



Text Analytics:

Risk Mitigation with Text Analytics

FOREWORD

There's a lot of fanfare about using text analytics to "hear" the voice of the customer and use this analysis to improve customer experience, drive customer satisfaction, predict the next best action, or determine what to offer a customer. Although I agree wholeheartedly that these are important endeavors and should be undertaken, there's a lot more that companies can do with text analytics. If corporate leaders thought about all the text data in their archives, they would begin to realize that there's more than just customer voices *and* more they could be doing with the richness of these data.

Think of the text data in corporate emails. Think of the text data in call center logs—both customer service and technical support. Think of the text data on customer forms. Think of the audio files that could be converted to text. The list goes on and on!

What if you could mine these data to expose areas of risk in your organization?

We [published an e-book](#) in which 28 industry experts provided advice on how to derive business value from text analytics. What emerged from these essays were several examples of how to use text analysis to mitigate risk. These examples range from preventing terror threats to insider trading to finding defective products before your customers do. We hope that you enjoy this subset of essays and that it gives your organization some ideas about how you can achieve business value from and minimize risk exposure with text analytics.



Vicki Rollins
VP Marketing
Angoss Software Corporation

About Angoss Software

Angoss is a global leader in delivering predictive analytics to businesses looking to improve performance across risk, marketing and sales. With a suite of big data analytics software solutions and consulting services, Angoss delivers powerful approaches that provide you with a competitive advantage by turning your information into actionable business decisions.

Many of the world's leading organizations in financial services, insurance, retail and high tech rely on Angoss to grow revenue, increase sales productivity and improve marketing effectiveness while reducing risk and cost. Angoss serves customers in over 30 countries worldwide.

INTRODUCTION

Two great forces are converging on businesses all around the world. One is a tidal wave of unstructured data in the form of text, audio, images, and sensor inputs. The other is a whole new generation of data processing technology, including low-cost, scalable cloud storage of almost unlimited size, and new techniques for quickly analyzing unstructured data. The result is an explosive growth in knowledge and insight.

Although analyzing text for insight is not new, what has changed in recent years is the ability to mine vast quantities of text—such as all the content on the Internet—and to do it quickly. This capability is profoundly changing how businesses use information to learn about markets, trade on knowledge, and refine their operations. Yet text analytics methods and techniques are rapidly changing. So, what are the best ways to extract value from text? With the generous support of Angoss, we posed the following question to 28 text analytics experts:

What advice would you give someone in your industry to get business value from text analytics?

The responses we received reflect the vibrant and evolving state of this emerging technology. One startling revelation that jumped out at me as I read these articles is that just as machine learning speeds the breadth and depth of analytical insight, machine-driven text analytics is having an extraordinary impact on the speed of human learning.

Even if you are not currently involved in text analysis, you cannot help but feel captivated by the insights this e-book contains.



All the best,
David Rogelberg
Publisher



Mighty Guides make you stronger.

These authoritative and diverse guides provide a full view of a topic. They help you explore, compare, and contrast a variety of viewpoints so that you can determine what will work best for you. Reading a Mighty Guide is kind of like having your own team of experts. Each heartfelt and sincere piece of advice in this guide sits right next to the contributor's name, biography, and links so that you can learn more about their work. This background information gives you the proper context for each expert's independent perspective.

Credible advice from top experts helps you make strong decisions. Strong decisions make you mighty.

Risk Mitigation with Text Analytics



Liang Zhou

Bloomberg L.P.....5



Shree Dandekar

Dell Inc.....7



Elizabeth Rector

Cisco Systems.....9



Daniel Angus

The University of
Queensland.....11



Alessandro Zanasi

Zanasi & Partners.....13



Luca Toldo

Ph.D., biology.....15



Steve Gardner

RowAnalytics Ltd.....17



Alexy Khrabrov

Nitro.....19

THE BUSINESS VALUE OF TIMELY, RELEVANT MARKET INSIGHT



**LIANG
ZHOU**

Head of Data Sciences,
Bloomberg Data,
Bloomberg L.P.

Dr. Liang Zhou leads the effort to revolutionize the data business at Bloomberg's Global Data division. Her work focuses on data strategy and "productization," but her research focuses on natural language processing, machine learning, predictive modeling, and quantitative research. Liang received her Ph.D. from the University of Southern California in computer science, with a specialization in artificial intelligence, and a master's degree from Stanford University in computer science.



Download the full e-book: [Text Analytics](#)

In the financial services sector, where Dr. Liang Zhou leads data science initiatives at Bloomberg's Global Data Division, it's critical to provide relevant insight into important market developments to financial professionals as quickly as possible. One example, Zhou explains, would be "If the government were trying to enact a policy change, perhaps by issuing banking regulations that would really enhance or restrict some banking practices." That type of information is typically published as real-time news, headlines, or articles. Headlines briefly summarize what's taking place, but longer documents explore the reasons why it's happening or what the resulting impacts might be. By performing text analytics on those documents, she notes, you can uncover that richer insight quickly and in real time.

Similar techniques can be employed to assess how certain developments will affect the market or, in some cases, a particular company or stock. Explains Zhou, "Say, for example, Company A and Company B are direct competitors and Company A rolls out a product before Company B does. We would parse out that information in news, and then reveal the resulting insight to our client so the client would immediately know that this development would not only affect Company A, perhaps causing Company A's stock to go up, but also Company B because Company B is behind in rolling out a competitive product." Revealing relevant insight quickly and accurately is the key to realizing the greatest possible business value.

“ We could convert a news article about a company into a number that expresses positive sentiment . . . and negative sentiment. ”

KEY LESSONS

- 1 Timely, relevant insight is key to realizing the greatest possible business value of text analytics.
- 2 Machine learning makes it possible to apply insight gained in the past to future documents.

THE BUSINESS VALUE OF TIMELY, RELEVANT MARKET INSIGHT

Another powerful method for driving insight involves converting textual content into numerical representation. For example, says Zhou, “We could convert a news article about a company into a number that expresses positive sentiment, as in this is good for the company, and negative sentiment, reflecting something that is bad for the company.” Her team produces those data for clients to embed into their time series analysis, making it unnecessary for the client to have to process the text information any further. “We produce a set of numbers to reflect what that article’s about, and the client can ingest that and incorporate it into its quantitative model,” she explains.

Machine learning powers many of these text analytics processes. It is possible to use insight from the past learn a technique, and then attempt to apply that technique to a new document. For example, Zhou says, “If we once saw a particular article that was really negative about Company A, then we might have discovered at that time that certain words and content carried a really negative connotation. We could capture that sentiment knowledge through machine learning and apply it to any documents we might encounter in the future as they arrive in real time.”

Language doesn’t change all that frequently, but it does evolve. Some words, once considered indicative of negative sentiment, can start to carry a positive meaning all of a sudden. “For example, a lot of young people now say, ‘This is killer,’ which means something is really good,” observes Zhou. Those words are not likely coming through in news articles, but they can be discovered over time through machine learning. You can apply this technique to social sentiment, market sentiment, or company sentiment. Some nuance is associated with each of these types of sentiment, but you can apply a core technique such as this one across those different sentiments.

By providing timely insight to key decision makers, ensuring that they are kept abreast of all the relevant information they need to have at their fingertips to shape their strategic direction and stay competitive as the market changes, you can demonstrate the business value of text analytics.

“

We could capture that sentiment knowledge through machine learning and apply it to any documents we might encounter in the future as they arrive in real time.

”

AVOIDING CATASTROPHE



**SHREE
DANDEKAR**

Executive Director, Product
Management, Predictive
Analytics, Dell Inc.

Shree Dandekar has been at Dell for the past 15 years in a number of roles, from software design, product development, and enterprise marketing to technology strategy. Currently, he is the executive director of product management and strategy, responsible for developing and driving the strategy for Dell's business intelligence and analytics solutions.



Twitter



| Website



| Blog



Download the full e-book: [Text Analytics](#)

There is a common misconception that text analytics is useful only in the social media space, says Shree Dandekar, executive director for data analytics at Dell. That is false, he asserts.

"Text analytics is almost like a foundational block for creating a solution that has any kind of semistructured or unstructured data," Dandekar states. "Nobody can deliver true solutions without it."

The technical and business goals for much of Dandekar's work at Dell are tailored around social media, but he looks well beyond that when offering his advice for deriving business value from text analytics. His suggestions:

- **Listen in.** Listening to customers means more than monitoring Twitter, Facebook, and LinkedIn. Those platforms contain only 40 percent to 50 percent of the data set, Dandekar says. Much of the conversation takes place elsewhere—on online community sites, chat sessions, customer support logs, and other vehicles. "Where the serious conversations happen is in some of the back-end communities," he states. "That is critical when you are talking about social media analytics."

“Where the serious conversations happen is in some of the back-end communities.”

KEY LESSONS

- 1 Listening to the pulse of the customer base requires more than social media monitoring.
- 2 You can glean critical business insights from any semistructured or unstructured data set. A hybrid model is the ultimate goal.

AVOIDING CATASTROPHE

- **Monitor trends.** Collect, record, and analyze textual data. “When you have your data set, you have to spend a lot of time modeling the data,” Dandekar states. Tools are available, for instance, that can assemble word clouds to help sniff out top trends and topics. Such tools are helpful, but Dandekar prefers building taxonomies to generate a hierarchy of relevancies. “You can filter your entire data set against a predefined taxonomy to help you understand where all the conversations are taking place,” he says. “That’s the first step in creating a high-level data model.”
- **Contextualize.** Unstructured data require context to produce actionable insights, according to Dandekar. Natural language processing–powered sentiment analysis, for example, can identify when conversations about your company or its products subtly take a negative turn. “Without the business context,” he says, “the data are not going to make sense.”

Dandekar describes an instance where context helped stave off a consumer revolt. At the time, Dell was set to release a notebook PC geared toward developers; it had announced that an open source operating system would be installed on the computer. Early social media feedback was glowing. Late in the process, however, a pricing manager failed to take into account the open source operating system when announcing the machine’s price. Online conversations began to sour, a fact that was revealed only because Dell was monitoring conversations with the aid of text analytics.

Dell had developed taxonomy for the machine that allowed it to drill down by topic, Dandekar says. When it came to the subject of price, the notebook’s Social Net Advocacy (SNA) sentiment scores were tanking. Realizing what happened, Dell took action. Within 24 hours, the price was reset and apologies were issued by email, Twitter, and blogs. The SNA scores rose.

For Dandekar, that incident’s implications extend well beyond social media. Critical business insights can be gleaned from any unstructured business data—not just social media, but also customer relationship management data, enterprise resource planning data, and transactional analytics. In fact, he says, the boundaries around customer data sets are collapsing.

Creating a hybrid data analytics model, then, is the ultimate goal, but for now, Dandekar observes, that goal remains elusive. It is the subject of a much bigger philosophical discussion that many businesses—Dell included—must have soon.

“

When you have your data set, you have to spend a lot of time in modeling the data.

”

LISTEN TO WHAT THE DATA ARE TELLING YOU



**ELIZABETH
RECTOR**

Sr. Manager, Strategic
Marketing, Cisco Systems

Senior manager in strategic marketing at Cisco Elizabeth Rector and her team developed an award-winning influencer intelligence program and an integrated voice of the customer practice used to drive strategic executive decisions. Prior to Cisco, Elizabeth consulted with top-tier multimedia agencies such as Televisa, Adobe, and Logitech. She has also held marketing positions at Oracle, XO Communications, and iUniverse. Elizabeth holds a B.A. in international business from Washington State University and an MBA from IESE, Spain.



Twitter | Website



Download the full e-book: [Text Analytics](#)

Companies often think about themselves and their markets in terms that make sense to them, even though that terminology may not be understood by the broader market. “Text analytics can help companies discover true market perceptions,” says Rector, “but only if the analysis is done in a way that reveals them.” If a company models its analysis on a narrow view of the market, then it will get a narrow answer. “You have to be thoughtful about looking at the data without the company lens. It’s the discovery aspect of this that is so powerful.”

Sometimes companies begin using text analysis in one group within a larger organization, where people are looking at just one type of data. “The most important thing is to integrate data from multiple sources as much as possible so you can get a more holistic view,” Rector advises. One type of data can answer a pointed question but not necessarily reveal a bigger picture.

Rector also points out that you have to be open to the possibility that you are not thinking about a question the way the rest of the market is thinking about it. “Big companies tend to approach questions from a company-centric perspective,” says Rector. To gain that broader perspective, you must strike a balance between framing and answering a specific business question and taking a broader look at what the data are saying.

“Text analytics can help companies discover true market perceptions, but only if the analysis is done in a way that reveals them.”

KEY LESSONS

- 1 Tap into imagination and fantasy to find creative new applications for text analytics.
- 2 One type of data can answer a pointed question but not necessarily reveal a bigger picture.

LISTEN TO WHAT THE DATA ARE TELLING YOU

Rector says, “I start with the question people are looking to have answered. We’ll look at the data through that lens.” It’s best to begin with a broad topic that is relevant and a broad selection of data sets without trying to impose a model on the findings before the analysis occurs. “By doing this,” says Rector, “you are able to see how others are talking about that topic.” Then, use iterative analyses to become more specific in what you are asking without losing sight of the broader market terminology.

There are advantages in applying the same analytics methodology to different data sets. For instance, by analyzing social media and analysts’ reports, you can identify discrepancies between what analysts are saying and what the broader market is saying. With follow-on analyses that drill into the significance of those discrepancies, you may be able to differentiate leading indicators of a trend from brief, transient events.

Another example is looking at differences between what chief information officers (CIOs) say in one-on-one interviews and what they are talking about in media and to analysts. “Without looking at the CIO media component, we wouldn’t get the complete picture of their market view,” says Rector.

When using data from multiple sources, it is important to balance their mix and integration, advises Rector. “We look at social and Twitter separately from analyst reports, for instance, because the volume of social data is so much higher.” Putting these two data sets together in the same data pool would produce results dominated by the volume of social media content. They should be analyzed using the same methodology but not necessarily in the same data pool, except when you are analyzing the relationship among the media types, such as amplification and engagement.

Rector says, “Valuable data analytics comes from not biasing the analysis with a narrow, company-centric view and being open to what the data are telling you.”

“

I start with the question people are looking to have answered. We’ll look at the data through that lens.

”

MEANINGFUL ANALYSIS DEPENDS ON CHOOSING THE RIGHT DATA



**DANIEL
ANGUS**

Lecturer in Computational
Social Science,
The University of
Queensland

Dr. Daniel Angus is a computational social scientist who specializes in the design, development, and application of advanced computational methods for the study of communicative practices. He is a co-inventor of the Discursis software technology, which has been used to reveal the dynamics of topical exchange in health, education, media, and political conversational contexts.



Twitter | Website



Download the full e-book: [Text Analytics](#)

Dr. Daniel Angus, who began his career as a computer scientist, came to text analytics in an interesting way. As a postdoctoral research fellow, he became involved in researching how animals navigate. Evidence suggests that many aspects of physical navigation are governed by the same part of the brain that processes language, the hippocampus. That revelation was the basis for a whole new approach to language and text processing. “We began thinking about processing communication data by looking at it through the lens of navigation,” explains Angus. This approach has turned out to be especially useful in analyzing conversational texts.

This kind of text analysis is used in many ways. For example, you can learn a lot that is relevant to your business by listening to how people talk about your brand. Text analytics allows you to isolate your brand name as a concept, then using a highly statistical approach to see what concepts are associated with it. Through co-locating concepts and words, you can quickly measure the kinds of words people are using in close proximity to your brand. “This not only provides insight into what people say about your brand, but it might reveal deeper insights. It could be a first step in exposing why people are thinking it,” explains Angus. Reasons “why” can have deep and not-so-obvious roots that may be different for different population segments.

“ This not only provides insight into what people think about your brand, but it can go a lot deeper. It can expose why people are thinking it. ”

KEY LESSONS

- 1 The real cost of text analysis is not in the analytical tools but in cleaning the data.
- 2 Analyzing unfiltered text frees the analysis from human biases that might creep into the analytical process.

MEANINGFUL ANALYSIS DEPENDS ON CHOOSING THE RIGHT DATA

Text analysis offers a different kind of insight than that provided by structured data analysis. “When we start talking about the difference between quantitative and qualitative research, this kind of text analytics blurs the distinction because it allows a very rich qualitative insight,” says Angus. It can have broad application in understanding deep sentiments and motivations behind population segments as they relate to complex issues, including public affairs, political issues, and many kinds of business questions.

When considering text analytics, Angus stresses the importance of knowing your goal. Only when you ask yourself, “What is a really interesting question or goal for me within my organization?” are you able to decide whether text analysis is appropriate for this kind of question; which are the right analytical tools to use in approaching that problem; and, if it involves text analysis, what text should be analyzed.

Angus points out that the question of what text to analyze is important. In instances where you are analyzing what people say, it’s best to analyze complete transcripts rather than notes because notes invariably introduce the bias of the note taker, and may miss seemingly mundane, but potentially important details. However, says Angus, “This is where the big cost question comes in.” The real cost of text analysis is not in the analytical tools but in cleaning the data. There is the challenge of accurate speech-to-text conversion, formatting issues associated with online data, and print text from news services, which comes in a variety of formats. “People often underestimate the time and cost of actually collecting and cleansing the data.”

For meaningful analysis, the text you chose is critical. Analyzing unfiltered text frees the analysis from human biases that might creep into the analytical process. “It’s not about removing human judgment,” explains Angus. “It’s about revealing true, genuine patterns, and then using human judgment to interpret those findings.”

“

People often underestimate the time and cost of actually collecting and cleansing the data.

”

IMAGINATION AND FANTASY



**ALESSANDRO
ZANASI**
President,
Zanasi & Partners

Alessandro Zanasi is a security research advisor and member of two European Commission boards (ESRAB and ESRIF) working to define European funding policies in security research. Before founding Zanasi & Partners, Alessandro co-founded TEMIS S.A. (now an Expert System company) and was an analyst for the META Group (now Gartner). He has been a researcher, professor, and IBM executive in Italy, France, and the United States and a Carabinieri scientific investigations officer. Alessandro is the author of *Text Mining and Its Applications to Intelligence, CRM and Knowledge Management*.



Twitter | Website



Download the full e-book: [Text Analytics](#)

As a pioneering text analytics researcher and business consultant, Alessandro Zanasi understands marketers' need to extract knowledge from text. As a security advisor to the European Union (EU), he understands the pressures on security operators to weed through voluminous textual data to help prevent the next terror strike.

Zanasi has just one bit of advice for those who wish to tap into the science of text analytics: tap into "imagination and fantasy" to find creative new applications for text analytics. "What I see generally is that people don't think of using their minds," he says. He tells several stories to illustrate how he and his colleagues have done just that.

The first happened more than 15 years ago, when Zanasi and his colleagues at IBM Research were seeking profitable ways to apply data-mining techniques to text. They discovered that by mining patent documents, they could predict investment opportunities for their clients. L'Oreal, Caisse des Depots et Consignations, and Électricité de France S.A. all used Zanasi's patent-scouring approach to invest in moneymaking start-ups, he recalls.

“ What I see generally is that people don't think of using their minds. ”

KEY LESSONS

- 1 Tap into imagination and fantasy to find creative new applications for text analytics.
- 2 Artificial intelligence is no substitute for human creativity.



IMAGINATION AND FANTASY

In 2000, Zanasi co-founded TEMIS S.A., a company that applied a more semantics-oriented approach to text analysis. Again, his team found an unexpected use for the science. They analyzed scientific and marketing papers to accurately forecast that Colgate and Nippon Electric Company were launching competitive pushes into their clients' commercial territories. TEMIS tipped them off.

In 2007, he formed the consultancy Zanasi & Partners (Z&P), which focuses on security. Part of Z&P's work has involved scouring social media, text messaging, digital metadata, and other sources written in languages like Arabic and Pashto to identify any radical at risk of joining terror cells.

Z&P continues to reimagine text analytics. Its antiterrorism work has recently morphed into an EU-funded project aimed at using social analytics techniques after earthquakes, floods, and even bombings to locate and rescue trapped and stranded victims.

Zanasi hopes that these examples spur others to imagine novel ways for text analytics to push their business opportunities forward. It might be a propitious time to consider it, he notes: the prices for text analytics technologies may soon drop.

Zanasi considers text analytics a branch of artificial intelligence, but he has seen people misconstrue what that really implies. "Once they have the technology, they decide to use it, thinking that so-called 'artificial intelligence' can substitute for their intelligence," he states. "And what I want to say is no, not at all."

It is far better to hire imaginative, creative thinkers and rely on them to get the best out of text analytics, he counsels. "We must try to use always human intelligence in combination with our technology so that our imagination and fantasy can be extended," Zanasi concludes. "That is the idea."

“

Once they have the technology, they decide to use it, thinking that so-called “artificial intelligence” can substitute for their intelligence.

”

NEVER LOSE SIGHT OF THE ESSENTIAL BUSINESS QUESTIONS



**LUCA
TOLDO**

Ph.D., biology; M.Sc.,
animal sciences

Luca Toldo is an industrial evangelist of text analytics in pharma and the author of more than 50 scientific, peer-reviewed publications in biomedical sciences. He has contributed to books on text mining, spoken at international text analytics conferences, and is rapporteur for the European Commission in e health IT and system biology. Luca created the patented knowledge discovery algorithm to predict gene-disease models based on published literature and is a co-developer of a pharmacovigilance application of relational machine learning methods.



Website



Download the full e-book: [Text Analytics](#)

In all the years Luca Toldo has pursued text mining and text analytics, he has come to realize the number one cause of failure when using text analytics is failure to define the right question. “The business question is the key which drives everything else,” says Toldo.

Text analytics offers a powerful methodology, but the business question must be the primary focus. A common error, which often starts in IT, is that when a company acquires this new analytics tool, people ask what it can do for them and how to set up an analysis. “But the question should not be about how to do the analysis,” advises Toldo. “The essential questions should focus on what the business needs to know.” This seems like common sense, but it becomes a reason for lack of adoption. By failing to ask the right questions in the beginning, a company fails to get satisfactory answers and often concludes that text analytics is not working for them.

Asking the right questions is a process of applying structure to unstructured data. “The key to that,” says Toldo, “is identifying subjects of discourse that matter most to the business.” So, the best way to focus the analysis on important business questions is to begin by laying out key concepts that affect the business, then map language relationships to those concepts.

“ The essential questions should focus on what
the business needs to know. ”

KEY LESSONS

- 1 By failing to ask the right questions in the beginning, a company fails to get satisfactory answers and can conclude that text analytics is not working for it.
- 2 The best way to focus the analysis on important business questions is to begin by laying out key concepts that affect the business, then map language relationships to those concepts.

NEVER LOSE SIGHT OF THE ESSENTIAL BUSINESS QUESTIONS

Defining language relationships begins with different levels of semantics that have different analytical value. Then, algorithms look at language patterns, such as how frequently certain language relationships are used, changes in semantic relationships, and concurrence of semantic occurrences within a sentence, a region of the document, or across documents. "Analysis of text in this way builds knowledge networks which become the basis for further analysis," says Toldo.

Much of the analysis Toldo performs involves mining medial research papers and electronic medical records (EMRs). Although EMRs are largely structured, Toldo points out that "80 percent of the content in an electronic medical record is unstructured data." Much of the structured portion of an EMR is set up for billing purposes and therefore is unsuitable for knowledge discovery, but the unstructured portions of EMR, which are notes from physicians based on meetings with patients, are full of useful information that can be used to develop protocols for drug trials or provide insights into better treatments.

One example is a recent project that developed an ontology to retrieve useful information from scientific literature and EMRs. The project specifically developed an ontology for multiple sclerosis (MS) that involved creating a dictionary of semantic synonyms in multiple languages. This project validated the ontology by analyzing the EMRs of 624 patients with MS. The study concluded that the MS ontology provided a valid semantic framework that could be used to automatically find useful information in both scientific literature and EMRs from MS patients. The same methodology could be applied to other types of research that analyzes both published scientific research and EMRs.

There are many other ways to apply text analytics. Toldo says, "It is quite possible to build an application that is able to analyze a conversation, identify questions, analyze a giant text database, and provide answers on the fly."

“

Analysis of text in this way builds knowledge networks which become the basis for further analysis.

”



**STEVE
GARDNER**

CEO,
RowAnalytics Ltd.

Dr. Steve Gardner has designed, built, and brought to market many innovative and commercially successful products in life sciences and health care. Steve is a former global director of research informatics for Astra AB, responsible for integrating, managing, and analyzing all of Astra's research and development information. He consults widely on large-scale informatics and analytics projects, digital health, personalized medicine and nutrition, and biobanking of human tissue.



Twitter



Website



Blog



Download the full e-book: [Text Analytics](#)

Steve Gardner, chief executive officer of RowAnalytics, has been using text analytics to discover knowledge connections since well before *text analytics* became a cool term. He explains what RowAnalytics does this way: “We have been trying to bring a degree of structure and the ability to use information from multiple different data sources to complete our picture of the particular problem space or domain we’re operating in.” He says they do that by “constructing very large-scale knowledge graphs, basically abstracting out pieces of information from the textual data sources, combining them—often with lots of structured data—and using that as a substrate to then answer quite sophisticated questions that rely on having a very good view of all of the information that’s known in a particular domain.”

Gardner offers two bits of advice to organizations seeking to achieve business value through the use of text analytics. First, he says, “From a technology perspective, I would suggest to people that it’s very unlikely that one single tool would fulfill all of their requirements unless they have a very straightforward set of things that they want to achieve. It’s much more likely that they will find themselves using multiple tools, and wanting to adapt them a little bit to the specific domain that they’re working in.”

In addition to general text analytics tools, Gardner suggests having specialized tools for validating the data generated. “Validation tools that can often be overlooked,” he says. “You can get very, very good results from automatic analysis of text if you think very carefully about the verification steps you’re going to go through for that data.”

“I think if you want to automate, the use of text analytics validation is absolutely essential to the process.”



KEY LESSONS

- 1 Multiple tools are often needed to generate the desired results from text analytics, including tools used to validate the results of the analysis.
- 2 Text analysis has the potential to expand the personal knowledge base across the organization.

EXPANDING KNOWLEDGE WITH THE RIGHT TOOLS AND CULTURAL UNDERSTANDING

Along with the analytics, he recommends using automated quality assurance systems that apply multiple tests to the data mined from source text to verify that those pieces of information are accurate. “I think if you want to automate, the use of text analytics validation is absolutely essential to the process.”

Gardner’s second piece of advice is related to the cultural impact that the availability of knowledge, enabled by text analytics, can have on an organization. He cites a major pharmaceutical company he worked with a few years ago. The company had been stuck on a couple of projects for about 18 months. During that time, 30 people were working to solve these problems—without success. “We did one of these very large-scale knowledge graphs,” he says, “combining data from about 75 sources—patent literature, scientific literature, clinical approval, and other data. We had basically taken in as much knowledge as we could find about the area we were operating in.”

Gardner explains that his team was called to present their findings. “We outlined a particular hypothesis that we generated based on the analysis that we had done and the correlations that we had found.” During the presentation, the pharmaceutical company’s topmost scientist in the field took exception to the hypothesis. He claimed to have read all the papers and said there was nothing to support the hypothesis. “Unfortunately for him,” says Gardner, “we then proceeded to walk through the chain of evidence and demonstrate that it was right and not only was it right, but it was supported by their own internal data.”

The lesson there, says Gardner is that “From a cultural perspective, it is not just that some people find this quite threatening to their power base, but the knowledge products that can be generated using text analytics are capable of being deployed to everybody’s desk.” In this case, he suggests, “The implication was that now everybody could be as knowledgeable as somebody who has spent a career developing this personal knowledge base and being the expert in their domain. The people who tend to find that most challenging are often quite senior in an organization. It destroys the information asymmetry argument on which a lot of their power is based.”

Gardner cautions, “If you’re planning to do this in a larger organization, it’s absolutely critical to take the company with you and to portray the systems that you’re building in their right light.”

“

The knowledge products that can be generated using text analytics are capable of being deployed to everybody's desk.

”



**ALEXY
KHRABROV**
Chief Scientist,
Nitro

Professor Alexy Khrabrov is a computer scientist working at the intersection of start-ups, big data, and functional programming. He is the first chief scientist at Nitro, bringing data mining, natural language processing, Scala, and Apache Spark together to drive the vision of smart documents. Alexy's previous roles include director of analytics at Clinkle; co-founder of Versal; senior research engineer at Amazon and Klout; and research scientist at Thayer School of Engineering, Dartmouth, and the NEC Research Institute in Princeton.



Download the full e-book: [Text Analytics](#)

We are approaching the point where both individuals and corporations can live entirely digital lives, says Alexy Khrabrov, chief scientist for Nitro, but we are not yet taking advantage of the opportunity. Khrabrov and his team of engineers, linguists, and mathematicians—even a few philosophers—want to change that by building reactive, document-centric workflows. Nitro's immediate aim is to help users collect documents and extract meaning from them. Then, in effect, they can stir meaning back in. “We almost want the document to have an application programming interface,” he says. “You should be able to query the document and ask it various questions. It should be enriched with all these data that we can extract and place alongside the document.”

He cites a nondisclosure agreement (NDA) is an example of what he means. As the founder of a start-up, Khrabrov has signed various NDAs with investors. These sophisticated legal documents can easily be misunderstood. A start-up might inadvertently grant the funder rights to share all its secrets while being sworn to complete secrecy, Khrabrov states.

“You have to work very hard to craft specific, very well-personalized messaging. You have to collect much more data to make these messages.”

KEY LESSONS

- 1 It is almost possible for individuals and companies to live fully digital lives.
- 2 Genuine predictive modeling through text analytics has finally become a reality.



AT THE FOREFRONT

A document, enhanced by artificial intelligence, should itself step in and help, Khrabrov asserts. “The NDA itself should automatically alert you and offer you a choice to make it a two-way agreement,” Khrabrov says. “You avoid the lawyer, improve your legal position, and get assurance that the system is watching out for you.”

Obviously, text analytics’ implications go well beyond businesses attaining better marketing insights into customers. As his NDA example suggests, it could—he insists it should—work both ways and someday will. Both consumers and producers will become adept at automatically analyzing text and both craft and filter messaging based on the analytics, he says. Text mining, text analytics, and natural language processing all are core to where he thinks things are headed.

Asked to suggest ways for businesses to gain value from text analytics, Khrabrov says:

- **Collect as much data as possible.** When it comes to data, it is important not simply to theorize, he says. You must collect, synthesize, and constantly evaluate data. “You have to work very hard to craft specific, very well-personalized messaging,” Khrabrov says. “You have to collect much more data to make these messages.”
- **Instrument everything.** Equipping yourself with instrumentation that measures and records data is a key step, according to Khrabrov. “Everything happening in your business, anything users are doing, you have to measure,” he offers. “You have to instrument user behavior so that it produces data—how long they stayed on the page, what have they looked at, where was their focus? If you can do that, you can model the user.”
- **Model the user.** Unless you can model users effectively, Khrabrov says, you will be hamstrung. He cites Amazon, which continually models its visitors over time to assess their interests and intent. If it did not, Amazon would fail to grasp when a visitor searches for a product but uses the wrong words to describe it. It would thus be unable to offer valid recommendations and would lose the sales opportunity—if not the customer. “That’s the crux of the matter,” Khrabrov says. “You should try to model what they are looking for, predictively find it for them, and offer it.”

At any other point in history, Khrabrov says, genuine predictive modeling through text analytics would have been impossible. “It is possible now,” he declares. “If you know how to do analytics of this kind, you are in the forefront.”

“

If you know how to do analytics of this kind, you are in the forefront.

”

Find the insight in the text

M C R U E B N R N Q S L Y G O
R P E M A V J L M Z O Y X O M
E Y D G Q Q I O Q Q F M R K L
M D A H Z D B T Q P A A P H S
O J E S U X P Y C V B Q R N S
T I R U M Y T H G I S N I U O
S O E S N B I F I G D L V P G
U C G C E D I S C O V E R Y N
C Z D I W N P A A D A D R L A
Z H E T N W T L F F J J X P Y
I C L Y U W Q I K B M Y C B O
K N W L H X K Q M W S S W K D
M Z O A O L S U X E F T E X T

INSIGHT ✓

TEXT

CUSTOMER

ANALYTICS

PREDICTIVE

DISCOVERY

SENTIMENT

ANGOSS

KNOWLEDGEREADER

KnowledgeREADER is an industry-first application that brings a new age of integrated customer intelligence by combining visual text discovery and sentiment analysis with the power of predictive analytics.

[Learn More About KnowledgeREADER](#)

