should apply every time. KPIs depend on each specific project, giving text analytics some flexibility but also introducing trepidation in initial deployments.

They must be relevant to the business problem at hand, whether that's increasing profit, better customer relationships or improved cost savings; whatever your challenge may be the new analysis needs a KPI that directly relates to it.

"You need to have a business sense as to why you're looking for what you're looking for. But then, you want to allow research and the ability to expand R&D around the topic. Then you can verify that the variable is the best bet and that it fits future research," **Farouk Ferchichi**, Executive Director, **Toyota Financial Services**



Introducing New Data

Onboarding new datasets has become much easier with today's systems, and our thought leaders believe that it this will be an ongoing trend. With each new data partner, employees get better at the work needed for integration. Growth in these individual cases will spur a larger industry growth, prompting some standardization of deployments and luring larger analytics players into the text space with common toolsets.

"Standard sources will have some variation between how people populate data fields, but they're much better than other instances where there is no uniform system or where different groups define different elements of information as text," **Mark Pitts,** Vice President, Enterprise Informatics, Data & Analytics, **Highmark Health**

Some of the data will naturally shy away from this standardization, especially data on review sites like Yelp, because it is tailored to users instead of analysts. The importance of these types of datasets will likely cause two challenges to the process of onboarding new data:

- More public and social sites may start restriction what information is available and want to get paid for the data they make available.
- Third-party data aggregators may try to increase the perceived value of their information by offering slightly processed data instead of raw data.

These present problems because having to pay for currently free information creates another financial hurdle as the text analytics group will have to seek money allocated to other business units.

Differences in value for data will also make pre-processed packages less useful, even as those aggregators raise prices for data that has been cleaned.

"Ideally, we wouldn't look to buy other people's mining, just their data. We want to use our own data mining services," said Ferchichi.



Is the Future Found in Off-the-Shelf Services?

"At least for today, the build-it-yourself is almost a necessity. You can find an off-the-shelf solution where someone has done of some of the heavy lifting of structuring the data and how to apply some modeling, but each specific deployment is going to be a little different and will require some custom adjustment," **Mark Pitts,** Vice President, Enterprise Informatics, Data & Analytics, **Highmark Health**

This amount of tweaking is what most companies will use as a judge of cost savings and usefulness, says Ruttley.

"The more data you have, usually the more complex it becomes and the smarter your algorithms and machine learning need to be."

As a whole, the industry seems to be heading down the eventual path toward off-the-shelf services providing an adequate baseline structure that scientists make adjustments from. That transition will take some time, however, because the industry will need to scale significantly to make developer investment worthwhile.

"You don't want your data scientists to be maintaining your algorithm. You eventually want to leverage out of the box systems, but that needs a large industry demand," says Ferchichi.

Open-source frameworks sometimes gain ground, but that may not be the case for text analytics.

"Lack of support or training plus the complexity of UI for open-source software is a deterrent," says Ravichandran, who sees the likely future as the "commercialization of open-source frameworks."

"If I pull out my crystal ball and take a guess, I think you're going to see evolution of niche product development and adoption in very specific cases for very specific industries. Longer-term, I think you're going to see a development of the technology itself that's better at processing natural language. Once that processing occurs you'll see it all become more mainstream," says Pitts.

Attitude and Aptitude

Companies performing well in today's landscape are doing so thanks to the right combination of "attitude and aptitude," explained Ferchichi.

Attitude is understanding that text analytics is part of a larger data evolution that companies can best take advantage of through their approach to infrastructure and systems. Companies must acknowledge that they need a scalable, open-source data infrastructure. This can be hard to implement now, so companies with an existing open-framework have a slight leg up.

"Data explosions are happening every day, and big data may not be relevant in the next few years. Does your company have the right mindset to manage data and make sure that its approach is welcoming to new data and services?" **Farouk Ferchichi,** Executive Director, **Toyota**

Farouk Ferchichi, Executive Director, Toyota Financial Services

Aptitude isn't as much about current staff abilities as it is the planning for future capabilities. The data explosion will come coupled with increased needs for data scientists and machine language specialists. Unlike many other business focus areas, there isn't a large available workforce with this skillset. That means there aren't enough specialists for all companies to expand as fast as they want to in this area. Part of the problem is educating the industry about the need for both text analytics and data scientists,

"Not everyone has a text analytics team, and that's something we should work toward remedying." Ramkumar Ravichandran, Director, Analytics, Visa





When Voice Becomes Text

As text analytics picks up steam, the thought leaders believe it will lead to some standardization of Web forms and other text-based data collection.

One area that may further this standardization is a greater adoption of automatic voice-to-text transcription. Calling a bank or credit card often delivers a recorded message – 'this call may be recorded for quality assurance purposes' – but soon that recording will automatically be entered and fed into a text analytics engine.

"I think you're going to see a lot of that kind of technology deployed when conversion becomes more robust," says Pitts.

Ruttley agrees with the benefits of speech-to-text use, especially when it comes to automatically categorizing sentiment or other aspects of the text that is not readily apparent.

"This analysis can generate lots of insight but things like training for different accents is highly experimental and so it can be difficult to accurately roadmap in terms of development time."

The climb may be worth it, especially as companies in the U.S. and UK continue to onshore call centers and take closer control of those operations. Ferchichi notes that call centers and product sales must adhere to strict ethical and legal standards. An internal quality control group can now use text analytics to monitor all calls instead of just listening to a sampling of recordings. Text analytics can allow teams to identify areas where compliance isn't being upheld properly, whether that's an individual case or a company-wide process.

Quick Tip for Demonstrating A Proof of Concept

For starting that first project, work with a survey that your company has already made and received responses on. This provides an easy way to judge results because it can quickly establish the benefit of text analytics. Using your own survey simplifies finding value because you know exactly what the survey is designed to show. For repeat surveys, a quick second project can be a historical review that will show if past efforts have paid off.





Lessons for the CIO New to Text Analytics

The growing importance of text analytics is gradually driving more attention from the C-suite. The convergence of marketing and technology will also prompt text analytics discussions because of its versatility in both domains.

To guide conversations and eventual deployments, this report presents 8 lessons and takeaways for CIOs or CMOs, generated with the support of our contributors:

The pursuit of perfection in the data environment is not realistic. Look for added value but not 100% control over data, metrics and outcomes.

Por goal-setting, frame wins in terms of driving a key business metric; e.g., 1% increase in net satisfaction; drive 5% repeat business; or satisfied consumers were 5X more engaged.

Trust data over your gut, but check twice. Text analytics will tap into data that no one looks at historically. It often competes with fundamental business operation knowledge and experience, and it may be at odds with intuition. If this happens, run the numbers again and then adjust policies accordingly.

It's not governance's job to determine where data should be. Text analytics is about sharing, so governance needs to allow for open channels.

5 Infrastructure is essential. If you don't have the right data infrastructure in place, text analytics becomes a lot of ad hoc experiments that don't all integrate.

- 6 Costs will quickly rise if you don't have a strategy that allows for growth and future data integration.
- Diversity in datasets adds to complexity and challenges. Work with teams to understand their capability before pushing forward with multiple additions.
 - Whenever you introduce commercial products, know how they integrate. Don't buy them in a silo.

O'o



Closing Remarks

The lessons provided in this report were garnered from the real-world experiences of our contributors, so it is only fitting to provide them with the final word. We asked them to look to the future growth of the industry and share their thoughts on upcoming needs.

Pitts:

"If you're a practitioner and you have textual data that you think might be valuable for analytics, the best thing you can do to get the ball rolling is to determine how likely this data is to solve a business problem. Work on a quick proof of concept and get it in front of your business process owners. Use it as a teaching tool to expose what these technologies can do for your business. The best thing you can do is to educate others."

Ferchichi:

"Throughout the process, it's important to include checkpoints that ensure a complete strategy around data analytics and infrastructure processes is being maintained. Whenever you introduce commercial products, know how they integrate. Don't buy them in a silo. You don't want to overwhelm an organization with an integration challenge. You need a very clear idea how the software and your pilot program will integrate with overall business processes."

Ruttley:

"Text analytics has a huge amount of potential. It can be applied to anywhere and everywhere that people input unstructured text. There's so much information that we can process and now have accessible; and it's not just in English. There are many other languages that we can use to supplement and understand a global presence."

Ravichandran:

"The onus is on all members - analysts, consumers, tool creators and evangelists - to develop and market this to the biggest audience pool including traditional analysts and metrics-driven executives. There is no need to create a separate organization focused on text analytics, just to train the current organization and its practitioners. This data is beautiful because it's unstructured and somewhat non-descriptive."

13th Annual Text Analytics Summit West

TWO DAY THOUGHT LEADERSHIP CONFERENCE, NOVEMBER 4-5, SAN FRANCISCO, CA

Be a part of the future of text analytics!

Meet and discuss strategy with experts and practitioners from major brands:



Find out more at http://textanalyticsnews.com/west/