

The State of the Modern Text Analytics Industry

With contributions from:

- Mike MacMillan
VP, Data Analytics
About.com
- Chi-Yi Kuan
Director, Analytics
LinkedIn
- Kristinne Gumbayan
Manager, Analytics
ABC News
- Howard Lyeth
Senior Marketing Analyst
L.L.Bean
- Anne-Marie Currie, Ph.D.
Senior Data Scientist
The Advisory Board

The whitepaper is prepared by:

Geoff Whiting
Principal
GWhiting.com

Brian Parke
Project Director
Data Driven Business
FC Business Intelligence

Intro/State of the Industry

Text analytics has stepped off the sidelines and is playing a major role in applications all across the size scale. As it reaches a pervasive deployment stage, text analytics continues to bring significant wins to the companies who employ it. From sentiment analysis of real-time social posts to long-term trend spotting via deep data mining, practitioners and deployments are now comfortable enough with the data to specialize for each business case.

The industry's maturation is most notable in the technology available to established and new analytics programs. The bevy of new off-the-shelf and open-source products can be tailored to each individual need. Instead of a long search for a single solution to fix a problem, today's analytics divisions are able to weigh multiple options and consider secondary benefits such as a program's ability to integrate or scale.

"If you had an unlimited budget, you could find something to fulfill all of your needs," said Howard Lyeth, Senior Marketing Analyst at L.L. Bean. "The big problem is finding a set of solutions that work together. There are many different types of tools and goals – from customer service focuses to deep data diving tools – and they're going to overlap at some points and gaps in others. The most difficult need is finding the right combination."

The text analytics industry is working hard to solve two main challenges:

- 1) **Data.** There is a rapidly increasing amount of unstructured social data available around the Web, but there is also a reverse trend on the challenge of data availability for us to acquire and analysis. The market is still trying to balance what is made available and what is useful.
- 2) **Techniques.** Finding right solutions for fulfilling the analytical needs of the industry is still a challenge. It's no longer a lack of a choice, but a struggle to understand what is available. There are many options and avenues to explore to find the solution that fits well within a company's requirements but this is often a long journey full of false-starts when the goal is not clear.

Analytics programs of all sizes are struggling to solve these concerns by trying to find the right mix of talent, technology and execution at a scale for their business.

"Our method to address these challenges was to have discussions with industry domain experts to get their suggestions," said Chi-Yi Kuan, Director of Analytics at LinkedIn. "Next it was to find the right talent with good experiences and right attitude."

Solving problems of this scale should never be a simple process. "This is a hard decision and there are often day after day debates. Ultimately it comes down to the quality, cost, and effectiveness of your solution that decides the winner," said Kuan.

To help others solve these challenges, our industry experts have taken a look at the culture, data and technology available to the market and offered up the thoughts you will find in the rest of this report.

Crafting a Text Analytics Mindset

What's Needed in a Data Culture?

The modern data-positive culture is one where information is available to every business stakeholder who relies on analysis to make daily business decisions. According to our experts, culture has proved itself to be one of the most important – if not the most important – aspect to developing a strong text analytics program.

A key element for a thriving data culture that drives business intelligence and proper decisions is a willingness to let data flow through an organization. More and more roles are placing an emphasis on data knowledge, so analytics teams can't keep too tight of a grip on information.

"Data teams sometimes take a stance as being a gatekeeper to data. They make individuals come with a request and then perform all of the analytics and provide a review of that for the business unit to fulfill their reports," said Mike MacMillan, Vice President of Data Analytics About.com. "In this situation, timeliness is often measured in days and weeks even if the business relies on it to make a split-second decision."

Not only does that approach make real-time decisions difficult, but it also introduces another long delay period if the business unit has a follow-up or clarifying question.

"In the modern enterprise there's no room to not be data-oriented. Things are just moving too quickly to always rely on someone else to provide you with every number at the end of the day," said MacMillan.

For many, having that data orientation requires in-house expertise for coding and development. A knowledgeable staff can help a company develop its own tools and work with more information before needing to make a specific investment.

"The little things that were doing now are pointing us to find out where the information can be useful now and where it can take us," said Lyeth. "We've come up with ways to do things on our own for now and that allows us to keep on working with the data. I think if at some point in time we had jumped on to one of the tools – like a customer service focus – we would have stopped thinking about all of those things."

On an individual basis, the culture around what analytics talent should do is also changing. Businesses are looking for someone who can help them make better decisions for the company. It's more than just crunching numbers.

From Report Writers to Storytellers

A key part of the new data culture is the shift to storytelling and insight sharing, instead of a focus on report generation. This means analytics teams must work with stakeholders to identify the aspects of the business that need improvement, and then learn the business language to share relevant insights.

"You don't want to be the person or department that crunches numbers. You want to be the department that's telling stories with the information," said Kristinne Gumbayan, Manager of Analytics at ABC News. "Being able to take the complexity out of the numbers and put it down into a clear and

concise manner that executives and product teams can understand is a trait that anyone should look for and companies are already demanding.”

In the realm of text analytics, this involves the ability to take overall sentiment and look at long-term results. Pattern recognition also plays an important role because it helps teams determine if a language shift has occurred that could skew past models or understanding.

“The process of adopting text analytics is iterative and checking in with the business stakeholders on progress should be done throughout the process to ensure that in the end, the business need will be met,” said Anne-Marie Currie, Ph.D.

Text analytics teams operate well when they’re essentially a shared service. They can tell a more holistic story than all of an organization’s data is pooled and there’s a focus on analytics that don’t just pertain to one realm of the business. Having both knowledge for a specific division you work in and providing competitive data that’s relevant for the whole industry is useful and will be necessary for future buy-in.

Buy-In Has Changed

More and more companies have learned that not only is the processing of digital information the future, but that analytics as a differentiator is already a market reality. The need for data points and the ability to make more decisions off of those data points has become much more apparent.

The market is shifting from analytics being a differentiator that allowed many to get ahead to a situation where companies now see the need to develop a program just to try and catch up.

“For smaller companies, there is a bit more talk-to with upper management and leadership to get investment. When you’re looking at larger companies, you’re typically seeing a stronger buy-in,” said Gumbayan.

Our thought leaders said that executive buy-in has shifted to a model where analytics teams need to understand the goals of different stakeholders across multiple business units. Buy-in now requires a stronger data understanding and the ability to determine which metrics can be tracked to provide the desired business intelligence.

“Understanding the business as a whole and what your different stakeholders want is a major challenge because sometimes they don’t even know what they want. It’s up to you as the analytics person to drive the story and mold it for them,” noted Gumbayan.

Part of delivering that insight must be a commitment to the truth, so that models are correct and expectations are realistic.

The Academic Mindset

Text analytics is becoming more academic in its processes and that holds significant potential for the market. This isn’t necessarily an inclusion of academics themselves – though some teams are bringing linguistics professionals on board – but it is an inclusion of their commitment to validated, truthful reporting.

“The mindset has been a blessing. It helps our team know they don’t need to artificially inflate numbers. They’re emulating academia where there’s a call to accurately report the numbers exactly as they are,” said MacMillan.

That focus gives teams the ability to spot trends on broader research and infuse individual skillsets and experience to better understand a context that's more subjective. Mapping sentiment with tone, colloquialisms, and abbreviations can take more time.

Adopting the academic mindset may provide this buffer and give insights not present in a more-traditional model focused on minimal experimentation related to matching actual information and forecasts.

"Having at least one senior, experienced text analyst or data scientist who has both leadership skills as well as proven technical skills in processing narrative text resources is important to this mindset," said Currie. "This will help to ensure a mature and stable text analytic program that can be leveraged to solve diverse business goals across stakeholders within the company."

Adopting this mindset comes with risk, but that too needs to be thought of differently for many companies.

Making Risk Tolerable

Successful analytics programs have realized the need to be data-driven and have embodied a desire to continually move past the status quo. Declining costs for technology and product development are accelerating these advances. However, a risk-averse culture can still get in the way of true innovation and smart modeling from text analytics.

"Organizations need to adopt a data science mentality that's willing to take a hit with analysis," said MacMillan. "There were definitely times we took a hit to Web traffic or other important metrics as a means to an end to really push our analytics."

When designed with analytics in mind, losses are often just a temporary consequence of change. Through small testing and sentiment analysis, organizations are able to see long-term goals and improve the customer experience in a way that drives future growth.

The information provided by text analytics can serve as a primer for that change and help guide organizations to keep harm to traffic, engagement and other touch points at a minimum.

"There is also the need to manage expectations for what can and cannot be achieved with stakeholders. Knowing what your business needs ahead of time and setting appropriate expectations for all involved will save you a lot of headache in the setup of the program and the choosing of the methodologies and processes employed," said Currie.

With a handle on risk culture, text analytics experts are better able to tackle data and technology challenges with less concern.

Developments in Data

The Prioritization Game

Historically, Web businesses were dependent on search engines to drive the business; it's how everyone got traffic from point A to B on the Internet. Social media has changed all of that.

"Social media is getting more and more influential. Large corporations, small and mid-sized businesses, and their marketers spend tons of time and money to hear the voice of customers through social media," said Kuan. "They want to build the closed feedback loop to improve products and services. It is simply not negligible."

Part of the shifting landscape from social media is that it's dominated by real-time information. Analysis of the flurry of text on platforms such as Twitter and Facebook provide a continuous shift in social sentiment. Such shifts mean companies can no longer afford to do their processing and provide batch data at the end of the day.

Revenue drivers and social spikes happen continuously on a mostly unpredictable cycle, and real-time responses are favored. This response is getting easier when companies have the right technology and the right priorities for their business.

"You don't have to optimize for every social network, just match your goals with their strengths. Facebook and Twitter are very valuable for up-to-the-minute details and current events; while authored content with a long lifespan can see incredibly engagement on networks like Pinterest," said MacMillan.

The strongest drivers of traffic, engagement and positive sentiment all have a role in this prioritization game. Social signals can provide great insights into what content and what parts of your content people engage with most. This gives insight into why they're engaging and the opportunity to optimize for that element moving forward.

Part of this understanding also includes monitoring sites and services that don't fall under the traditional guise of social networks, suggests Gumbayan. "For example I look at BuzzFeed and if I'm in the mood to laugh. For many, Facebook has become a news and link aggregator and people click on the items that are interesting, and they're on these secondary social media sites."

Businesses with strong text analytics can quickly find out on which networks their strength lies, but will still need to stay on top of social developments and ask tough questions.

Social's Ever-Moving Goal Posts

Where does an emoji fall in the realm of text analytics? What does each of the multitude of smiley faces mean to your system? Do all of today's consumers have the same understanding and desired usage?

"I definitely think analytics on its own is going to get more difficult as time progresses, especially with social media," said Gumbayan. "That's not because the analytics behind it is incorrect, but it's about keeping up with social change. Look at social media and the way kids, really the whole internet, are writing these days. This conversation style is progressing at such a far enough pace that it's unlikely we'll ever be able to tell everything about what a person is trying to say based on what they write."

And managing that change is a main focus because the rewards appear large when starting. Part of that excitement is also the ability for analytics to explore brand new ground.

“This type of data has its own distinctive style and content and provides a fertile ground to develop new areas of expertise within the field of text analytics. It also presents unique challenges due the volume and nature of the discourse style,” said Currie.

And at the end of all this effort, some brands may not find social to be extremely valuable. For those where customers are continually providing feedback through other means, there is already a wealth of text to analyze. Adding the uncertainty of social media – for example, not being able to parse out actual customers from a crowd of commentators – can make it an area where an ROI is hard to either obtain or discern.

The work in social may eventually turn into a renewed focus on classification, as opposed to trying to always play catch-up. The industry may move away from sentiments such as “a person talked about this and it was positive,” to trying to understand the other elements of language that tell you what kind of a person they are and what insights can be created or changed with this persona development.

Sentiment may see a new focus that changes the way it is classified and changes what the industry is looking for as it collects and processes information.

“When someone speaks, there are accents, intonation, tone, dialect and all kinds of things that we don’t really pay attention to now, except to translate it into something else. It’s possible to go beyond what a customer is saying and use these other elements of language to abstract more information. The industry isn’t sure what it’s going to find, but it will be a future focus,” said Lyeth.

One thing that the wave of social media data has begun to cause is a pause on some open-ended data. Analysts are starting to look for a balance between that information and structured sets.

Easing Responses

One surprising trend in the text analytics space has been to adjust and even limit the areas where text is collected. Open-ended questions and data gathering have shifted and is growing at a slower rate than other collection paradigms at some brands.

According to our experts, this is due to a shift in limitations and priorities. It’s less of a concern about adding cost with each new box – like it was in the past – and more about speed of analytics and usefulness. The return and usefulness of the data is at the center of the conversation, as opposed to the cost or infrastructure to collect it.

“If we introduce a new study, we have to keep things manageable,” said Lyeth. “In the past we often would have provided the opportunity for an open-ended response, but as we start to think more and more about the ease of analyzing text responses we work to identify areas where we can guide these responses.”

Technology's Three Questions

Are Systems Mature Enough?

Digital information is growing at an amazing rate and a significant amount of the data that can be analyzed has strong text elements. As the Internet evolves and the analytics space progresses, the need for analytics is poised to grow significantly.

In the past, technology was a limiting factor of that growth. Off-the-shelf and third-party tools have developed to become good enough for most deployments and there is significant momentum in toolkit development.

Now there is a maturity to the market that shifts away from growth limitations to growth guidance. Instead of focusing on where existing technology will limit analytics, text teams are now considering a software purchase versus growing the analysis skills within their teams.

"The challenge of third-party systems is that they take the hassle away from you and provide strong results, even in the realm of text analytics, but they charge for data in to the system, out of the system and all of the processing in between," said MacMillan. "It's hard to tell, especially if you're growing rapidly, if it is more costly to have your analysis performed in house or having third-party costs grow as your business grows."

Our experts found that there are multiple avenues to getting the right blend of research or modeling tools with product-focused analytics. The right toolset is often a mix of open-source, third-party and internal development.

"In addition, actively evaluate the needs of your text analytics program before it is setup. Will you be needing a research-based text analytics program or a product-related text analytics program? Ask those questions," said Currie.

To optimize the selection process, there should be more discussion among all of the available stakeholders. The better the text analytics team understands the problems they're being asked to solve or the business insights desired, the higher the likelihood that the text analytic solution will meet the need.

This has provided new insights such as company wins for the customer journey. Aligning tools with customer experience teams can help to remove pain points and increase retention rates. This new realm of insight has significant value, but it requires a blend of technologies that must operate together.

Can You Beat the Integration Hurdle?

Almost every text analytics problem or metric has multiple solutions on the market. The concern is that these open-source and third-party programs are often built to live in a data silo. They're often difficult to integrate without significant work from your own development team.

"It all has to work together. You need to do a lot of thinking beforehand and make sure you're investing in the right tool," said Lyeth. "If you're in a large company and you've got a complex system that's been in place a long time, it's a big boulder you're not going to be able to move without a lot of help."

Integration also crops up as a significant problem when organizations lack future planning for metrics, tools and needs.

"One thing that's good to do prior to starting an analytics implementation is talking to all of the stakeholders. Talk to every part of the business to be aware of what could be asked of you in the future," suggests Gumbayan. "That way you can get that type of analytics implemented or it can be on your radar to acquire new software to match the needs of each part of your business."

The long-term integration hurdle is less about the existing products themselves and more about their interaction with a database. As information grows and expectations about what can be done increase, there's a need to keep data pristine. However, data has its own concerns for many text analytics programs.

Is There Too Much Data?

"Up until a certain point you can start simple and continue to add information to your analytics. You'll need to judge where it makes sense once you get started because there's only certain areas where the scale provides enough benefit," said Lyeth.

However, taking this little-by-little approach can cause problems later on if a business intelligence focused storage paradigm isn't in place. "If you're growing and adding without organization, it catches up and it's a pain to go back, assemble it and keep it all in one place," warned Lyeth.

If this mindset is adopted, then there is little reason for a growing organization to let some data fall away.

"Track as much as possible. There's never such a thing as too much data. If you can track it, then why not?" asks Gumbayan. "You never know where the future need will come from or what your stakeholders will need in the future."

Where this differentiates is in the age of the data. Many companies have databases of past information, though it is often stored in a way that's hard to look through. It was collected before a company could have reasonably developed a text analytics mindset.

Sometimes it's worth digging in to older information, but not always. It's an eventuality, but not a major immediate need for most brands with consistent customer communication.

What these technology questions do highlight is that the market is making high-level considerations about its growth. That thinking makes it an exciting time for text analytics and the future is full of promise.

Areas of Coming Improvement

There's a lot of excitement for what the future of text analytics holds. Some of it comes in new areas of modeling and analytics, while others look to improved insight sharing and an easier method of connecting data to business wins.

Areas where multiple experts say they see improvements coming include:

- A shift from analytics as reactive to sentiment to a proactive method where future sentiment is predicted by current trends. Essentially it would be analysis of the precursors to thoughts, mood changes and customer interactions.
- Blending of technical and scientific elements to better integrate platforms. The hope of many is that systems will combine under something akin to a unified dashboard to relate data directly to business intelligence, customer experience and predictions on the interplay between the two.
- The conversation is expected to shift about what's needed, instead of what's available. As new tools come into play, businesses will make a judgement of how many different analytics tools your team can handle versus what your budget can handle.
- Businesses are expected to place a renewed focus on service tickets, in-product feedback, and customer satisfaction survey in new areas. Our experts placed a specific emphasis on reviews within the apps space.

"Gold standards to ensure appropriate accuracy testing are sometimes overlooked or not setup properly and can negatively impact text analysis results," cautioned Currie. "If the data you have access to is rich in meaning, the opportunity to dig deeper to draw out the subtle yet valuable patterns and trends should not be overlooked."

All of the challenges and hope of the market can lead us to takeaway one keep point: Never underestimate the value of text data or of any analytics program's ability to keep digging deeper for new, meaningful insight.

Conclusion/Lessons from Our Leaders

The lessons and approaches detailed in this report were forged by the real-world experiences of our contributors' developed in the text analytics arena. It's only right to provide them with the final say on how the market looks today.

We asked them about their hopes, concerns and what advice to give to the industry to help it grow through today's shifts.

MacMillan:

Text analytics technology and business intelligence are changing at a breakneck pace, especially over the past two years. Innovation means you should look for opportunities to not reinvent the wheel. Small third-party or open-source tools can give you a lot of punch for a relatively low cost. Text analytics is the backbone of learning and focusing on new insight helps to prevent any wasted money, time or resources on processes.

Lyeth:

Almost every time you talk to someone else or see some new kind of tool there's something that you haven't thought of; you'll be glad to learn it. It's helpful just to be around and talk with people from different companies as well as vendors. Even if you already have some idea of what you're looking for, these conversations can keep you thinking about it in new ways. You'll improve your understanding when you pair it with that internal map you've created of your needs.

Gumbayan:

An analyst shouldn't just be a data person, they need to be a social person as well. I think a great trait to have is the ability to speak concisely and clearly with a lot of different groups. On the technical side, exposure is essential. Excel is definitely something you should work hard to learn. If you're like me it'll become more fun as you get older, I promise. Anytime you get a chance to get exposed to industry numbers or insights, take advantage of it.

Kuan:

Dream big and make it scalable. With recent advances in data infrastructure, machine learning, and text mining, scalability is becoming more realistic in many cases. The industry is headed to a place where it can leverage text analytics to enable every business opportunity to take advantage, especially in areas such as content marketing, advertising, and real-time feedback loops.

Currie:

The future of text analytics is already heading in and will continue to head in the direction of more tightly linking multiple data sources together from different industries. That will include social media, financial, retail, marketing, criminal justice, educational, and medical data. A stronger focus on population identification and identification of behavior trends and prediction within all of these areas is

already growing. With more connections that can be mined, privacy issues will become an increasingly fundamental part of the text analytics landscape.