Mighty Guides

Text Analytics

28 Experts Share How to Achieve Business Value

Angoss

TABLE OF CONTENTS

FOREWORD

The increasing pressure to remain competitive and deliver revenue growth has channeled companies to focus on ways to better understand their customer interactions. In an age of big data, however, companies are finding it excruciatingly difficult to extract meaning from unstructured data. The increasing rate of data accumulation is making this task even more complex and overwhelming.

So, how do companies extract meaningful insights from incoming customer data and process the findings along the exploratory trajectory to turn insights into action?

Successful companies today both listen to and understand what customers are saying and are taking action in response to customer feedback, incorporating the voice of the customer (VOC) into business strategies for sales, marketing, and customer service through text analytics.

To further derive business value from text analysis and transition customer insights into action, companies are opting for applications that provide more exploratory power. As companies demand greater control of their customer intelligence decisions, a new age of *integrated* customer intelligence has been born. This evolution introduces text analytics solutions that can combine visual text discovery and sentiment analysis with the power of predictive analytics, thereby enhancing predictive models with the output of text analysis for greater exploratory power and return on investment.

This e-book was created to demonstrate the value of text analytics to a vast array of companies, customer intelligence professionals, and marketers. We hope that the first-hand text analytics experiences and best practices chronicled in the book can help companies build tactical strategies around customer-related programs and deliver unparalleled customer intelligence and VOC insights to support customer experience management, regardless of where these companies are in their text analysis journey.

Rick Makos
CEO
Angoss

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

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.

Angoss

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

Angoss
Predict. Act. Perform.

BETTER THAN YOU WERE

META S. BROWN
President,
A4A Brown, Inc.

Meta Brown's work helps organizations build analytics competence and improve communication between technical and nontechnical professionals. She is author of *Data Mining for Dummies*. You can find links for many articles on analytics and communication on her website, metabrown.com.

Twitter | Website | Blog

For Meta Brown, the definition of text analytics is simple. "It means converting text into some simpler, more conventional data structure—period," she says. The difference between that and all the other definitions she hears is that the others start there, but then plug in additional layers of meaning, which she thinks really just refer to conventional analytics.

Her guiding principle when it comes text analytics is similarly straightforward: "You will not make any money on text analytics if you do not start with a plan to do that." With that in mind, she offers three pieces of advice:

- **Present a convincing business case.** Identify a case for text analytics that solves a business problem and saves money. Executives often do not want to hear a business case built on revenue increases because they do not believe it will happen, Brown says. They prefer cost savings. Perhaps they have been burned with false promises based on past technology investments and process improvements, she suggests. "Revenue is not the strongest way to make the case that justifies someone agreeing to take on the costs that go into your plan," she adds.
- **Address a specific problem.** Look for a small payoff initially. Some retailers that see a lot of customer service inquiries still deal with those inquiries manually, which creates labor costs, potential opportunity costs, and possible customer satisfaction issues. Instead, she suggests, "We could use text analytics to direct those inquiries, sort them, and deal with them more quickly than through the process we have now." That would result in cost savings and resolve a simple problem, in turn encouraging the organization to wade deeper into text analytics. "Each tiny success is worth something," Brown says.

KEY LESSONS

1. **Executives tend to prefer the cost-savings case to the revenue-generating one.**
2. **Text analytics can be core to a business—but it is no panacea.**

> "You will not make any money on text analytics if you do not start with a plan to do that."

BETTER THAN YOU WERE

- **Follow through.** Say that you have proposed investing in text analytics to log and categorize open-ended, written Web survey responses. Do not deviate from your plan. Make sure that any required employee training is completed. Then, use the automated tools. Brown is stunned at how often software subscriptions lapse because companies never get around to using them. Instituting a new system need not require replacing your current processes—at least not immediately—but ultimately it should be out with the old and in with the new. "It's not enough to buy text analytics software," Brown says, "you've got to learn it, use it, and use the results to guide the business. No change, no benefits."

Brown has seen first hand how badly her advice is needed. For example, she spent her early days as a trainer, teaching survey research and data analytics to mid-career professionals. In that training, she always asked if their businesses conducted surveys that used open-ended written responses. If they did, she asked which companies sent the responses out to be coded, which merely glance at them, and which ignore them. "In every single group, multiple people admitted that they did nothing with the text," she says. "They were spending money to collect data that they weren't using." That is a waste of resources and customer goodwill, she states.

In contrast, approaching text analytics as a fail-safe answer would also be a mistake. Businesses sometimes abandon text analytics because of that misunderstanding. In one case, Brown notes, an organization uncovered a single mistake when an automatic categorization turned a positive term into a negative one. After that, the company rejected text analytics altogether.

Text analytics can become a core part of your business process, but make peace with its imperfections, she says. "You are not going to be perfect—nothing else in life is perfect. The objective is to be better than you were before."

> "In every single group, multiple people admitted that they did nothing with the text. They were spending money to collect data that they weren't using."

STARTING SMALL CAN BE THE KEY TO SUCCESS WITH TEXT ANALYTICS

THOMAS KUNJAPPU

Senior Product Manager, Twitter

Thomas Kunjappu loves working on novel products that use data to solve problems. He is currently working on corporate tools @ Twitter, where he imagines a more productive workplace enabled by software. He graduated from Williams College and lives in the San Francisco Bay area.

Twitter | Blog

Businesses should have clear goals in mind when they are considering text analytics technology, according to Thomas Kunjappu. Rather than getting drawn in by the hype cycle that surrounds text analytics tools, it's essential that business leaders have clear success metrics in place so that, prior to any engagement or initiative, they know what the end state or desired outcome should be.

It's also important to be organizationally ready for text analytics and to know what you're getting into when you're talking about this type of project, he believes. This includes questions of talent resources, industry-specific skill sets, and linguistic approaches as well as software requirements.

In terms of business goals, simply aiming to discover and understand sentiment are not ends in themselves. "We need to go one step further and get to the insights. It's better to start with a smaller area of focus, and then widen your scope," advises Kunjappu. "Then, you'll start to build momentum for the technology and its uses within the company, and you will be confident of the benefits of using the technology."

KEY LESSONS

1. **When considering a text analytics initiative, it's essential that you define clear, focused business goals.**
2. **Starting small can be a smart way to demonstrate the value of text analytics and build momentum for future endeavors within the company.**

Kunjappu once worked with a telecommunications client that began with a broad imperative: to know what people were saying about the company and to acquire the necessary software to achieve that one goal. This approach met with mixed results, he says.

> "It's better to start with a smaller area of focus, and then widen your scope."

STARTING SMALL CAN BE THE KEY TO SUCCESS WITH TEXT ANALYTICS

"They came up with their ontology and hundreds of topics in a hierarchical format that the central team thought was relevant and would provide something useful for the rest of the organization. There were lots of pretty reports, but the decision makers and key stakeholders didn't think that they answered any of the questions that were top of mind for them." Unfortunately, when those stakeholders voiced the questions that they felt were critical to success, the central team wasn't ready to handle them.

So, the team adjusted its approach, picking one unit within the company and solving a problem for it. This unit had a specific question about customer churn involving a particular region and product set. The team captured the stakeholders' hypotheses about why churn was happening, then trained the staff to put in the linguistic rules to accurately build up topics for analysis. Because it had a defined goal that a clear business question facilitated, the organization was able to achieve the results it was looking for.

This development completely changed the conversation about text analytics within the company, Kunjappu says. "The business stakeholders were finding value in what they were doing, and they wanted to invest more in that team. Word got around, and other units were clamoring to be next." They then performed different analyses for other types of units. "The really interesting insights come when you marry text analytics with quantitative data, operational data, financial data, and—in this case—network data because it was a telecommunications company," Kunjappu observes. The firm was able to combine those varied data sets in increasingly sophisticated ways, leading to far more innovative insights than were originally possible.

Kunjappu notes that building a global ontology on customer review data is useful for describing what is happening but it often leads to obvious answers that leave the business thinking, "Tell me something I didn't know." Talk to stakeholders early to understand their and hypotheses, they use text analytics to prove or disprove them. In this way, you increase the engagement level and the importance of the project immensely.

> "Really interesting insights come when you marry text analytics with quantitative data, operational data, financial data, and—in this case—network data."

Businesses can reap bountiful rewards from text analytics initiatives but only when they approach both the strategy and the technology that supports it in a careful, considered manner. By defining clear business goals and accurately identifying the resources needed to achieve them, it is possible to gain such valuable insight. For companies new to the process, starting with a small, focused initiative can often be the best approach.

WHERE DO YOUR HIGHEST-VALUE CUSTOMER ENGAGEMENTS OCCUR?

BILL SHELDON

Chief Solutions Officer, Angoss

Chief solutions officer for Angoss' predictive analytics professional services division, Bill Sheldon has more than 20 years of experience in sales and marketing effectiveness. Prior to joining Angoss, Bill was vice president of Marketing Solutions for Paradyz Analytics, a multichannel marketing services agency, and chief operating officer for CFM Partners, a leading provider of Software as a Service compliance risk-management solutions.

Website

Bill Sheldon's view on the business value of text analytics is shaped by his role as a business-to-business (B2B) provider of predictive analytics to the financial services industry. "We work with a lot of asset management companies, companies that sell mutual funds through a network of financial advisers," Sheldon explains.

Mutual fund companies have internal salespeople who call on financial advisers to promote different funds that achieve different financial objectives. Some companies use an analytics model to tell them which advisers are most likely to purchase and which products they will buy and to identify upsell and cross-sell opportunities. "When we engage with customers," says Sheldon, "we look specifically at their challenges, which are mostly in sales and marketing." The problem you need to solve points to the kind of data you need to solve it, and text is a valuable part of that data set. But it is just one of many data sources. "Text is not typically stand-alone data," says Sheldon. "It's an important companion to other structured business data."

KEY LESSONS

1. **The problem you need to solve points to the kind of data you need to solve it, and text is a valuable part of that data set.**
2. **Increasingly, companies are recording their customer conversations and using speech-to-text tools to convert them for text analysis.**

Text analytics is especially valuable in discovering, issues, opportunities, and trends. For instance, you can look at customer service notes to see who's complaining about what. Or you can look at customer relationship management (CRM) activity that salespeople enter after sales conversations and emails between salespeople and their customers. "When we look at email, we look at sales and customer emails separately," says Sheldon. By looking at emails in this way, you can remove the sales bias and get a better sense of the true customer voice. Text analysis can also serve as a broader market-sensing tool to reveal trends, such as a growing interest in certain kinds of investment products.

> "Text is not typically stand-alone data. It is an important companion to other structured business data."

WHERE DO YOUR HIGHEST-VALUE CUSTOMER ENGAGEMENTS OCCUR?

"These kinds of analyses can be brought back to business performance, and then we can look at this in several different dimensions," says Sheldon. For example, from a product dimension, you can look across all those interactions and see how people are feeling about specific products. Or you can look at sentiment in the sales dimension. This could mean looking at all the conversations of one salesperson to see which phrases that person uses and perhaps do a comparative analysis of different salespeople and tease out differences in what they are saying that might give insight into differences in their performance. Sheldon says, "Text analytics used in this way becomes a valuable sales coaching tool."

Many companies, particularly in the B2B segment, spend an extraordinary amount of money putting information such as meeting notes into their CRM systems, but they spend little money extracting insights from those data. Mostly, they use call meeting notes as memory refreshers before their next engagement, but Sheldon points out, "Those meeting notes contain a lot of valuable business insight."

Many companies don't know the value of their text, or if they do, they don't know which text they should analyze. "When companies are looking to extract value from their text data," says Sheldon, "they should begin by looking at where the highest-value customer engagements occur."

Increasingly, companies are recording their customer conversations and using speech-to-text tools to convert them for text analysis. This practice is especially important in B2B businesses like financial services, where highly compensated, skilled salespeople speak with highly skilled investment advisers. "These are high-value conversations that can give you a lot of insight into all the dimensions of that engagement," says Sheldon. "In text analytics, the greater the volume of text you have to work with, the more insight you can derive from it."

> "When companies are looking to extract value from their text data, they should begin by looking at where the highest-value customer engagements occur."

AVOID DRAWING INCORRECT CONCLUSIONS

JEFF CATLIN
CEO,
Lexalytics, Inc.

Jeff Catlin has more than 20 years of experience in search, classification, and text analytics products and services and has held technical, managerial, and senior management positions in such companies as Thomson Financial and Sovereign Hill Software. Prior to forming Lexalytics, Jeff was general manager of LightSpeed Software, where he was responsible for sales, marketing, and development efforts for the Knowledge Appliance and iFocus products. He holds a bachelor's degree in electrical engineering from the University of Massachusetts Amherst.

Twitter | Website | Blog

The value of text analytics is not in the *what* but in the *why.* "You can dig out facts, but if you don't go deep enough, the facts won't tell you much," says Jeff Catlin, whose business specializes in analyzing various kinds of social content. In fact, you are likely to draw incorrect conclusions if you do not find the deeper context behind what you are seeing in a text analysis.

A good example comes from a project in which a hotel chain wanted to examine customer feedback to see how two Las Vegas hotels—Bally's and Bellagio—fared with their customers. The project involved applying text analysis to the customer feedback, with particular attention to sentiment about the hotels in general, and specific features like the rooms, the pool, valet service, and location. Such an application is typical of text analytics on social media. An interesting and unexpected finding was that consumer sentiment around Bellagio, a luxury venue, was showing higher negatives than sentiment around Valley's, a low-cost venue. Valley's customers seemed to be more satisfied. "We knew the data were accurate," says Catlin, "because the hotels scored the same on location. They are directly across the street from each other."

KEY LESSONS

1. **It is quite possible to draw incorrect conclusions if you do not find the deeper context behind what you are seeing in a text analysis.**
2. **You need to use text analytics tools that can understand and correlate different ways of expressing a sentiment. Only in that way is it possible to lift the *why* answers out of the top-level results.**

> "You can dig out facts, but if you don't go deep enough, the facts won't tell you much."

AVOID DRAWING INCORRECT CONCLUSIONS

A first look, at the data suggested that Bellagio had a problem. Here was a cut-rate hotel scoring higher in customer satisfaction than the luxury hotel across the street. A deeper dive into the data, however, revealed that the people complaining about Bellagio were complaining almost exclusively about the price of things at the hotel. Because Bellagio caters to the very high end of the market, they realized that negative ratings based on complaints about price were not a problem for them. "This is a good example of how the data were misleading," says Catlin. "If you were doing a superficial social media look, those negative sentiments could be worrisome," but by looking more closely at the data and viewing them in the context of what was actually driving the negative sentiment, you could see an answer to the *why* questions behind the negative ratings. In this case, it became clear that the negatives were not something Bellagio needed to worry about.

"To do this deeper analysis, you need to be able to understand what's being talked about, even if it is being talked about in seven different ways," says Catlin. You need to use text analytics tools that can understand and correlate different ways of expressing a sentiment. Only in that way is it possible to lift the *why* answers out of the top-level results.

Catlin says, "Most people who come to text mining think in terms of buckets they understand and problems they know they want to solve. In those cases, it's usually viewed as a categorization and sentiment problem." A lot of the more interesting data come out of a "what are people talking about" kind of analysis. You can't precategorize or build a taxonomy to capture what people are going to be talking about; you have to be able to extract the bits of information that are relevant and hot at that moment. Customers don't tend to think about text analysis in terms of things they hadn't expected, but that could harm them. "In my view, that is one of the best uses of this technology. It is totally data driven. You don't have prebuilt biases about what is going on. The data are just telling you what people are saying."

> "To do this deeper analysis, you need to be able to understand what's being talked about, even if it is being talked about in seven different ways."

FIRM, NOT FROZEN

SUDIP CHAKRABORTY
President,
Xypress LLC

Sudip Chakraborty founded Xypress to help companies grow profits by applying sophisticated analytics to big data. He and his team have helped companies uncover insights from structured and unstructured data in the financial services, technology, information services, and pharmaceutical industries. Sudip holds two patents in the areas of approximate text matching, text analytics, and association discovery: System and process for discovering relationships between entities based on common areas of interest and System and method for searching and matching databases.

Twitter | Website | Blog

As an expert in applying sophisticated analytics in business-to-business (B2B) settings, Sudip Chakraborty, president of Xypress LLC, offers four recommendations for getting the most out of text analytics:

- **Develop a set of hypotheses regarding how you can solve a business problem.**
- **Apply text analytics and review and refine your original set of hypotheses.**
- **Leverage humans to evaluate, refine, guide and apply – machines are not sufficient.**
- **Deploy a cross-functional team of highly trained professionals with skills in business problem solving, text analytics and structured data analytics.**

KEY LESSONS

1. **Develop a set of three to five hypotheses for text analytics, then be prepared to iterate.**
2. **The real business value of text analytics comes from analyzing both structured and unstructured data.**

Develop a set of hypotheses. As a first step, generate three to five hypotheses about how text analytics might help your business to solve one or a set of business problems. The challenge, he states, is that there is no end to the possibilities: "Text analytics consists of many different techniques including sentence segmentation, parts of speech tagging, named entity recognition, sentiment analysis among others - you can, therefore, spend days, weeks, even months, applying text analytics to unstructured data without realizing business value from your efforts." He also notes that the analysts who understand and can solve text analytics problems often have different educational backgrounds and experience from those who understand and solve business problems. It's up to business decision makers to set direction and time limits. "Analytics professionals can spend a lot of time looking for things," Chakraborty observes. "Unless they get direction, their findings are unlikely to be of business value."

> "We actually needed business analysts to evaluate the quality of the results and to ensure that they would be valuable to our client's customers."

FIRM, NOT FROZEN

Iterate. Your theories should be firm, not frozen. In particular, if you have little prior experience, you might be far off target at first. Redefine your hypotheses periodically. Chakraborty illustrates his point with a story about how he once helped a $50 billion tech company explore the wearables market. The client directed its own considerable resources to interviews with experts and focus groups. Chakraborty's team, meanwhile, performed social media analytics to unearth trends in wearable technology. While performing this analysis, the team uncovered a problem no expert had yet spotted: competitors' wearables caused skin irritation for some customers. However, this was not part of the original hypotheses. Chakraborty's team worked with the client closely to refine the original set of hypotheses; they decided to apply a broader set of text analytics techniques to dig deeper into the skin irritation issue. The client, therefore, had an opportunity to reassess its device before it ever hit shelves.

Software cannot do it all. Human assessment is crucial. Again, Chakraborty speaks from experience. He holds a patent on a sophisticated B2B search engine designed for a global wholesale website that performs product searches through association discovery and uncovers purchase trends. During that project, he felt compelled to engage business analysts beyond the original team of data scientists and software developers. "Just looking at the data that our text analytics threw up was not sufficient," Chakraborty says. "We actually needed business analysts to evaluate the quality of the results and to ensure that they would be valuable to our client's customers."

Deploy a cross-functional team with complementary skills. People tend to approach data analytics as an either/or proposition, he observes—it's all about structured or unstructured data. That approach risks sacrificing the real business value of text analytics, which comes from analyzing both structured and unstructured data, Chakraborty states. "In my experience I see that organizations tend to start with one knowing that they will get to the other," Chakraborty says. "It's just that it is often challenging to do both together, and they don't want to take on more than they can handle." Humans, he adds, remain indispensable. "Machines are increasingly able to perform significantly more intelligent tasks, but be prepared for humans to assess, refine, and guide the execution of your text analytics program," Chakraborty concludes. "That is the way that you will actually maximize the business value from your efforts."

> "Machines are increasingly able to perform significantly more intelligent tasks, but be prepared for humans to assess, refine, and guide the execution of your text analytics program."

MATCH THE DATA AND ANALYSIS METHODOLOGY TO THE BUSINESS CHALLENGE

SETH GRIMES

Principal Consultant,
Alta Plana Corporation

Seth Grimes specializes in strategic IT analysis, architecture, and planning, with a focus on business intelligence (BI) and text and sentiment analysis. Seth is a widely followed industry analyst. He consults via Alta Plana Corporation, organizes the Sentiment Analysis Symposium and LT-Accelerate conference, and writes frequently for a variety of publications. For a more extensive biography, visit sethgrimes.com and follow Seth on Twitter at @SethGrimes.

Twitter | Website | Blog

There are two essential challenges in seeking to discover business value via text analytics. One is choosing the right text for analysis; the other is choosing an analytical methodology that is suitable for the text and insights you are seeking.

It all begins with a business need. "This means defining broad goals but also identifying and focusing on particular elements that are key to reaching those goals," advises Seth Grimes. Determine what kind of insights will address those break-out elements, and then the text (and other data sources) that, properly analyzed, will produce those insights. "Your technical choices – data, analytical methods, user interfaces – should always be clearly linked to business goals" according to Grimes, "for text analytics and for every other analytics initiative."

"Any business process that involves text in high volumes can fruitfully use text analytics to gain insights." The challenge, however, is that unlike structured data, where the data elements have defined meaning, meaning captured in text is not predetermined. Instead, meaning derives for context and perspective. "A given piece of text can mean different things to different people in different situations," says Grimes. For example, someone who runs a hotel property would read a review of his or her business for a different reason than someone who is reading the review while looking for a place to stay. The person looking for a place to stay has his or her needs in mind. The person who is running the hotel is more interested in problems people who have stayed at the hotel have reported and what he or she can improve.

KEY LESSONS

1. **It all begins with the business need. This will give you the basis for deciding what text to analyze and what kind of analysis is appropriate.**
2. **By combining analytics from multiple data sources, it becomes possible to correlate events with behaviors and other business activities you are measuring.**

> "Different sources may provide different, divergent insights about a given topic."

MATCH THE DATA AND ANALYSIS METHODOLOGY TO THE BUSINESS CHALLENGE

Further, all text is not created equal. "Different sources may provide different, divergent insights about a given topic," says Grimes. For instance, in finance, ever since trading desks have become highly automated, people have been working to find signals that give early indications about market movements. It's common wisdom in the financial world that you can make money by trading on information from a reliable source. "But a big question in the financial world is how to analyze social media," says Grimes. Social media are full of adversarial signals such as misleading or mistaken information or deliberately false information. "There's opportunity to be discovered in social-derived insights," say Grimes, "but if you trade on those insights, you are taking a very large risk due to the information's unreliability. Don't ignore social signals; instead, understand what they're good for – unmatched early warning of breaking events and early trend indicators – and their limits."

When analyzing text, keep these basic principles in mind:

- "The changes that you see in data are much more interesting and important than the absolute values of what you are measuring," For instance, the fact that the price of a stock jumps from $80 to $90 is much more interesting than the fact that the stock is trading at $80 at a particular moment. Something happened to trigger that jump, which means that a longitudinal analysis of data can put "events" into context.
- "You can get significant analytical lift when you combine information from the variety of sources available to you." This can mean combining different kinds of text analysis with other metrics. For example, companies use omnichannel analytics to analyze dozens of inputs from many customer touch points. To analyze inputs from these channels, you need to adapt to the different ways people say things in the different media, and the fact that they are using these different media along different points in the customer journey.

By combining analytics from multiple data sources, it becomes possible to correlate events with behaviors and other business activities you are measuring. "Integrating and correlating across channels gives a more complete and useful insight into what is happening in a business."

> "You can get significant analytical lift when you combine information from the variety of sources available to you."

COMBINE STRUCTURED AND UNSTRUCTURED DATA TO IMPROVE CUSTOMER RELATIONSHIPS

ROMAN KUBIAK

Senior Consultant

Roman Kubiak holds an M.S. degree in economics from Gdansk University and an MBA from Baruch College. He has worked on analytic teams for such financial services firms as American Express and CityCards as well as for e-commerce companies 1800Flowers.com and PetCareRx.com. He moved to a telecommunications company to pursue his interests in natural language processing and text analytics. His other interests include online marketing attribution and net lift purchase propensity modeling.

Twitter | Website

For Roman Kubiak, senior consultant to the advanced analytics team at a telecommunications company, text analytics helps businesses learn more about their customers. We derive the most value, he says, by mining both structured and unstructured data. "Today, many of the interactions are long distance. People are either interacting with our website or exposed to our ads; they see something on the Web. Other than in retail stores, we don't have much of an opportunity for face-to-face interaction," he says. "Instead, we have the customer service phone transcripts, the customer support emails, and what people are saying on social media, and we use that information to learn what customers' impressions of our business are."

It's not a simple process. Kubiak says that he uses tools like topic modeling, text classification, and text tagging to merge the output from unstructured data with the output they receive from numeric databases and customer records. "In my experience, the biggest challenge is capturing and storing the text data generated by customers," he says. He recommends doing so as a good starting place. "You need to be able to keep records for extended periods to be able to analyze changes over time."

KEY LESSONS

1. **The ability to capture customer data over an extended period of time is essential to developing a successful text analytics strategy.**
2. **Preprocessing data—cleansing, tagging, classifying them—is a time-consuming element of text analytics that cannot be overlooked.**

> "We have the customer service phone transcripts, the customer support emails . . . we use that information to learn what customers' impressions of our business are."

COMBINE STRUCTURED AND UNSTRUCTURED DATA TO IMPROVE CUSTOMER RELATIONSHIPS

It's not enough to have historical records, however. Kubiak says that the next-greatest challenge when trying to derive business value from text analytics is that "records need to be stored with the time stamp or the customer identifier for use later in merging the unstructured data with structured data. Of course, in sources like social media chatter, there is no customer identifier for you to work with. In that case, the only option is to look at the time series, look at the date when certain messages are popping up, and try to align that information with changes in the business or changes in marketing."

"For instance," Kubiak says "When we introduce major changes to our marketing campaign, we want to know how customers react. To do that, we can look at the social media chatter and see what the impressions are, but where we find the most value is when we are able to combine the results of the text analytics with structured data. Potentially, we can use the topics customers are discussing with us as early indicators for shifts in customer preferences and sentiments. We are also trying to improve our customer satisfaction by listening to the voice of the customer."

Although the field of text analytics offers many potential benefits for creating deeper relationships with customers, Kubiak warns that "it's a new and sexy field, but it's not really glamorous. Whoever wants to venture into this area should be patient and ready to spend time preprocessing the data—on things like text parsing, tokenizing. This is a time-consuming and painstaking process but it may have great impact on the results of the analysis."

"They also need to understand the other aspects of preprocessing, like regular expressions, removing special characters, removing some phrases and punctuation from the text," he adds. "That takes a lot of time. Software can do some of it for you, but to become better and better, it's good to have a better understanding of how these techniques work and the cautions and benefits of some of these approaches." To achieve that better understanding, Kubiak suggests, "Text analysts should try to spend time learning what has already been done in academia and look for business applications for these techniques."

"We can use the topics customers are discussing with us as early indicators for shifts in customer preferences and sentiments."

THE VOICE OF THE CUSTOMER

KAENAN HERTZ

Executive Director, Ernst & Young

Kaenan Hertz specializes in helping clients use strategic information and analytics to drive innovation in products and services. He has extensive experience helping companies innovate their customer experience strategy; enhance offerings through digital, traditional, contact center, and third-party channels; and understand the implications of changing market and competitive dynamics on customer strategies. Kaenan has held leading roles in marketing and decision management for top insurance, banking, and finance companies, including USAA, Citibank, and Sallie Mae.

Twitter

Before taking his present job at Ernst & Young, Kaenan Hertz was a market intelligence executive at a top-tier financial services company. There, he once headed up a "voice of the customer" project that identified more than 100 points of entry for customers to offer feedback, including letters to the chief executive officer, conversations with customer service representatives, emails, surveys, and touch points across all departments and product categories. Hertz's project aimed to generate a 360-degree customer view by building the infrastructure to gather, synthesize, and analyze customer feedback into a central database. "You can then start to use that database to drive insights," he says.

Some of those insights were unexpected. Hertz's former company offered several automobile insurance products, for example. After analyzing customer feedback, the company was surprised to discover that some of its insurance feature sets were not important to customers. For Hertz, the moral of that story is that you ultimately cannot do text analytics in a vacuum. "It is a collaborative partnership among many different departments in any given company," he says.

KEY LESSONS

1. **Text analytics will ultimately require broad interdepartmental collaboration.**
2. **It would be wise to start with a small pilot project and grow your capabilities from there.**

> "They never drive value . . . because they haven't articulated the business problem they are trying to solve."

THE VOICE OF THE CUSTOMER

In that spirit, Hertz offers some advice for getting your text analytics journey underway:

- **Start with a real business problem.** Only when you know the issue you want to resolve can you find the appropriate tools to answer it, Hertz states. At Ernst & Young, he has seen companies become so enamored by new tools and technologies that they forget to connect them to any real business objectives. "So they never drive value," Hertz observes. He is personally aware, he says, of several companies that are now limping through big, unstructured text data projects. Why? "Because they haven't articulated the business problem they are trying to solve," he states.

- **Start small, and grow from there.** He calls this approach the *crawl, walk, run methodology*. Although ultimately, text analytics cannot be done in isolation, you can begin modestly. An example might be collecting open-ended feedback and distilling it into root causes for a customer's most recent transaction. That would answer a single problem using one well-defined source of information, he says. You would not need to merge data across multiple systems and different departments. Small successes like that can generate corporate enthusiasm to keep pushing forward. Conversely, he warns, "If you try to boil the ocean, you will never succeed."

- **If you collect data, use them.** If you solicit the customer's input, you need to process it and act on it, Hertz states. That is what customers expect. After all, he says, performing text analytics is really about listening to the voice of the customer. "Don't just collect information and do nothing with it," he concludes.

> "If you try to boil the ocean, you will never succeed."

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.

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.

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.**

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."

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."

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

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.

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.**

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

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.

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

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."

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

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.

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.**

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

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."

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

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."

HOW DO YOU MEAN, "DUCK"?

JON LEHTO

Unstructured Data Analytics, Constant Contact

Jon Lehto has developed tools and analysis of customers, empowering informed product management planning and decisions. These strategic tools allow rapid analysis of big data. Jon has evaluated many machine learning algorithms; is an expert on search engines, taxonomy, information extraction, email archives, image processing, and database internals; and has consulted on these technologies at large international, commercial and government defense and intelligence agencies.

Website

As an unstructured data analyst, Jon Lehto finds that clients sometimes ignore his advice because they want analysis that validates their preconceived notions. What they should do, in his opinion, is focus on what the data has to say. To that end:

- **Begin with the end in mind.** Lehto's job involves "shredding" millions of emails and survey text into actionable information by using natural language processing (NLP) and other techniques. Instead of looking for real insights, however, clients often want analysis that makes them look good. They *should* be asking how their company makes money and how can they make more, he says. Do people want to buy their products? How can the organization make email solicitations more attractive and competitive? In other words, "What are you doing to drive customer interest? That's where the text analytics comes in."
- **Do not ignore the outliers.** The 80/20 rule suggests that 80 percent of transactions come from 20 percent of customers. Your job, then, is to identify the 20 percent. "Some needles in the haystack are golden," Lehto states. You are looking for the results presented on your site 1 million times, not the average number of times viewed, Lehto says, especially if 22 the average. If you regard the extremes automatically as bad data, he warns, you might end up losing the benefit of the hard work of collecting, cleansing, and analyzing your data. Don't let it happen: "Expect to spend a lot of time with fertilizer and thorns while collecting roses."

KEY LESSONS

1. **Don't ignore outliers in the data or be lulled into complacency by smoothed-out averages.**
2. **Ad hoc, on-the-fly categorizations could be the next big leap in text analytics.**

> "What are you doing to drive customer interest? That's where the text analytics comes in."

HOW DO YOU MEAN, "DUCK"?

- **Avoid "some assembly required" thinking.** When your boss wants actionable data, text analytics should provide them. Instead, some people offer smokescreens in the form of 30- to 60-page written reports. The analyst can respond to the boss's request by simply saying the data are in the report. You can wind up with so many reports and so few answers that you practically need to run them through NLP algorithms just to squeeze out any meaning. Discontinue unused reports, he advises, "especially if no one notices."

Lehto has recently focused on taxonomy development using machine learning, His projects include a huge search index taxonomy that contains massive supporting documents for the federal Congressional Register. To truly learn, he states, machines generally require additional contextual knowledge than might be obvious at first blush. If you input the word *duck*, for example, a machine will want to know what you mean by "duck". Terms must be described and supported with sufficient data, he says. When that takes place, machine learning becomes easier.

But Lehto is already looking beyond that. He thinks that the next leap in text analytics should not focus on machine learning. In effect, it should focus on machines' ability to forget. To him, flexible, ad hoc categorization is the "holy grail." The key to that concept—the "ad hoc" part—is that categorizations should be done on the fly, and the underlying text analytics should discard old categories as they become irrelevant.

That capability has not arrived quite yet, but Lehto is hopeful. "I want that categorization to go away quickly so I can make another one as I move along in my learning," he states. "Tomorrow, I won't care because I already answered that question."

> "Expect to spend a lot of time with fertilizer and thorns while collecting roses."

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

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."

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.**

> "Where the serious conversations happen is in some of the back-end communities."

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."

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

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.

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

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.

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.**

> "The essential questions should focus on what the business needs to know."

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."

Text Analytics Case Study

"I would recommend KnowledgeREADER™ to others...
With Angoss, it's more of a partnership than a sell."

Find out how this iconic brand is using Angoss and text analytics to better understand their customers

Download the case study now!

ROBUST TEXT ANALYTICS RESULTS REQUIRE A HUMAN ELEMENT

HELEN CLEGG

Social Analytics Manager, A.T. Kearney

As social analytics manager at A.T. Kearney, Helen Clegg uses her library science background and expertise in taxonomy development to add value to the text analytics capabilities of the company's Data Science Team. She is a regular presenter at Taxonomy Boot Camp and Text Analytics World and a guest lecturer at the Grenoble Graduate School of Management and the London School of Business and Finance. Helen is an organic gardener and studying for her Permaculture Design Certificate.

Twitter | Website

Helen Clegg, social analytics manager at A.T. Kearney, had no idea when she chose to pursue a degree in library science that she would eventually be using that knowledge to improve text analytics results. That is exactly what happened, however, when she began working with text analytics at A.T. Kearney, and now she has a unique view of how to gain the greatest business value from text analytics efforts.

"I think companies are struggling to get value from text analytics for a number of reasons," she says. "First of all, it's not easy to translate a text analytics concept into business terms. It's really unfamiliar territory, and if you think about it, the language that you use when you talk about text analytics is unfamiliar language—terms like *semantics, lexical chains, algorithms*. When you start talking about these concepts and use these terms, people don't understand it. It's unfamiliar, and so they don't know how to get business value from it."

It's doesn't have to be overwhelming, however. Clegg suggests, "I would recommend going for a project-oriented approach. You pick a project or a use case, if you will. You start small, and you work out what the business problem is. What is the problem that the project team is trying to answer?"

> "I would recommend to go for a project-oriented approach . . . . You start small, and you work out what the business problem is."

KEY LESSONS

1. **A good approach is to take one small element of a larger project and work toward finding the right solution for that element, expanding text analytics efforts over time.**
2. **Machines don't have common sense. They don't understand subtly and nuances. For this reason, the human element is still one of the most important elements of text analytics.**

ROBUST TEXT ANALYTICS RESULTS REQUIRE A HUMAN ELEMENT

"By doing that, it makes the concept of text analytics much more tangible and demonstrates, if you're successful, the speed and just how quickly text analytics as an application can help solve that problem," she says. Clegg also says that this type of approach makes it easier to determine which text analytics tools to use. "This is important," she says, "because there are lots of tools in the marketplace. We started with a couple of open source tools, and now we've added a couple of licensed products, so we have a mix. It's only by taking this approach that it's allowed us to work out the tool set that we need."

Although Clegg finds value in analytics tools, she warns that for truly robust results, text analytics must also have a human element. "Don't discard the importance of the human element. There are many proponents out there who say use a machine-only approach. They let the big data tools and the algorithms do everything, but we've found that our results are even better and they're more robust if we leverage human intelligence during the process. We're leveraging classical library science skills and that's where I fit in. I have the library science qualification, and that skill is really useful for designing a custom taxonomy for whatever project or use case we're working on."

Clegg points to an example that was used as an illustrative exercise for a project team. "This particular team wanted to know what was top of mind for a particular group of consumers when buying a specific brand of car," she explains. "This required mining a vast and complex array of social media posts. Leveraging human intelligence, we developed a custom automotive taxonomy based purely on what the project wanted to find out. Once we had the custom taxonomy worked out, we were able to mine these social posts with this taxonomy. That enabled us to gain rapid insight into consumers' buying behavior."

> "Sentiment analysis doesn't always provide the answer to the business problem. Saying that something is negative or positive doesn't really give you deep insight."

One last bit of advice from Clegg, "I would also say don't get hung up on sentiment analysis. There are quite a number of sophisticated sentiment engines on the market but what we would say from the work we've been doing is that sentiment analysis doesn't always provide the answer to the business problem. Saying that something is negative or positive doesn't really give you deep insight," she says. "That's why we have developed more of a contextual approach where we leverage human intelligence around the taxonomy and combine it with text analytics on the machine side."

EXPANDING KNOWLEDGE WITH THE RIGHT TOOLS AND CULTURAL UNDERSTANDING

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

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."

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.**

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."

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

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."

IN TEXT ANALYSIS, SCALE MATTERS

GERSHON BIALER

Sr. Data Science Engineer, CrunchBase

Gershon Bialer is the senior data science engineer at CrunchBase, where he spearheads scalable efforts to integrate big data into CrunchBase and ensure high data quality. He has been programming since attending National Computer Camp in elementary school and has extensive industry experience applying machine learning and natural language processing.

Twitter | Website | Blog

As an aggregator of news and information serving the business start-up ecosystem, CrunchBase analyzes large amounts of text every day from many different sources. For many businesses, text analytics is research related or designed to answer essential business questions. CrunchBase has an established model for finding certain kinds of relevant information, extracting it, and preparing it for publication on its website. Gershon Bialer explains, "It varies depending on the day, but we process between 10,000 and 30,000 articles each day."

When processing text, one of the first questions Bialer asks is how much text the analysis considers. It could be anything from one document to the entire Internet. If you are analyzing one document, a person could do that. "If, on the other hand, you're analyzing the whole Internet, you're going to need a highly automated solution to process all those data," says Bialer. The size of the analytical task is one factor that dictates the analytical methods and technologies you use.

KEY LESSONS

1. **The size of the analytical task is one factor that dictates the analytical methods and technologies you use.**
2. **After you have a sense of the size of your data set, you must develop a baseline understanding of what you can reasonably extract from it.**

> "If, on the other hand, you're analyzing the whole Internet, you're going to need a highly automated solution to process all those data."

IN TEXT ANALYSIS, SCALE MATTERS

After you have a sense of the size of your data set, you must develop a baseline understanding of what you can reasonably extract from it. Not all text is the same. If you have a structured piece of text, such as a corporate filing that follows a specific format, it will be a lot simpler to analyze that than something that is freeform, such as an article. Even within more freeform content, however, the nature of the text can vary. "We look at a lot of news articles that are reasonably well written, and people generally use proper words for the names of things," says Bialer. But for content in forums, blogs, and social media and for short-form text such as Twitter, the task of identifying meaningful entities becomes more difficult. "One simple way to establish a baseline metric is to look at the text and ask what could a human reasonably tell from it," advises Bialer. If the text obviously contains relevant elements, that is the beginning of defining searchable entities. If the text is totally impenetrable, it may require machine learning algorithms to define meaningful elements.

> "One simple way to establish a baseline metric is to look at the text and ask what could a human reasonably tell from it."

With the text data set defined and a baseline expectation of what you can extract from it established, it becomes a process of defining entities and concepts the algorithms use to identify relevant information. If you are using a machine learning strategy, these entities become the basis for training the algorithm in a supervised learning approach.

"One of the big challenges in analyzing articles," says Bialer, "is distinguishing between article content and imbedded advertising." A common way to weed out advertising is to look for entities that have no relationship to most of the other entities in the article. So for example, if it is an article about a pharmaceutical product, and suddenly there are a few references to Ford pickup trucks, that's likely spurious information from an advertisement. "But it's not always so easy," says Bialer. "They can be pretty sneaky about inserting context-related advertising into the middle of an article, and that can be hard to detect."

Text analytics is about a lot more than the algorithms and the tagging. "As interesting as that is, a lot of what I do is actually just dealing with scalability issues," says Bialer. You can build a great algorithm, but then you have to make it work reliably at scale. And then of course you must extract the results into some kind of useful structure.

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

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.

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.**

> "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."

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."

PROACTIVE, HOLISTIC PLANNING IS KEY TO MAXIMIZING BUSINESS INSIGHT

TIM KUHNS

Principal, Innovation Research and Analytics Evangelist, ACT, Inc.

Tim Kuhns is responsible for corporate initiatives sponsored by ACT's Office for Innovation. His primary focus is to advance data literacy at ACT by developing analytics capabilities within communities of practice, which include a cross-section of staff who practice applying quantitative data analytics, text analytics, and multichannel analytics. ACT realizes the advances in data literacy through research, training, experimentation, and prototype development.

An analytics evangelist, Tim Kuhns advises that every text analytics initiative must begin with a holistic plan that speaks to the needs of the entire business. In his role as principal, innovation research and analytics evangelist at ACT, Kuhns observes that organizations that don't have a clear plan for their analytics can often end up with *siloes*, or separate spaces, not working in a coordinated fashion that could maximize overall insight to the entire company. Fortunately, however, there's a way to address that challenge.

Says Kuhns, it begins with "really getting out into the organization and doing intraorganizational research to find out who it is that has an interest in this space or that has a need or a problem to solve and is working there." Why is this important? "Because," explains Kuhns, "working from this holistic view, in which you bring people together from across the organization, there is a certain amount of commonality. You'll find there are a certain number of tools and platforms that are applicable in one area that are also applicable to others. It's also a way to gain efficiency, to save costs, to share experiences, which collectively advance the literacy and the knowledge and the capability across the organization to achieve success."

KEY LESSONS

1. **Every business or organization needs a proactive, holistic plan for engaging in text analytics.**
2. **Organizations without a clear plan for their analytics can end up with silos and not working in a coordinated fashion that could maximize overall insight to the entire company.**

> "Working from this holistic view, in which you bring people together from across the organization, there is a certain amount of commonality."

PROACTIVE, HOLISTIC PLANNING IS KEY TO MAXIMIZING BUSINESS INSIGHT

Last year, Kuhns led an initiative at ACT to move beyond analyzing structured data—neat columns and rows and databases—to focus on the growing realm of unstructured text. The task force he created was able to build a common understanding within ACT of what unstructured text analysis really meant over four main areas of importance in the organization. From there, they delved into each area and documented their existing practices in the 17 main application areas that they discovered for text analytics. They also took care to assess the current state of their analytics capability, regardless of whether it was as robust as it could be and whether there was more capacity-building work to be done.

Kuhns took away several key learning points from his initiative. "If you're really going to thrive in this area," he says, "you're going to need to find what is unique to your business, what someone else may not need to deal with in this space. We're working very diligently to find ways to make sense out of the areas that are unique to our business." He also notes that although he often hears of an emphasis on social listening platforms at other organizations and believes that they are indeed important for keeping a finger on the pulse on the business, his experience has shown him that those platforms are just one part of the total analytics picture and, as such, only form a portion of the rich insight to be tapped.

Ultimately, a proactive approach toward soliciting feedback is essential for achieving analytics goals. If you are looking to uncover the best, most actionable insight that you possibly can, "Get out there and ignite the conversations," Kuhns stresses. "A lot of times, unless you get people talking, they won't talk." You may be surprised at what you learn by inviting your stakeholders to share their experiences with you in an open-ended fashion. By comparing the data that you receive with honest, direct feedback about their expectations surrounding their customer experience, you can point the way forward to true business value.

> "You're going to need to find what is unique to your business, what someone else may not need to deal with in this space."

MANAGE EXPECTATIONS

PETR KNOTH

Senior Data Scientist, Mendeley Ltd.

Dr. Petr Knoth is a senior data scientist at Mendeley, where he develops text-mining tools to help researchers' workflows. Petr is also the founder of the CORE system, which aggregates millions of open access publications from repositories and journals and makes them freely available for text mining. Previously, as a researcher at the Open University, Petr acted as the principle investigator on many national and international research projects in the areas of text mining and open science.

Twitter | Website

An important step that motivated text analytics of research papers was made in 1986 when Swanson argued that knowledge that can constitute scientific discoveries can be already publicly available, but hidden in large volumes of text. Two years later, this idea was used to reveal the relationship between magnesium deficiency and migraine. Since then it has been demonstrated that these hidden connections can be discovered using text mining techniques more easily and systematically.

Petr Knoth is a senior data scientist at Mendeley, a subsidiary of the Reed Elsevier publishing company, part of the RELX Group. He calls the migraine discovery one of the "holy grails" of text mining. "Everybody would like to achieve this," he states.

Knoth has dedicated his career to helping researchers. "I saw that there was so much that could be done with the data," Knoth says. "I have really wanted to help researchers improve their workflows."

Toward that end, as a research fellow at the United Kingdom's Open University, he developed COnnecting Repositories (CORE), a meta-repository that aggregates and enables the text-mining of more than 25 million open access research documents from 700 databases worldwide. Recently, he has lead a team developing a new infrastructure that will allow CORE text miners to take advantage of cloud-based algorithms.

> "Text mining has so many application domains, it is absolutely incredible."

KEY LESSONS

1. **Develop a framework, and identify how you will obtain data before you set out on the text analytics path.**
2. **Communicate clearly, and manage expectations realistically.**

MANAGE EXPECTATIONS

The promise of text analytics is not limited to Knoth's core constituency of academics and medical researchers, of course. He sees implications for brand-reputation management, market research, competitive intelligence, as well as other areas. "Text mining has so many application domains, it is absolutely incredible," Knoth says. "That is because the portion of unstructured textual data on the Web is enormous."

Before you can derive business value from text analytics, however, you must lay the groundwork. With that in mind, Knoth offers the following advice:

- **Develop an evaluation framework.** When faced with a text analytics problem, the first imperative is defining your goals and identifying the specific metrics you will use to measure success. "Unless you define what this value actually is—the metric that measures this value—you are unlikely to succeed," Knoth warns. "It is not even guaranteed that you will find anything."
- **Identify how you will obtain the evaluation data.** Imagine that you have millions of documents that contain citations that reference other documents. You want to isolate and extract just those citations. How will you know how well is your method doing at scale? If you can't answer that question cogently, Knoth warns, do not proceed. "You need to have some sort of gold-standard data set that already contains these references," he says. Otherwise, you will find it difficult to demonstrate any value in your framework.
- **Communicate clearly and realistically.** People who solve text analytics problems should be both solid data scientists and good communicators, Knoth advises. They should be able to state clearly both how they will execute their plans and how they will manage situations if things go off track.

> "You need to have some sort of gold-standard data set that already contains these references."

Text analytics is increasingly popular, Knoth notes. Nonetheless, it is not always easy to explain to investors, customers, and management why they should invest in it. Clearly, it has reached a state where it can be put into production and offer useful applications, but it has its limits. Don't oversell it, Knoth warns.

"The point is that you need to manage these expectations well," he states. "When you are talking to senior management or to a customer, you need to understand that this person doesn't necessarily know what is realistic."

THERE'S MORE TO TEXT ANALYTICS THAN DATA

JAMES J. NOLAN

Vice President, Analytic Technologies, Decisive Analytics Corporation

James J. Nolan is the vice president for analytic technologies at Decisive Analytics Corporation (DAC) in Arlington, Virginia, where he leads a world-class team of mathematicians, computer scientists, and engineers in the development of patented text, video, audio, and data-fusion capabilities that DAC deploys worldwide in military, intelligence, and commercial systems. James holds Ph.D. and M.S. degrees from George Mason University and a B.S. degree from the University of Maryland.

Dr. James J. Nolan, vice president for analytic technologies at Decisive Analytics Corporation, works with clients challenged with turning volumes of data into actionable information. There is no easy solution, he says. "A client may be convinced that there's some kind of shrink-wrapped solution that they can buy, install, and run to solve their problem. What we've found is that you really need some expertise to understand how you can apply an existing text analytics tool set to your workflow."

However, Nolan points out that it's not enough to have a great tool set. Organizations, he says, "also need to dedicate some internal resources to working with a text analytics vendor to transfer that subject matter expert knowledge into the system so they can get a solution that's going to 100 percent meet their problem."

In one commercial example, where a customer wanted to identify events, relationships, and topics in the news that could help it get ahead of the news cycle from a sales perspective, he says, "they wanted to be able to broaden their coverage so that they could look at more news sources, or they wanted to be able to do it faster." To accomplish these goals, Nolan says, a combination of off-the-shelf tools and internal knowledge was necessary.

KEY LESSONS

1. **Set aside internal resources to work with your chosen text analytics vendor. Transfer existing data and knowledge into the new system so that it performs to expectations when the integration is**
2. **Spend time preparing data before bringing in a text analytics vendor or consultant.**

> "What we've found is that you really need that expertise to understand how you can apply an existing text analytics tool set to your workflow."

THERE'S MORE TO TEXT ANALYTICS THAN DATA

Off-the-shelf tools alone won't always provide the desired capabilities, however. Unstructured data are one challenge businesses face. Nolan explains, "The biggest challenge that we've found with the unstructured data is all the different formats in which data can be represented. All the different ways that data can be captured. You hear people talk about the big data problem. We don't see it as a big data problem. We see it as a diversity of data problem."

"The first question to ask," he says, "is what format the data is in? Look at it, what does it look like? A lot of time the answer is it's in PDF format. PDF can mean a lot of things. Within PDF, you have all the different ways you can represent PDF, and what we find is that most of the time, what PDF means is they have images."

"My point is," he continues, "while somebody might say all this data is in technically standard formats, when you start to dig down into them, there's all kinds of funny business that people can do with those standard data formats that are completely legit and legal, but from a text analytics pipeline perspective can be a real nightmare." He says businesses spend a lot of time trying to build a pipeline, but the first part is to get all those data into a normalized form, and that can be challenging.

Normalizing data before bring in a text analytics consultant or deploying text analytics tools would be ideal, says Nolan, "but it might be a little unrealistic because most of the companies are not text analytics experts. If you could get everything into a common format that's well documented and understood, that would make the text analytics process easier and frankly would probably reduce costs."

"If you can't get to that bar, I would say for each type of data that you have, you should first have a good example of that data to show to the potential vendor," he advises. "And second, for every type of data, have some kind of description of what's in the data so that we can understand it."

> "You hear people talk about the big data problem. We don't see it as a big data problem. We see it as a diversity of data problem."

THE VALUE OF TEXT ANALYTICS COMES FROM BUSINESS CONTEXT

RAMKUMAR RAVICHANDRAN

Director, Analytics & A/B Testing, Visa Inc.

Ramkumar Ravichandran is the director of analytics and A/B testing in the Digital Product organization at Visa. His core responsibility is to enable a data-driven decision-making framework for Visa's leadership and stakeholders. Ram is passionate about the value of data and insights, and he loves sharing what he learns at conferences, in blogs on LinkedIn, and through Slideshare. He is also writing a book titled *Practical Predictive Analytics*, an overview of the analytical techniques that help deliver value to organizations.

Twitter | Website | Blog

Ramkumar Ravichandran's great revelation about the power of text analytics came when he joined a company in which research, the purview of text analysis, and business analytics fell under one manager. "That was the first time I realized the power of text analysis in complementing what we get from business analytics," says Ravichandran. Analytics is a numbers story that comes out of an analyst's head, but the reasons *behind* that story are based on what customers say. "Only when we put these two together can we get the complete picture of customer actions that show up in the analytics as numbers." Text analytics is not going to answer all business questions; it is a complementary technique that completes a picture, not a solution by itself.

KEY LESSONS

1. **Analytics is a numbers story that comes out of an analyst's head. You must look to what the customers say to find the reasons behind that story.**
2. **You must be able to translate the analysis back into actionable recommendations.**

A good example of how text analytics fits into a larger business process is how it serves a product development life cycle, which encompasses many business activities and types of analyses. It might begin with an envision stage, where people decide what they are doing and why they are doing it. There would be a review of business reports that provide metrics related to the business activity under consideration. There would be analysis of text from a variety of sources, such as social media, call center feeds, surveys, and reviews. This analysis would provide deeper insight into justifications of the product's development from the customer's perspective.

> "It's only when you look at other sources of information and insights that can you turn text analysis into a business recommendation that will benefit the company."

THE VALUE OF TEXT ANALYTICS COMES FROM BUSINESS CONTEXT

Based on this analysis, there might be several recommendations for possible approaches to the problem, and then A/B testing to narrow the choices down to one solution. The final decision would be based on analytical and customer data that show why this is the optimum approach. "It's only when you look at other sources of information and insights that can you turn text analysis into a business recommendation that will benefit the company," says Ravichandran.

To make text analytics really impactful, you must:

- **Consider your business model.** For instance, a retail business will rely on different kinds of text data than a business-to-business organization.
- **Understand the business need.** Is the analysis related to product development or user experience management? Is it going to improve an operational process?
- **Understand the questions you are trying to answer.** This involves considering actions you might take based on answers the analysis provides. You must also have a good framework for setting expectations. Text analysis does not answer all business questions, and a lot goes into preparing clean data for analysis. Finally, you must be able to translate the analysis back into actionable recommendations.

Text analysis can reveal unexpected results. Ravichandran cites an example in which an analyst was analyzing product reviews from customers. He decided to look at just five-star reviews to identify positive experience drivers. In the course of looking for what made customers happiest, he discovered that although these customers really liked the product, there was a pattern of comments revealing a problem with shipping. "This research, which was designed to find positive experience drivers, identified a problem that could easily be fixed through an operational process change," says Ravichandran.

Ravichandran advises that when doing business analysis, it is important to connect analytical dots from different places, then complement them with text analysis to answer the question, "Why."

"This research, which was designed to find positive experience drivers, identified a problem that could easily be fixed through an operational process change."

**Participation in this e-book is purely on personal basis and does not represent Visa in any form or matter. The advice in this article is based on learning from work across industries and firms. Care has been taken to ensure that no proprietary or work-related information of any firm is used in this material.*

TEXT ANALYTICS OFFERS POWERFUL TECHNIQUES FOR REVEALING HIDDEN INSIGHT

SETH EARLEY

CEO,
Earley Information Science

Seth Earley is CEO of Earley Information Science, an information management strategy consulting firm and the editor of data analytics for *IT Professional Magazine*. He also serves on the editorial board of Applied Marketing Analytics. His interests include knowledge strategy, data and information architecture, search-based applications, and information findability solutions. Seth conducts workshops for senior leadership on aligning information management strategy with measurable business outcomes and develops information governance programs for clients in multiple industries.

Twitter | Website | Blog

There are myriad use cases for text analytics in verticals ranging from retail and customer service, to life sciences, telecommunications and far beyond, according to Seth Earley. Many techniques and tools are available that allow practitioners to make sense of large amounts of content – such as email messages, customer service tickets or processes large numbers of research studies to identify and understand scientific trends. The key to realizing value is in applying structure to unstructured data and knowing what questions your company seeks to answer in the data and content.

If you are looking at a body of content that is well written or vetted, it is easier to find the structure. A news article is well organized (compared with a typical email message for example) and usually is about a specific thing or event. One way you can draw insight from it is by performing entity extraction. Entity extraction works best on well-formed content but can also work on less structured information like social media posts. "By creating a list of terms that are of interest – such as your products or those of a competitor (also known as a *controlled vocabulary*)," says Earley, "you can group comments according to those term occurrences and focus on the things that people say about your product or those of your competitor. You can also identify whether comments are positive or negative using sentiment analysis – another text analytics approach."

KEY LESSONS

1. **Text analytics involves finding structure in unstructured data so that you may properly analyze them.**
2. **There are many powerful techniques and tools that practitioners can use to uncover valuable business insight from their unstructured data.**

> "You can parse the information according to a process or problem, and then correlate that data with internal processes."

TEXT ANALYTICS OFFERS POWERFUL TECHNIQUES FOR REVEALING HIDDEN INSIGHT

You can process a wide range of content, including blog posts, discussion forums, trouble tickets, customer service calls, and transcripts of call center conversations. "You can parse the information according to a process or problem, and then correlate that data with internal processes," Earley explains. "If a user posts a complaint and describes how your product arrived late, was damaged in shipment, that they then did not hear back from customer service, then called and waited on hold for 20 minutes to reach an agent, and it took 4 weeks to get their refund, there are several internal process issues which that one piece of feedback identified. Pulling these kinds of insights from dozens or even a hundred email messages is one thing. Trying to process thousands or tens of thousands requires text analytics to identify themes and process issues. This is how voice of the customer data can be mined from unstructured text and used to identify problems in various departments."

There are specific business problems, even seemingly insurmountable ones, to which text analytics can be effectively applied, notes Early. A field service organization he once worked with was taking days to service equipment because the technical documentation they used was difficult to find and it was located in several different places. The organization had attempted to solve this problem several times before with no luck. It wasn't until they took a structured approach to analyzing the text and the content that they were able to reduce the amount of time that their engineers were looking for information by 50 percent.

The field service technicians were spending 40 percent of their time—approximately 16 hours a week—searching for information that they could not find. After categorizing the data so they could retrieve the specific content they needed, they were able to save eight hours per week per technician. With 3,000 service engineers across the organization, that amounted to $50 million per year in savings.

> "By doing some entity extraction, they were able to find underutilized content that allowed them to pursue some new research avenues."

In another example, Earley once discovered incredible business value derived from text analytics during a project for a life sciences firm that was trying to locate research about a particular drug. "They didn't have visibility into it because it was deep on the 30th page of search results," Earley says. "But by doing some entity extraction, they were able to find underutilized content that allowed them to pursue some new research avenues for some medications that were going off patent." Through applying auto-categorization techniques, they were able to mine the organization's deeper content and reveal vital, actionable information.

Text analytics is a powerful tool for unlocking hidden business value and solving complex business problems across a range of industries. By applying a set of best practices to a clearly defined business problem or question, you can achieve dramatic outcomes to move your business forward.

QUALITY TEXT ANALYTICS DEPENDS ON LARGE QUANTITIES OF DATA

RONEN FELDMAN

Full Professor at the Hebrew University, Amenity Analytics

Ronen Feldman is a professor of information systems at the Business School of the Hebrew University in Jerusalem. He received his B.Sc. degree in math, physics, and computer science from the Hebrew University and his Ph.D. in computer science from Cornell University. He is the chief scientist of Amenity Analytics, a New York–based company that specializes in the development of algorithmic trading tools based on its customizable text mining engine.

Twitter | Website | Blog

Text mining and text analytics have been going on for many years, but only in the past few years have systems and analytical engines made the task economically scalable. Ronen Feldman points out that some years ago he was able to run a financial text analysis on his high-performance Dell laptop: the analysis took nine hours. Today, the same analysis running in a low-cost public cloud takes a couple of minutes.
This is significant, because many highly insightful text analytics systems depend on analyzing vast quantities of data, and the results become valuable only if they are available immediately. To illustrate this point, Feldman describes the work he has done using text analytics to provide stock market projections for financial analysts.

"One of the most important steps in using text mining in finance is to identify every possible business event that relates to a company," says Feldman. "You can track hundreds of business event types." For example, even though a stock goes down considerably, an analyst or an industry expert might say they feel there is a huge opportunity in that stock. This expert's opinion is a valuable event. Part of the assessment involves knowing whose opinion that is and his or her historical track record in making predictions about that stock. It is possible to rate the sources based on their accuracy. And it's not just the source: where the source appears is important, too.

KEY LESSONS

1. **Many highly insightful text analyses depend on looking at vast quantities of data. The results become valuable only if they are available immediately.**
2. **It is important in this type of analysis not to interpret the meaning of any particular event, because one business event may have positive or negative effects in the context of other business events.**

> "One of the most important steps in using text mining in finance is to identify every possible business event that relates to a company."

QUALITY TEXT ANALYTICS DEPENDS ON LARGE QUANTITIES OF DATA

For instance, if something is published in *The Wall Street Journal*, it carries more weight than if it appears someplace no one has ever heard of. Other event examples include news or rumor of an acquisition, announcements or rumors of layoffs, rumors that employees are not happy, and rumors or announcements of new products.

It is important in this type of analysis not to interpret the meaning of any particular event as it relates to the stock value, because one business event may have positive or negative effects in the context of many other business events. Feldman's approach identifies the event, looks at past occurrences of that event for this company or similar companies in the same sector, and then looks at what happened to the stock in those cases. "What should we expect to see after such an event? We utilize the history to tell us what to expect." says Feldman.

"The whole approach is information fusion. You want to combine all the information from every possible source for analysis." This approach depends on a large body of historical data as well as powerful Web crawlers that continuously go to all the different websites to find every relevant business event for every publicly traded company. "The more information we collect, the better the analysis becomes," says Feldman.

There are several approaches to this kind of analysis. One is to define business event rules up front. Such an approach is not dynamic, however. Feldman uses an unsupervised machine learning framework to constantly analyze relevant sentences to see what kinds of relationships it can find. Every time the system finds a new instance, it examines that event, and then humans decide if they want to edit it into the list of business events for tracking. "This is a semi-supervised approach to machine learning," says Feldman. This approach makes it possible to work with hundreds of different event types and to quickly adapt as new instances appear in the market.

"The more information we collect, the better the analysis becomes."

KEEP IN MIND THAT TECHNOLOGIES CHANGE QUICKLY

VISHAL DESHPANDE

Data Analytics Solution Architect, Day1Solutions

Vishal Deshpande is an experienced IT professional with a record of developing and supporting successful, high-return IT projects and solutions that incorporate a range of applications and technologies. He is consistently called on to improve organizational effectiveness and efficiency through leadership that aligns the business processes and IT infrastructure to realize cost savings, accelerate performance, and sustain strategic flexibility. A natural team leader and mentor, Vishal excels in dynamic, demanding environments while remaining pragmatic and focused.

Twitter | Website

A great deal of mature analytics technology is available today to extract information and intelligence from structured data, but with the rise of the Internet and social media as well as the prevalence of email, business process documentation, and other kinds of audio and video information, most new data generated today are unstructured. This development is driving changes in the technology used to store and process data. "Many organizations are starting to look at text data, but the tools and methods are not standardized," says Vishal Deshpande.

Organizations are beginning to recognize that they have sources of intelligence within their archives that can provide practical operational benefits through text analytics. For instance, most organizations fail to use intelligence they could extract from their own email archives and content management systems. "Text analysis can be used to do everything from generate business intelligence from this company-owned text data to simply manage those resources more effectively," says Deshpande. For instance, text analytics can be used to automate the process of deciding which emails should be archived and which can be discarded.

KEY LESSONS

1. **Most data generated today are in an unstructured format.**
2. **Organizations fail to use intelligence they could extract from their own email archives and content management systems.**

> "Text analysis can be used to do everything from generate business intelligence from this company-owned text data to simply manage those resources more effectively."

KEEP IN MIND THAT TECHNOLOGIES CHANGE QUICKLY

Analyzing any kind of text is an iterative process. "Whether you are analyzing social media, feeds, HTML and PDF documents, or any other kinds of text, the first step involves converting the content to raw text." When the data have been reduced to raw text, they are available for analysis. Many organizations have no idea what they have in their text data. For instance, if they decide to mine their entire email or content management system archive, an initial text analysis can help identify and categorize the content. Topic modeling and semantic analysis can develop a taxonomy of key words. "Once this work is done, it becomes possible to form questions that relate to specific business needs, which can then be researched through deeper analysis of the content," says Deshpande. It's a discovery process that can begin with, "I have no idea what I need to look for," to a recognition of the kind of data you have, which allows you to drill deeper.

Technologies that are making it cost-effective to process high volumes of text are only a few years old. The technologies are combinations of lower-cost data storage and new tools capable of analyzing extremely large data sets. In addition, techniques are changing. New machine learning methods are being employed to analyze raw text data and give the data a structure so that they can be fed into forward-looking machine learning systems. "These systems generate probabilistic predictions of how a trend is going to shape up in the future," says Deshpande. As the technology matures, Deshpande sees text analytics moving beyond "event" or "question" analysis toward a continuous monitoring and streaming analysis tool that continuously analyzes vast quantities of text.

"Whether you are analyzing social media, feeds, HTML and PDF documents, or any other kinds of text, the first step involves converting the content to raw text."

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.

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.

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.**

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

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.

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

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."

KEEP IT SIMPLE

JOSÉ MARÍA GÓMEZ HIDALGO
Analytics Manager,
Pragsis Technologies

José María Gómez Hidalgo has been a lecturer and researcher at several universities as well as director of research and development at security firm Optenet. Currently analytics manager at Pragsis, he leads and develops business analytics on big data projects in finance, CRM, insurance, and e-commerce. José María has published research results in text analytics applied to security, biomedicine, e-commerce, and information access. His research interests include natural language processing and machine learning, text and data mining, and information retrieval.

Twitter | Website | Blog

When it comes to sentiment analysis, Pragsis Bidoop analytics manager José María Gómez Hidalgo says, people tend to make a common mistake early on: they focus on technical sophistication when a simpler approach might actually be better. Here's his advice for avoiding that gaffe:

- **Focus on the business problem.** People work with text analytics because they find themselves awash in valuable text data that they want to exploit. Gómez Hidalgo observes that solution providers often push them toward sophisticated automation that may be several strides beyond what they actually need. "The customer just wants to get knowledge from the sentiment information," Gómez Hidalgo states. "Users need a good interface that allows them to get information about sentiment instead of the ultimate sentiment classifier."
- **Run lots of data.** But keep the process simple. This advice comes from Gómez Hidalgo's research experience and applies to all forms of data, though he calls it an imperative for text analytics. Automated solutions that rely on simple algorithms to process big textual data sets are a smarter bet than pushing small amounts of text through "complex and clever" algorithms, he asserts. Why? Because even when text data are analyzed *en masse*, algorithmic accuracy levels remain roughly consistent regardless of whether they're based on simple or complex formulas. "The simplest algorithms will get better results long term," he states.

KEY LESSONS

1. **It's wiser to run a lot of text through simple algorithms than to run a little text through complex algorithms.**
2. **It's vital to complement unstructured text analytics with structured data.**

> "The simplest algorithms will get better results long term."

KEEP IT SIMPLE

- **Complement text with other analytics.** Unstructured text analytics opens a lot of interesting insights about customers, Gómez Hidalgo notes, but eventually you must complement unstructured text with structured data. For instance, he is currently developing a sentiment analytics system for an e-commerce company that's building a recommender module. By compiling and analyzing social network inputs, the company will get a good gauge on brand and product sentiment. Those insights alone won't be sufficient, however. "You need the *where* and the *when*," Gómez Hidalgo states. "Where is the best place to make a recommendation, and when do you have to do it to get the best return?" Those are vital structured data elements that must be included. "If you don't, you won't be solving the customer's full problem," he warns.

Gómez Hidalgo describes two main approaches to sentiment analytics. You can give weight to sentiment-related words that linguistics experts have evaluated and tagged, or you can deploy machine-learning text classifiers based on previously labeled text collections. Whichever you choose, he advises, keep well-trained experts close at hand. At best, machine learning systems offer an imperfect snapshot of sentiment. The inherent ambiguities of such linguistic elements as irony and sarcasm are still too much for unaided machines to master.

"You need a human being to retrain the system, to make it cleverer and also to complement the analysis that the automatic system does," he says. "You need the humans at the end."

"You need a human being to retrain the system, to make it cleverer and also to complement the analysis that the automatic system does."

WHEN "MAD MEN" AND MATHEMATICS COLLIDE

PAUL HOFMANN
CTO,
Space-Time Insight

As CTO, Paul Hofmann draws on more than 20 years of experience in enterprise software, analytics, and machine learning. He has held executive roles at BASF and SAP, where he was vice president of research and development, and conducted academic research at the Massachusetts Institute of Technology. He taught at the Technical University in Munich, and Northwestern University. Most recently, Paul served as CTO of Saffron Technology, which has been acquired by Intel.

Twitter | Website | Blog

Paul Hofmann, a statistics, machine learning, and semantics expert, offers several insights to help people in business benefit from text analytics by enlightening their thinking on the subject:

- **Understanding humans by understanding text.** Data volume on the Internet and elsewhere is growing exponentially, and most of it is human-generated text, Hofmann notes. "This is why text analytics is necessary," he says. "If you want to understand people, especially your customers and how they use your product and engage with your organization, then you have to be able to possess a strong capability to analyze text."
- **Context matters.** Marketers (notionally the "Mad Men of Madison Avenue") spend a lot of time and money understanding the touch points customers reach on their pre-purchase journey. Attribution, in that sense, equals context, and text analytics is a key ingredient in establishing that context. The concept is important not only in marketing but in government, research and development, finance, and elsewhere. "Attribution is in effect a form of predictive modeling," Hofmann states. "It is a way to predict, based on past behaviors, what a customer likely will do. For that we have to combine human sentiment with the trace customers leave on the Internet for example" he says. "To understand sentiment we need quite complex text analytics."

KEY LESSONS

1. **Humans generate most data, so understanding text is crucial.**
2. **The days of intuition-driven 'Mad Men' are over.**

> "If you want to understand people, especially your customers... then you have to be able to possess a strong capability to analyze text."

WHEN "MAD MEN" AND MATHEMATICS COLLIDE

- **Live not by text alone.** If not for combining text and statistical data, it is unlikely he would have made the breakthroughs he did for an aeronautics client. You will see, at best, half of the picture. After all, combining the two, he states, is essentially what the human brain does. "You have to combine semantics with statistics," Hofmann says. "I am very deeply convinced of that."

Hofmann considers the television show "Mad Men" a snapshot of marketing in transition, one that underscores his point. By the end of the show's run, its cynical ad men were conceding, if grudgingly, that there was business value in the primitive computer that had been installed in their agency. Of course the insights gleaned from a computer in 1970 would have been limited, but today, he says, those limits have all but disappeared. In fact nowadays compute power and the plethora of available and capturable data provide tremendous insights. In today's world, and tomorrow's too, it is an absolute necessity to take advantage of the ability of machines to predict a customer's behavior, he remarks.

"We have been transitioning from these guys who are a little bit chauvinistic, who infer or assume behavior while they smoke the cigars and drink the booze—and then make a commercial," Hofmann says. "Gut feeling is being replaced by actual insights derived by analyzing big data."

Hofmann points to the exploding field of marketing attributions as a crossroads where text data and statistics merge to generate deep insights that Madison Avenue could scarcely have dreamed about five decades ago. Customers create a recording and story by leaving digital markers along their unique engagement with vendors and products. Today, we have fine-tuned tools to read and analyze those markers. The ad man's intuition is supplanted by machines that learn.

> "Today, we can really analyze and predict people to understand the feelings and intentions they have about a certain product."

"Before, those guys on Madison Avenue had it all in their head and they had to figure it out by inference and lengthy validation cycles," Hofmann says. "Today, we can really analyze and predict people to understand the feelings and intentions they have about a certain product. This enables the marketer to speak directly to buyers and consumers."

"That's the holy grail of marketing," Hofmann concludes. "We all want to predict why and when a customer will buy, and when the part will fail before it actually breaks, and to do that, text analytics is a critical necessity."

What are your customers saying about you?

Receive a complimentary analysis of your text data!

Unstructured data can be overwhelming and difficult to interpret. Our exclusive Mighty Guides FREE offer will help you make sense of your data so that you can make better business decisions!

Click here to access this offer!

Angoss
Predict. Act. Perform.