

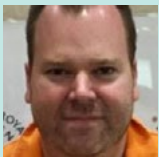


# FOREWORD

You have no doubt read an article, seen a conference speaker, or listened to an expert on a podcast tell you that business people need to become “data-driven decision makers.” There’s no question about the value of this approach—it works—but I don’t think that marketers get enough credit: we’ve been basing our decisions on data for years. We conduct focus groups for new products, we assess gross ratings points for our media, we look at the response rates of our direct marketing campaigns, and we conduct post-campaign analysis. Even though our line of work is often labelled *creative*, we are results driven and judged on figures. To this end, many of us have successfully made the jump and seen amazing results from not just looking backwards and measuring effectiveness but also looking forward and predicting effectiveness.

Until recently, analytics (even predictive analytics) has largely been viewed as a numbers game, but natural language processing and the ability to conduct analysis on unstructured data—that is, words—are opening new doors for marketers. Evaluating your brand equity is no longer confined to the number of responses to a survey; it’s as vast as Facebook, Twitter, and YouTube. The customer experience is no longer just anecdotal sales feedback but direct analysis of thousands or millions of customer emails, calls, and comments. Of course, these ideas are just the beginning. As “creative types,” we know that we’re limited only by our imaginations, even in the areas of analytics.

Angoss Software has [published an e-book](#) containing interviews from 28 experts on their experience—the value and the possibilities of text analytics. The shorter edition that follows here is a selection of these interviews that get to the heart of the potential of text analytics specifically for marketers. We hope it gives you the insight and inspiration you need to take the step onto the leading edge.



Mathew McLarnon  
Senior Campaign Marketing Manager  
Angoss Software

## 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.

# Applications for Text Analytics in Marketing



**Jon Lehto**  
Constant Contact.....5



**Jeff Catlin**  
Lexalytics, Inc.....7



**Helen Clegg**  
A.T. Kearney.....9



**Seth Grimes**  
Alta Plana Corporation.....11



**Gershon Bialer**  
CrunchBase.....13



**Alessandro Zanasi**  
Zanasi & Partners.....15



**Paul Hofmann**  
Space-Time Insight.....17



**Roman Kubiak**  
Senior Consultant.....19

# 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



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

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."

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



## 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.

# 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.  
”



# 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



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

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 Bally’s, a low-cost venue. Bally’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.”

“You can dig out facts, but if you don’t go deep enough, the facts won’t tell you much.”

## 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.



# 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.

”



# 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



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

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.”

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.”

“

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. ”

# 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



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

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.

“ Different sources may provide different, divergent insights about a given topic. ”



## 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.

# 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.

”

# 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



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

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.

“ 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. ”

## 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.





# 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 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.

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.

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





# IMAGINATION AND FANTASY



**ALESSANDRO  
ZANASI**  
President,  
Zanasi & Partners

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



Twitter | Website



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

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

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

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

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

## KEY LESSONS

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



# IMAGINATION AND FANTASY

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

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

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

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

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

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

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

”

# 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



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

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.”

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

## KEY LESSONS

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



# 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.

“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.”

“

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

”

# 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



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

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.”

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

## 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.

# 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.

”





Thank you for reading a selection of essays from our Text Analytics e-book.

If you'd like to download the full version of the e-book in which **28 Experts Share How to Achieve Business Value** please click the button below.



[Download the Full E-book](#)

# 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!

