

# **Business Intelligence Network™ Research Report**

# Voice of the Customer

# Text Analytics for the Responsive Enterprise

By Seth Grimes

BeyeNETWORK • 1790 30th Street • Suite 310 • Boulder, CO 80301 • 303-339-7255 • www.BeyeNETWORK.com



# TABLE OF CONTENTS

Introduction	1
The Voice of the Customer	2
Needs and Challenges	2
Sources and Voices	3
VoC Elements	7
VoC Processes	9
From Findings to Action: User Experiences	14
Technical Insights	16
Technical Context	16
Text Technology Basics	17
"Unstructured Data," BI and Predictive Analytics	18
Sentiment Analysis	20
Deployment Options	23
Market Outlook	24
Study: VoC Text Analytics Best Practices	25
The Value of Best Practices	25
Survey Findings	25
Advice to End Users	29
Business Objects Voice of the Customer Solution Profile	32
IBM Voice of the Customer Solution Profile	40
SPSS Voice of the Customer Solution Profile	45
IBM Customer Case Study: NTT DoCoMo	51
SPSS Customer Case Study: Cablecom GmbH	54
Author Information	57



# INTRODUCTION

Text analytics helps business users discern and capture the Voice of the Customer from online media such as blogs, forum postings and news articles; from email, chat interactions and contact-center dialogues; and from surveys and other mechanisms for collecting customer feedback: from the totality of enterprise information sources.

What are your organization's customers – and your prospects and the media – saying about your company and your products and your competition? Voice of the Customer (VoC) initiatives can answer that question and help you formulate your response. VoC is not just an extension of customer relationship management (CRM) to the customer experience. It is an approach that can guide enterprises in meeting the spectrum of sales, marketing, customer support, brand and reputation management, product and service design, and quality demands.

Businesses seek to understand the totality of customer needs and opinions, whether explicitly stated or indirectly implied. Businesses have faced the obstacle that the majority of relevant information is found in "unstructured" text outside an organization's databases. Text technologies overcome this obstacle by making relevant information accessible to business analysts, managers and executives who fill critical sales, marketing, support and product management roles regardless where the information is found. Text analytics – solutions that convert human language into data for business intelligence (BI) and predictive analysis – has breathed new life into the time-tested VoC business concept.

VoC analysts probe both individual views and collective, market thinking, which we might term Voice of the Market. Important information is no longer found only in corporate-sponsored and internally held sources. Text analytics creates the ability to discern and capture the voice of the customer from online media, such as blogs and forum postings; from email, chat interactions and contact-center dialogues; and from surveys and other mechanisms for collecting customer feedback, complementing traditional transactional sources.

This report describes how text analytics has become an essential part of Voice of the Customer solutions. It describes VoC techniques and processes and explains the fundamentals of text analytics technology and solutions. It covers implementation options and best practices, addressing the basic question: How do organizations get started with VoC text analytics? And it presents the findings of a best practices survey that asked end users and consultants about their goals, information source, return on investment (ROI), and advice for organizations looking into VoC text analytics.

The report concludes with profiles of leading Voice of the Customer text analytics suppliers – Business Objects, IBM and SPSS – accompanied by in-depth customer case studies.



# THE VOICE OF THE CUSTOMER

"I have learned, based on my experience, that everything is dominated by the market. So whenever we are struck with any obstacles or difficulties, I always say to myself: 'Listen to the market, listen to the voice of the customer.""

- Yoshio Ishizaka, Toyota Motor Company, Executive VP

Voice of the Customer is a time-tested business concept that has gained new life through the application of text analytics. The idea is simple: listen to customers, understand their needs, respond and measure the effectiveness of the response. Execution is not necessarily simple, however, given the wide variety of ways in which customers – and prospects and the collective voice known as the market – express themselves. The responsive enterprise starts by listening, by determining what customers are saying and seeking ways to turn this information into paths around business obstacles and difficulties.

## **NEEDS AND CHALLENGES**

Organizations seek answers to common, fundamental business problems in the voice of the customer:

- Are our customers satisfied with our products, services and support?
- What do customers and prospects think of our competitors and their products and services?
- What do they like, what problems have they had, and what additional offerings or features would they like to see... and be willing to pay for?
- Can we trace issues to root causes?
- Who are our most profitable customers?
- Can we build models that will predict when a customer is likely to drop our products or services?
- Can we profile the most promising prospect and understand how to reach them?
- Who influences the market, and how do opinions propagate?

More generally, organizations ask:

• How can we improve customer satisfaction?



- How can we maintain and increase our competitive edge?
- How can we boost profitability?

Some of the answers to some of these questions may be found in transactional systems, in records of purchases and service utilization; returns, complaints and warranty claims; and website clickstream including diverse data on page referrals, dwell time, abandoned shopping carts, and the like. The bulk of answers, however, are to be found in non-transactional "unstructured" sources, and information from those sources provides far more explanatory power than information captured in transactional systems. Further, organizations do not capture at all the opinions of people a business would like to win as customers and of market influencers in transactional system.

The need is to hear, understand and act on customer, prospect and influencer views. But important "unstructured" information sources are difficult to access and analyze because of their volume and diversity. Those sources include non-transactional, operational systems such as contact center transcripts and notes, instant-messaging logs, point-of-service notes, complaints and service requests transmitted via web forms and email, and warranty claims. They include feedback collected via surveys. And they include stakeholder "talk" that is directed to (other) customers rather than to business analysts: forum postings, news and review articles, and blogs and other social media. This information is conveyed in text, in narrative form, rather than in structured, encoded records.

The biggest VoC challenges are to identify and access all relevant sources and make sense of their contents. These are information retrieval, natural-language processing and text analytics challenges.

## **SOURCES AND VOICES**

"We know these [customer-satisfaction] initiatives are making a difference because our customers are telling us. Through our Voice of Customer survey, where we hear from over 200,000 customers a week, we see improvements in likelihood to recommend, find and buy, and our clean and uncluttered metrics."<sup>1</sup>

Paul Raines, Home Depot, Executive Vice President of U.S. Stores

**The voice of the customer** is expressed in different forms through different channels. Every organization should understand the voice channels available and should assess the coverage, value and relative importance of each source.

<sup>&</sup>lt;sup>1</sup> http://seekingalpha.com/article/54060-home-depot-f3q07-qtr-end-10-28-07-earnings-calltranscript?page=-1.

**b**•eye Business Intelligence Network™ The Global Vision for BI and Beyond

We can classify voice sources in three basic categories:

- Operational systems
- Solicited feedback
- "Open source" intelligence

#### **OPEN SOURCE INTELLIGENCE**

The open-source intelligence category includes the stakeholder talk cited earlier: forum postings, news and review articles, blogs and other social media. We refer to this information as "open source" because it is available to anyone: not to only customers and prospects and your company, but also to your competitors. This information appears in print and in online publications, but unfiltered, unedited sources are equally important. While conventional publications have brand authority, the social web of forums and blogs and micro-blogs (e.g., Twitter) convey a sense of authenticity and immediacy conventional channels can't match.

Minqing Hu and Bing Liu, in a 2004 paper, "Mining and Summarizing Customer Reviews<sup>2</sup>," explain the challenge:

In order to enhance customer satisfaction and shopping experience, it has become a common practice for online merchants to enable their customers to review or to express opinions on the products that they have purchased... As a result, the number of reviews that a product receives grows rapidly. Some popular products can get hundreds of reviews at some large merchant sites. Furthermore, many reviews are long and have only a few sentences containing opinions on the product. This makes it hard for a potential customer to read them to make an informed decision on whether to purchase the product. If he/she only reads a few reviews, he/she may get a biased view. The large number of reviews also makes it hard for product manufacturers to keep track of customer opinions of their products. For a product manufacturer, there are additional difficulties because many merchant sites may sell its products, and the manufacturer may (almost always) produce many kinds of products.

Hu and Liu describe mining review sites, but what they write applies equally to forums, blogs and other social media. These non-establishment sources are universally online. ("Online" effectively equals "on the Web," currently. That equation is changing as group mobile connectedness becomes more pervasive.) They are valuable for brand monitoring, reputation management and competitive intelligence: all VoC applications.

Given that both conventional publications and non-establishment sources are universally online, the VoC analyst's first job is information retrieval (IR):

<sup>2</sup> Mingqing Hu and Bing Liu, "Mining and Summarizing Customer Reviews," KDD '04, August 2004, http://www.cs.uic.edu/~liub/publications/kdd04-revSummary.pdf.



identifying and monitoring sources and fetching items as they are posted. Text analytics automates the process of making sense of the retrieved information and extracting meaning and trends.

Emerging brand-monitoring tools and services subscribe to syndicated RSS data feeds and also crawl the Web, both operating continuously and supporting ad hoc inquiries. While many of these tools and services rely primarily on manual methods, automated systems can deliver greater monitoring reach and speed and flexibility, for instance to apply semantically rich IR that understands concepts and context and not just keywords.

#### SOLICITED FEEDBACK, A.K.A. SURVEYS

Solicited feedback is collected via systematic surveys and via quick, ad hoc questions, the latter akin to former New York Mayor Ed Koch's famous and frequent "How'm I doing?" The point in each case, systematic or quick, is to actively seek out stakeholder views.

Surveys may be qualitative, which traditionally has meant relying on focus group discussions, and they may be quantitative, based on a questionnaire, a survey instrument, distributed to a carefully designed population sample. A "census" that seeks responses from the totality of stakeholders is another option. And a survey may be repeated periodically to produce a longitudinal picture, to detect and illuminate trends.

Survey research is a long established practice although industry surveys are frequently non-scientific and informal. If the goal is to draw a statistically accurate picture of a population or to create a predictive model of stakeholder behavior, formal methods help. But if the goal is to identify potential problem cases, perhaps by evaluating survey responses against some form of predictive model – and this evaluation can be as simple as looking for certain keywords and response patterns that hint at issues – then rigorous sampling would not be in order.

Organizations can survey stakeholders at the point of service, which is particularly easy to do via a simple form when products and services are delivered over the Web, or they can solicit survey participation by email, telephone or post. Analyses typically involve frequency counts for coded-response questions and crosstabulations that illustrate relationships among variables. The traditional approach has been for people to read and summarize free-text responses, which are known as verbatims, but clearly this approach doesn't scale when more than a few responses are to be analyzed.

Text analytics offers the ability to identify and extract topics, opinions and sentiments from survey verbatims. The technology can also classify responses according to pre-established topics. As part of what is sometimes termed *Enterprise Feedback Management*, text analytics can produce extremely high survey-analysis ROI figures by eliminating the need for humans to read individual verbatim responses.

The major shortcomings of solicited feedback are two. The request for feedback reaches only individuals with whom an organization has directly interacted. And the



information gained is not unprompted – it is therefore expressive of an intensity of feeling that would lead someone to comment without being asked – so that it represents a skewed view of a stakeholder population. These forms of *bias* mean that surveys, if they are used for predictive purposes, should be used in conjunction with other means of hearing the voice of the customer.



#### **OPERATIONAL SYSTEMS**

Operational systems record an organization's interactions with customers and prospects. Some capture transactions – purchases, returns, service utilization, reservations, inquiries, etc. – but provide sparse qualitative information. Others track encounters, for instance, a customer's or prospect's (or competitor's) website visit, which may or may not have led to a purchase, captured as a clickstream in webserver log files, similarly accompanied by little qualitative information. But other operational systems record information-rich narrative that can serve as a valuable source of customer-experience information. Such systems are one of the first sources an organization should analyze to identify product and service quality issues as well as opportunities.

Contact-center notes and transcripts are at the top of the analysis list for most organizations embarking on VoC text analytics. The reason is clear: contact (or call) centers are a business's front line in handling customer inquiries, complaints and service requests. They are also, often, a business's last opportunity to placate and hold on to dissatisfied customers. Indeed, some VoC systems, or rather analytical systems that score models derived from VoC data, are integrated into contact-center and CRM software applications. These integrated systems can indicate churn propensity and suggest remedies as well as up-sell and cross-sell offers in real time.

Other forms of input come into play as well, albeit more in business-to-business (B2B), and less in business-to-consumer (B2C), settings: requirements documents, product specifications, contracts and reports.

## **VoC ELEMENTS**

Let the customers talk, give them a chance to tell you in their own voice the reasons and provide you with suggestions. It works better than you guessing what the answers might be and suggesting those.

- Avinash Kaushik, Google analytics evangelist

It is essential that Voice of the Customer solutions capture stakeholder voices in the forms and venues where they are expressed. Yet not all *utterances* carry equal value. Preferred VoC solutions will concentrate on information that is focused, useful, clear and usable.

#### FOCUSED FEEDBACK

Avinash Kaushik, formerly Director of Research & Analytics at Intuit, currently analytics evangelist at Google, and known for his Occam's Razor blog, says there are "three questions no [website] survey can live without (and often a survey can be pretty awesome with just three questions)." Those questions, condensing from a blog



posting<sup>3</sup>, are:

1. What is the purpose of your visit to our website today?

[This question] can also be framed as "What is the reason for your visit today?" or "What task are you looking to accomplish on our website today?" or "Why are you here today?"

- 2. Were you able to complete your task today?
- 3. If you were not able to complete your task today, why not?

[This question] can also be framed as "If you were able to not complete your task, please explain." or "Why were you not able to complete your task on our website today?" or simply as "How can we improve our website to ensure you are able to complete your task?" etc.

The answer to this [number 3] question is open-text VoC, Voice of the Customer.

Kaushik notes that "clickstream data is often missing the context, in that absence we overlay our own opinions / experiences / perspectives to make sense of it all."

On the analysis front, he offers a helpful start:

Analysis for this question is done by categorizing the responses into common themes and then rating the % of times each theme is occurring in the open ended VoC for those who are not able to complete their task. This is your simple and direct to-do list of issues directly from the horse's mouth about what you should work on in order to improve your website experience for your customers.

#### THE NET PROMOTER SCORE

Business strategist and customer-loyalty guru Fred Reichheld would have organizations focus further by asking customers "one simple question": Would you recommend us to a friend or colleague? Respondents are classified as promoters, passives and detractors. The percentage of promoters minus the percentage of detractors equals the Net Promoter Score (NPS). Reichheld claims NPS to be "the single most reliable indicator of a company's ability to grow."

The Net Promoter Score is attractive as a satisfaction measure that indicates growth potential. It is designed to be utilized as part of an improvement program: "The business goal here isn't merely to delight customers, it's to turn them into promoters – customers who buy more, and who actively refer friends and colleagues." <sup>4</sup> The score on its own, however, does not illuminate the causes of customer satisfaction or dissatisfaction, nor does it suggest remedies that can lead to improvement.

<sup>3</sup> Avinash Kaushik, "The Three Greatest Survey Questions Ever," http://www.kaushik.net/avinash/2007/04/the-three-greatest-survey-questions-ever.html.

<sup>&</sup>lt;sup>4</sup> http://www.netpromoter.com/improve/index.php



In the VoC context, the score, when applied to segments of the customer population, clusters happy customers and also points to customers whose needs have not been met. NPS provides direction for further study of primary materials – materials that more extensively capture customer voices – supporting root-cause analyses that can lead to process, product and service improvements that will ultimately boost the singular NPS rating.

#### **CUSTOMER LIFETIME VALUE**

Just as all survey questions are not created equal – they don't all provide the same value illuminating customer satisfaction, nor in pointing to root causes of customer issues – different customers have different business value. Modeling and estimating *customer lifetime value* (CLV) allows an organization to weight the response to issues experienced by a particular customer by the profit to be expected in retaining that customer. More profitable customers are, simply put, more important to most organizations and are thus, once basic service-level agreements and product-quality standards are met, deserving from a business point of view of higher levels of attention.

CLV calculations take into account many factors that typically include an estimate of churn propensity; retention cost; and past and projected customer spending and the cost of provided goods, services, and support. The latter factors are weighted by money cost over time.

More sophisticated models may take into account indirect value to be derived from a customer, for instance, the likelihood that he or she will influence acquisition and retention of other customers' business: whether he or she is a promoter or a detractor.

CLV findings may be used in marketing and not just in support and retention decisions. For instance, Home Depot has indicated that 2% of customers drive nearly 30% of sales. Through analytics, the company is "gaining a better understanding of who these customers are and what their unique buying patterns are."<sup>5</sup> Understanding CLV allows a forward-looking organization to design products and services and tailor support to its highest-value customers. And it allows such organizations to craft marketing and customer-acquisition strategies that appeal to the prospects with the highest value potential.

CLV calculations take into account VoC text analytics findings and are used in crafting VoC responses.

## **VoC Processes**

Voice of the Customer programs can and should be implemented via well defined, structured approaches that start with business goals and then determine what

<sup>&</sup>lt;sup>5</sup> http://seekingalpha.com/article/54060-home-depot-f3q07-qtr-end-10-28-07-earnings-call-transcript?page=-1

**b**•eye Business Intelligence Network™ The Global Vision for BI and Beyond

information is needed to enable decisions that help businesses meet those goals. Once a business understands information needs, staff can decide what data to collect and work to discover where that data resides and determine the analyses they need to carry out.

It is not enough to know that, as Sid Banerjee, CEO of text analytics firm Clarabridge, explains, industries such as travel, hospitality and retail "live and die on customer experience." This statement is a call to action, but to make a difference, businesses must understand:

- How and where to detect actionable customer-experience points.
- How to assess findings and trace issues back to root causes.
- How to plan for change and estimate its impact and analyze cost/benefit.

Businesses must then:

- Implement.
- Measure effectiveness.
- Evaluate return on investment (ROI).
- Improve products, services and support to raise satisfaction levels.

An effective VoC program is data driven with clear goals and expectations.

#### SETTING GOALS

Voice of the Customer program goals reside somewhere between the very general questions presented in the NEEDS AND CHALLENGES section –

- How can we improve customer satisfaction?
- How can we maintain and increase our competitive edge?
- How can we boost profitability?

and the fairly specific business-problem statements that precede them in that section. Goals that are frequently offered, albeit not one-size-fits-all, include:

- Increased sales to existing customers
- Improved new-customer acquisition
- Lower average cost of sales, new and existing customers
- Higher customer retention/lower churn
- Higher satisfaction ratings



• Fewer issues reported and/or service complaints

Achieving focused goals translates into answering high-level business needs such as improving satisfaction, increasing competitiveness and boosting profitability. Progress toward focused goals is decidedly measurable. For instance, instead of talking ethereally about improving satisfaction, we use an indicator, in this case satisfaction *ratings*, and set a time frame for progress. We track change over time to demonstrate the hoped-for progress.

#### **INFORMATION GATHERING**

Information gathering and filtering – classifying and filtering source materials based on value – constitute the first stages in what is essentially a processing pipeline.

It's fine to talk about a totality of enterprise information sources – online media such as blogs and forum postings; email, chat interactions and contact-center dialogues; and surveys and other mechanisms for collecting customer feedback; as well as traditional transactional sources – but not all sources are equally accessible or promise equal value. Analysts should choose sources that promise to be of use in meeting business goals. For instance, an organization looking to identify customer issues and boost customer satisfaction might look first in contact-center materials while a brand/reputation-management initiative would look at material posted in news and social media.

A variety of sources means a variety of access methods. Operational systems– callcenter, email, and CRM systems and the like – are accessed via some form of enterprise application integration (EAI) or extract, transform and load (ETL) from operational to analytical databases. Software can subscribe to some web sources such as blogs and news feeds via RSS, while other web sources, once identified via search or other means, are accessed by programs that regularly visit host sites. Survey management systems are another frequently important component, often freestanding but sometimes integrated with web and call-center applications. The information they collect is typically available in database or file formats.

When analyses cross system/source boundaries, we have to consider integration issues. That is, we need "key" fields that allow us to join information from disparate sources. Multiple sources means richer, more accurate analyses (via the enrichment and triangulation approaches described in later report sections), but they also add complexity.

#### FACTS AND OPINIONS

Text analytics researcher Bing Liu remarks that there are two main types of textual information on the web, facts and opinions. That remark extends to other VoC sources. Even surveys, which are designed to capture opinions, may contain factual information.

Facts and opinions are subject-predicate-object *triplets* that also contain descriptive modifiers. What distinguishes opinions from facts is that opinions convey sentiment, they convey *attitudinal* information.



"I bought a Motorola RAZR V3 last October" is a factual statement (although it may not, of course, be true). "Motorola rates the RAZR V3's talk time at up to 7 hours, but I've been getting only 4 hours" is a statement that contains two facts, about the phone's rating and about my experience. "I'm disappointed in the phone's talk time" is an opinion.

Facts and opinions are both important to VoC analyses but really in different contexts. Opinions and opinion sources come to the fore in VoC market and competitive analyses, in product design, and in applications such as brand and reputation management. The technologies used in identifying opinions are very similar to those used to discern facts in text, with the added twist that understanding the nuances of human speech related to sentiment is a difficult task. Factual information, whether extracted from text or derived from transactional sources, can provide much-needed context that can help in making sense of opinions.

To get a sense of approaches to opinion processing, consider an excerpt from the Hu-Liu paper, "Mining and Summarizing Customer Reviews," cited in the SOURCES AND VOICES section of this report:

Given a set of customer reviews of a particular product, the [mining] task involves three subtasks: (1) identifying features of the product that customers have expressed their opinions on (called product features); (2) for each feature, identifying review sentences that give positive or negative opinions; and (3) producing a summary using the discovered information.

There are many examples of applications of opinion and sentiment extraction in VoC initiatives that encompass review sites, blogs, forums and the variety of online social media and feedback channels. Take as an example Unilever, a forward-looking producer of consumer packaged goods (CPG) that has applied text technologies to understand consumer opinions and sentiment. According to consultant Tom H.C. Anderson of Anderson Analytics, the analysis process applied in studying the Dovebrand pro.age campaign starts with surveys and web scraping, via software that automates access, from online consumer forums. Sites such as *www.campaignforrealbeauty.com* and *www.doveproage.com* host many thousands of messages with potential value. Anderson's approach codes and characterizes data, looking for sentiment polarity – positive, negative and neutral – seeking to understand emotions and attitudes. It applies a *triangulation* process with a 43-attribute "psychological content analysis" and with human coding of random sampling of records that validates results discovered through automated text analytics.

#### TRIANGULATION

VoC analysts may need to consider multiple data points to estimate a quantity that cannot be measured accurately with a single mechanism. Such an approach, which dates back half a century in the research context, is called *triangulation*. Per a paper, "The Point of Triangulation" by V.A. Thurmond:

Triangulation is the combination of at least two or more theoretical perspectives, methodological approaches, data sources, investigators, or data analysis methods. The intent of using triangulation is to decrease,



negate, or counterbalance the deficiency of a single strategy, thereby increasing the ability to interpret the findings.

Related to VoC, to provide one example, we consider that business spends huge sums shaping brand image and promoting brand awareness.<sup>6</sup> To gauge the effectiveness of particular campaigns, corporate marketers will study transactions, for instance sales made in response to direct mail or using coupons, web-page visits and ad clickthrough, etc. But study of past transactions is of limited use in understanding potential buyers who are not responding to market messaging, in understanding competitive positioning and in picking up on nascent trends. Surveys and social-media mining, especially for attitudinal indicators, and utilizing text analytics, can fill the gap.

Kate Niederhoffer, who is VP, Measurement Science at Nielsen Online, explains,

Triangulation (in text analysis) means validating insights and substantiating the narrative derived from conversational data. It is critical to dissect this data in varied ways (e.g. what's being said, when, by whom, how, etc.). One must take multiple perspectives, combine manual and automated approaches, perform quantitative and qualitative analysis, and layer qualified interpretation.

Consumer generated media has endless depth, if not only because it offers a combination of structure (e.g. dynamic links) and content (e.g. text). With such varied venues for consumers to express themselves—and the inherent complication of naturalistic conversation (e.g. slang, sarcasm, jargon), it is vital to have similarly varied research methodologies. Organizations should take a triangulated approach in order for insights built on naturalistic opinions to be trustworthy.

#### **DATA ENRICHMENT AND ANALYSIS**

Another form of refinement relies on various forms of record linkage: on demographic data enrichment and on matching VoC data to transactional records. Data enrichment, on the one hand, attaches descriptive statistics to VoC-extracted information, whether we know the precise identity of the individual subject or not. All we really need to know about the subject – the person who anonymously posted an opinion in an online forum or chat room or who was tracked visiting a website – are basics that may be available to us where details are not. For example, we can generally infer someone's geographic location from her or his IP address. With a geographic location in hand, we can associate a spectrum of information such as (typical) income range, marital status, number of children, age, politics and education to the poster's attitudes.

In those cases where we know the identity of our interlocutor (caller, website visitor, forum poster), we can link the expressed attitudes and issues to the person's

<sup>&</sup>lt;sup>6</sup> Material excerpted from Seth Grimes, "Sentiment Analysis: A Focus on Applications," *http://www.b-eye-network.com/view/6897* 

**b**•eye Business Intelligence Network™ The Global Vision for BI and Beyond

transactional record to know, for example, where that person shopped, what that person bought, what attempts there were to correct an emerging product or service issue, etc., that led to what that person had to say.

In both scenarios, we provide context that improves our understandings of the voice of the customer.

## **FROM FINDINGS TO ACTION: USER EXPERIENCES**

"We've uncovered concepts and relationships in text that would be too costly – or even impossible – to detect by any other methods. We can now combine multiple data sources to evaluate customer expectations and improve customer satisfaction by employing more one-to-one customer contact and preemptively resolving customer complaints to keep our retention rates high."

– Federico Cesconi, Cablecom, head of customer insight and retention

Few businesses would view Voice of the Customer as an end in itself, simply of academic interest, but VoC initiatives that only monitor – initiatives that are not used in actual attempts to improve products, services and support – fall far short of their potential. VoC can help analysts discover the root causes of issues in order to enable them to develop and test process improvements. VoC can help marketers evaluate campaign effectiveness. And it can help with acquisition of new customers and retention of existing customers. The key in all these cases, the key to turning findings into action, is predictive modeling and integration into customer-facing systems and processes.

Seen in this light, VoC is part of the product-management, marketing and customer relationship management (CRM) toolboxes.

The VoC experience at NTT DoCoMo, Japan's leading mobile communications operator, illustrates how findings derived via text analytics can be used to create better products. DoCoMo VoC work identified unusual customer requests regarding use of cellular phones at international airports. Customers wish to use their cellular phones until take-off for overseas trips. And customers wish to use their same number abroad, despite having been too busy to make prior adaptive arrangements. These requests, mined from customer interactions, led DoCoMo to open shops at major airports in Japan. The popularity of these airport shops led the company to open multiple shops at some airports to handle customer demand.

On the CRM front, customer relationship management is an enterprise application that automates customer interactions – bringing into play a customer's transaction history as well as supplementary demographic and marketplace information. Advanced CRM utilizes predictive models that identify cross-sell and up-sell opportunities, customer-retention risks and conditions that are likely to lead to



desired goals. VoC text analytics can provide "lift" – it can enhance results – in all these facets of CRM.

VoC builds on traditional transaction-based CRM to create a fuller picture of customer and market attitudes about an organization's products and services.

Federico Cesconi, head of customer insight and retention at Cablecom, Switzerland's largest cable network operator, reports that his company's VoC program has led to a tremendous improvement in churn and customer rates. The company generated a predictive model from VoC data that showed that customers often decided to terminate their service months in advance of actual termination. Cablecom instituted a proactive customer engagement initiative where operators call customers determined by the company to be high churn risks. An assessment showed that this VoC-driven step would reverse customer-satisfaction problems in 55% of cases. Customers who rated their satisfaction at 0-6 on a ten-point scale could be boosted to an 8-10 point score.

Catherine Cardoso, associate insights manager at Unilever, is responsible for the Dove-brand pro.age campaign described in the FACTS AND OPINIONS section of this report. Cardoso says text analytics is a new methodology for Unilever but reports, "We were very pleased with the results and the depth of insight. The results were helpful beyond understanding reactions to our campaign. We also gained an understanding of what motivates people on discussion boards, which issues are most important to women in our target group, and how to create better products and messaging for them." Cardoso sees great potential for text analytics. She says, "We've been thinking about other ways to utilize this technology which would allow us to not only continue to listen to and understand our consumers, but to create a more real-time two-way communication."

VoC text analytics helps organizations determine what is important to the customer. This is likely different from what is important to the firm and may not be what business analysts expect. VoC efforts do not stop there. Findings are key to designing better products and services and to improving processes that affect the customer experience. And they are an essential ingredient in modeling customer behavior in order to better anticipate and respond to customer needs, to act and not just to study.



# **TECHNICAL INSIGHTS**

Text Analytics is exciting technology, opening up new applications and approaches to solving information needs and supporting decision making for an improved customer experience.

- Michael House, Maritz Research, Division Vice President

Traditional analytical approaches handle structured data well, but prior to the advent of text analytics, discerning customer voices in textual sources was laborious and expensive, providing only qualitative, subjective, small-sample findings. Text analytics additionally enables systematic and automatic text processing of individual cases and also the derivation of scientifically valid statistics from larger-scale data collection efforts. Moreover, prior to the advent of text analytics, it was difficult to integrate the analysis of these text-discovered, qualitative findings with quantitative analyses of numerical operational data. It was therefore quite difficult to deliver much talked about but seldom achieved *360-degree customer views*. Text analytics helps organizations overcome these limitations.

Text technologies – text analytics, natural-language processing (NLP), and analytically rooted search – are at a maturity stage where they have evolved from research tools into business solutions. While appeal is stronger than ever for lifesciences researchers and intelligence analysts and other long-time users, the center of market attention has shifted toward business applications such as customer relationship management (CRM), media and publishing, competitive intelligence, ediscovery and compliance, and financial analysis, and to Voice of the Customer initiatives that support sales, marketing, customer support and product management functions.

Text technologies extend established analytics applications by providing the ability to source information from documents and the Web. They simplify information access via natural-language (NL) interfaces, especially in the form of advanced search and NL data query. Service-oriented architecture and hosted and on-demand deployment options speed deployment time and lower entry costs by allowing business users to concentrate on business questions rather than on IT. The result is a core role for text analytics in VoC initiatives.

## **TECHNICAL CONTEXT**

Best-in-class companies are 55% more likely to have text analytics in place than the industry average.



#### - Aberdeen Group<sup>7</sup>

VoC efforts benefit hugely from text analytics, boosted by advances in integration of text technologies both with business intelligence (BI) and predictive-analytics tools and by adaptation to industry sectors such as hospitality, retail, health care and communications services. These advances answer usability concerns; they make the technologies accessible to end users. Hosted and on-demand text analytics and adaptation to service-oriented architecture (SOA) – Text Analytics as a Service – have further reduced barriers to adoption, creating flexibility for business focused end users and lowering startup costs.

Marketplace indicators are uniformly positive. Sales are booming at established technology and service firms. A steady pace of acquisitions of text analytics companies by larger firms demonstrates the strategic value of the technologies. And the market is steadily producing new and innovative products and services, extending VoC to Voice of the Market and Voice of the Patient concepts that cover competitive intelligence, health-services delivery and similar fields.

## **TEXT TECHNOLOGY BASICS<sup>8</sup>**

The term *text analytics* describes a set of linguistic, lexical, pattern recognition, extraction, tagging/structuring, visualization and predictive techniques. The term also describes processes that apply these techniques, whether independently or in conjunction with query and analysis of fielded, numerical data, to solve business problems. These techniques and processes discover and present knowledge – facts, business rules and relationships – that had been locked in textual form, impenetrable to automated processing.

Text analytics starts with document acquisition, either targeted retrieval of all material identified by a search or blanket intake of email, web pages, scientific papers, corporate reports, news articles and the like. The next step is typically linguistic processing: determining sentence and phrase boundaries, stemming words, determining parts of speech. This step is followed by tagging and extraction of features – entities and their attributes, terms, concepts, sentiments and relationships – with some form of term normalization and use of lexical analysis to provide frequency counts and the like. Use of taxonomies, lexicons and gazetteers, and machine-learning techniques facilitates this work.

Text-mining tools annotate, extract and analyze associations among identified entities and concepts and the documents that contain them. They create categories or they may apply existing taxonomies – hierarchical knowledge representations – to classify documents; data extracted to databases may be analyzed via business

<sup>&</sup>lt;sup>7</sup> Alan See and David Hatch, "Data Management 2.0: Making Sense of Unstructured Data," Aberdeen Group, July 2007, *http://www.ql2.com/news/articles/aberdeengroup\_datamgmt2\_0707.pdf*.

<sup>&</sup>lt;sup>8</sup> Excerpted from Seth Grimes, "Text Technologies in the Mainstream," *http://altaplana.com/TextTechnologiesintheMainstream.pdf*.



intelligence (BI), data mining and visualization. They apply statistical techniques to cluster documents according to discovered characteristics. Lastly, they deliver both interactive exploratory capabilities and hooks to allow classification to be embedded in applications to add automated text processing.

The ability to stem words, identify phrases, and extract terms and entities is shared in degrees by search tools, which are, however, built for document retrieval rather than analysis and exploration of document sets and their contents. Information extraction, statistical analysis, visualization – none of these functions is present in typical search or content management offerings. Knowledge discovery – pattern recognition – via application of linguistic, statistical and machine-learning techniques, and via data mining and visualization, is a key differentiator of text analytics from those latter technologies.

Because text analytics looks at document sets and identifies inter-document relationships, it supplies context that enables far greater relevance in search results than is provided by search tools. Contextual relevance – the ability to apply domain knowledge to match patterns and cluster results – is a second key technology differentiator. Lastly, text analytics tools can be embedded in applications that produce and consume significant amounts of textual data and often pose real-time operational demands. Content management and enterprise search tools do not offer the same potential for operational integration.

## **"UNSTRUCTURED DATA," BI AND PREDICTIVE ANALYTICS**

Early adopters have deployed [text analytics] in conjunction with their data mining and BI solutions. In fact, the vast majority [of 2007 survey respondents] stated that they were extracting text and combining it with structured information for use with specific data mining tools or BI solutions. <sup>9</sup>

#### – Fern Halper, Hurwitz & Associates

Text analytics is an answer to the "unstructured data" challenge. The scope of the challenge is expressed in the truism that 80 percent of enterprise information originates and is locked in "unstructured" form. How complete can an organization's business intelligence (BI) or predictive-analytics program – or its Voice of the Customer initiative – be if it does not accommodate and exploit a majority of business-relevant information?<sup>10</sup> Faisal Mushtaq, CTO of media and market intelligence solution provider Biz360, goes so far as to say that "only a solution that

<sup>9</sup> Fern Halper, Hurwitz & Associates, "Text Analytics: The Road to Understanding Your Company's Unstructured Information," 2007.

<sup>10</sup> Seth Grimes, "Text Technologies in the Mainstream," April 2008, http://altaplana.com/TextTechnologiesintheMainstream.pdf



leverages a combination of information extraction, data mining and business intelligence technologies can deliver true actionable intelligence."

The "unstructured data" challenge has been recognized for decades. Consider the first recorded definition of BI, which appeared in an October 1958 *IBM Journal* article by H.P. Luhn, "A Business Intelligence System<sup>11</sup>," that describes a system that will:

...utilize data-processing machines for auto-abstracting and auto-encoding of documents and for creating interest profiles for each of the "action points" in an organization. Both incoming and internally generated documents are automatically abstracted, characterized by a word pattern, and sent automatically to appropriate action points.

We see that the earliest BI focus was on text – on extraction, categorization and classification – rather than on crunching numerical data! The reason BI refocused on numerical data is obvious: organizations of all sizes harnessed computers to run their businesses, generating volumes of fielded transactional information that was (and is) directly tied to business operations. Database-managed operational information was similarly the go-to source for VoC effort for similar reasons, supplemented by manual analysis of textual information and by qualitative research tools such as focus groups.

With the availability of text analysis tools, large-scale, automated analysis of textual sources is finally attainable, just in time to refocus the analyses on the huge volumes of text on the Internet and within the enterprise. This shift has occurred for BI, for predictive analytics and for VoC efforts.

BI and predictive analytics vendors have taken a variety of steps to build text analysis capabilities into their tools. Some have extended ETL (extract, transform, load) capabilities to "unstructured" sources. Some have extended BI interfaces and predictive analysis workbenches to directly access text. Some have created partnerships with text analytics vendors, have acquired text analytics companies or prioritized in-house development of text-analysis capabilities.

BI and predictive-analytics tools are widely used in the enterprise. Extensions and enhancements that enable them to access textual source mean that business analysts can now apply familiar software tools to encompass VoC analyses and also that marketers, product managers and customer-support managers have a much wider set of tool choices and methods available than previously for VoC analyses.

Conventional BI and predictive analytics applications, for financial planning, sales analysis and the like, work off factual data. What differentiates VoC analyses is the need to tackle attitudinal information: sentiments and opinion. The FACTS AND OPINIONS section of this report describes the difference between facts and opinions. Technical approaches to dealing with sentiment merit a closer look.

<sup>11</sup> http://www.research.ibm.com/journal/rd/024/ibmrd0204H.pdf



#### SENTIMENT ANALYSIS

Sentiment analysis is one of the most exciting applications of text analytics today. Because sentiment – with opinions and other attitudinal information – is central to human communications, it is also of key importance for Voice of the Customer study.

The first step is to discern sentiment, which (per FACTS AND OPINIONS section) may be part of an opinion statement or may be implied rather than directly stated. For instance, "a Prius averages 46 mpg while a Hummer H3 averages 15" contains two factual statements, but the juxtaposition implies a negative view of the Hummer and a positive attitude toward the Prius. However, sentiment exists not only at the phrase level; sentiment can be attached at the document level as well, e.g., in the case of a positive movie review.

Sentiment analysis, technically, comes down to a classification and informationextraction problem, and extension of techniques applied for classification by topic and for entity and fact extraction but with a much higher degree of complexity involved.

#### LANGUAGE COMPLEXITY

Sentiment analysis is particularly challenging because of the huge variability and subtlety of spoken and written language. Meaning that humans readily grasp from context is very difficult for computers to detect. How can software reliably discern facts and feelings in light of not only abbreviations, bad spelling and fractured grammar, but also sarcasm, irony, slang, idiom and, well, personality? How is a computer to understand when a high school senior doesn't grasp language basics per the excerpt blog from a Language Log blog posting<sup>12</sup>:

Sridhar Srinivasan asks about an amazon.com review (emphasis added):

"...After seeing the reviews, I bought this book in new condition at a really cheap price. I couldn't be less satisfied. The book explains all details in clear detail, so that even talented high school students could understand the text. The images that accompany the text greatly reinforce the main ideas...."

Sridhar observes that the author seems to mean "I couldn't be more satisfied," and wondered whether there's any connection between this and the idiom "I could care less," that is used to mean "I couldn't care less."

#### SENTIMENT COMPONENTS

Another example, taken verbatim from Dell's *IdeaStorm.com*, illustrates key sentiment concepts, and it also shows a wide set of issues. Start with misspellings

<sup>&</sup>lt;sup>12</sup> Mark Liberman, "Ask Language Log: more or less," *http://languagelog.ldc.upenn.edu/nll/?p=290* 



and irregular spacing and punctuation, which can make it difficult for software to correctly identify words and discern sentence and word boundaries:

Dell really... REALLY need to stop overcharging... and when i say overcharing... i mean atleast double what you would pay to pick up the ram yourself.

There's irregular capitalization: lower-case "i" where a capital letter is called for and "ram" rather than the conventional "RAM" as an abbreviation for random-access memory. The other capitalization that stands out is "REALLY," and that's a first sentiment-related concept, use of capitalization that indicates intensity of feeling.

This posting is about overcharging, and it's about RAM. These two topics would be picked up by looking for those two terms in a dictionary or perhaps in a taxonomy – a hierarchical organization of terms and concepts – that would position RAM as a hardware component and "overcharging" as a negative sentiment indicator.

According to researchers Livia Polanyi and Annie Zaenen, "The most salient clues about attitude are provided by the lexical choice of the writer." <sup>13</sup> Lexical choices: those are words. Boost, benefit and brave indicate positive valence – that is, tone or polarity – while *conspire, catastrophe* and *cowardly* are negative.

It is dangerous, however, to judge sentiment only by the presence of valence words. Throw in a negator such as *not* or *never* and you flip the valence. Intensifiers – for instance, very and most (and all-uppercase "shouting" as seen in the Dell example) – indicate the strength of the sentiment expressed. Modal operators such as *might, could*, and *should* distinguish hypothetical from real situations and weaken intensity, as in Polanyi's and Zaenen's example sentence "If Mary were a terrible person, she would be mean to her dogs." Other, "presuppositional" terms such as *barely* and *even* similarly relate what the speaker/writer observes to his or her expectations. They can also help us distinguish subjective statements from objective ones.

#### **HYBRID APPROACHES**

A variety of approaches can be used to tease meaning from textual sources. Statistical techniques that focus on pattern recognition sit at one end of the spectrum with linguistic technologies, tailored to particular languages and discourse (subjectmatter) domains at the other. According to Faisal Mushtaq of Biz360, "No single technology or technique works the best. Automated analysis of unstructured text poses unique technology challenges requiring an interdisciplinary approach to text analysis. A good solution is a combination of the 'right' technologies to solve a real/immediate customer problem."

A particular hybrid approach can lead to more accurate sentiment analysis for surveys and other sources that contain both fielded (usually numeric) information and free text. For instance, stars associated with Internet Movie DB comments hint at

<sup>&</sup>lt;sup>13</sup> "Contextual Valence Shifters," in *Computing Attitude and Affect in Text: Theory and Applications*, James G. Shanahan, Yan Qu, Janyce Wiebe, editors, 2006.



polarity, at the direction of overall sentiment. One *Alvin and the Chipmunks* reviewer gave the movie 8 stars out of 10. It is likely that his sentiments captured in the text were generally positive and moderately forcefully held. It's not surprising that a 5/10 review has the title, "A huge disappointment for fans of this memorable series" and 10/10 is coupled with "I just LOVED IT!" (Note the "shouting" indicating sentiment intensity.) Similarly, a hotel guest who chose a Fair rating in a satisfaction survey is likely to have posted more complaints than praise in free-text response fields.

There are further challenges. For instance, a particular movie review may contain opinions of various polarities – some positive, some negative and some neutral – and intensities. How do you decide the overall sentiment of the review and similarly understand the aggregate picture, the voice of the market rather than just of individuals? Can you discover relationships between sentiments and the characteristics of the people who expressed them as well trends over time and how opinions propagate through social networks? Can you forecast quantities like boxoffice receipts from opinions extracted from movie reviews? These analytical steps are the province of traditional data mining and descriptive statistics, which can be (and is being) applied to extracted attitudinal information.



## **DEPLOYMENT OPTIONS**

Voice of the Customer text analytics users may choose from a number of deployment options:

- Software installation
- Outsourcing
- Information-/research-service subscription
- Hosted/on-demand

While software installation is the conventional approach – it is well understood and therefore won't be discussed here – hosted/on-demand text analytics is the preferred service-delivery model for a rapidly growing segment of text analytics users. This approach is a variant of the increasingly popular software as a service (SaaS) model. It comes in a number of flavors. With hosting, you contract another organization to install and maintain software for your use. With on-demand, you pay for only the services you use. Both styles of SaaS may be delivered via web services interfaces or via more-traditional integration approaches (that is, where the client in effect outsources software automation of selected business processes).

The outsourcing option applies particularly for survey analysis, which has long been handled by specialized survey research firms.

Subscription to an information or research service is a particularly attractive option for organizations doing competitive- and market-intelligence research using web sources such as user forums, blogs, news articles and the like. Providers enjoy economies of scale by being able to sell the same or similar services and information to multiple clients, potentially leading to cost savings for users.

Outsourcing and adoption of "as a service" text analytics and subscription models are a response to efficiency, cost and rapid-delivery business drivers. Users can focus on business goals rather than on IT. They can get started quickly, relying on provider expertise, without incurring full software-licensing and training costs. SaaS adoption and service subscriptions therefore represent a change in how software is used and paid for. Along with the ability to outsource, they are key enablers accelerating the adoption of text analytics by organizations that would otherwise lack the resources to license, design and manage in-house solutions.

A number of factors influence an organization's choice. They include:

- Project timing
- Scope of the initiative: users and information sources
- Analyst capabilities
- IT capacity



• Budget and funding

The tendency is still for enterprises to house larger scale projects in house, that is, projects that extend to large numbers of users or that cross departmental boundaries. And organizations are still reluctant to send information from operational systems. Organizations that are just getting started with Voice of the Customer text analytics, with limited deployments and information sources and enterprise integration, will benefit from the widest variety of options.

## **MARKET OUTLOOK**

The \$250 million 2007 text analytics market – one can estimate annual growth rate at over 25% – includes product vendors, technology suppliers who offer linguistic and natural-language technologies, industry vertical solutions and an allotment of revenue attributable to text analytics at broad-market analytics vendors.

Text analytics is doing very well in the areas where it first emerged – life sciences and intelligence – but the biggest growth areas are customer and market intelligence: as a tool that lets companies directly understand stakeholder views – the Voice of the Customer – about their products and services and those of their competitors. On the technical side, there is huge interest in sentiment analysis and opinion mining and, in particular, on mining web information sources.

On the technology front, growth at new and established companies integrating text with BI and with predictive analytics, and in hosted / "as a service" offerings, has been especially strong. Also, several vendors focus on delivery of competitive and market intelligence as an information service, where text analytics provides a key back-end enabling capability.

Overall, the commercial landscape reveals both stability and opportunity, byproducts of the rapid pace of growth in adoption and interest. Market development will continue accordingly with further consolidation, clarification, alliances and new entrants.



# **STUDY: VOC TEXT ANALYTICS BEST PRACTICES**

This section reports the findings of a study designed to investigate best practices for the implementation of Voice of the Customer text analytics initiatives. Report author Seth Grimes has been looking at text analytics best practices for several years and designed and conducted the present study.

## THE VALUE OF BEST PRACTICES

The 2006 white paper "Finding Value in Text Analytics<sup>14</sup>" describes best practices as follows:

"Best practices" are generalized principles, techniques, and methodologies derived from theory, academic and industrial research, direct experience, and case studies that guide those who apply them in assessing requirements, evaluating options, and devising strategies for the implementation of new technologies. The application of best practices recognizes that while each enterprise has unique circumstances and requirements, there are very broad areas of commonality among organizations with similar purposes and needs that we should seek to exploit."

The current best practices study avoids covering points that are, in reality, common to many IT domains and particularly to BI and predictive analytics. Further, the intention was to be very practically focused rather than theoretical. The study therefore centers on lessons solicited from individuals with experience applying VoC text analytics to real-world business problems at their organizations and from a number of vendor representatives and industry analysts.

## **SURVEY FINDINGS**

For this study, the author had free-form discussions with five industry analysts, six end users and a variety of vendor executives. The author additionally conducted a formal, small-sample survey of VoC text analytics practitioners – end users and consultants – that attracted seventeen responses, not a bad number given the relative newness of the application of text analytics to Voice of the Customer challenges.

#### LENGTH OF RESPONDENT EXPERIENCE

A plurality of respondents, 35%, report more than two years of experience using text analytics for Voice of the Customer or related needs (e.g., Customer Experience Management). The "still evaluating/not using" category covered 31% of respondents, 15% were users of "less than 6 months," 8% "6 months to less than one

<sup>&</sup>lt;sup>14</sup> Seth Grimes, "Finding Value in Text Analytics," June 2006, http://altaplana.com/FindingValueInTextAnalytics.pdf.



year," and 12% "one year to less than two years."

#### SATISFACTION WITH VOC TEXT ANALYTICS

All current users responded to request to rate experience/satisfaction (E/S) with VoC text analytics. On a seven-category scale:

- None rated their E/S in the three categories below neutral
- 18% of respondents were neutral
- 47% were satisfied
- 24% were very satisfied
- 12% were completely satisfied



#### INFORMATION ANALYZED: CURRENT AND PLANNED

Of the 70% of respondents who are currently engaged in VoC text analytics:

- 92% are analyzing contact-center notes,
- 68% surveys,
- 63% contact-center notes or transcripts,
- 53% website feedback, and

**47% email and correspondence.** These are "inside-the-firewall" sources that capture operational information and feedback shared with organizations by established customers and prospects.

Looking at "outside-the-firewall" sources, online forums top the list, analyzed by 47% of current users, followed by blogs at 42%.

News articles, point-of-service notes or transcripts, review sites and chat and/or instant-messaging text are tapped by a minority of current user respondents. While chat and/or instant-messaging text is analyzed by only 16%, that information category tops the list of sources respondents plan to harvest with 26% reporting plans. Smaller proportions of respondents plan to harvest the other information sources listed as choices.

The following chart presents current and planned information analyzed as a proportion of all respondents, current VoC text-analytics users or not:





#### **MEASURING ROI**

When asked how they measure ROI, return on investment:

- 58% of current-user respondents chose higher satisfaction ratings,
- 53% chose fewer issues reported and/or service complaints,
- 53% chose higher customer retention/lower churn, and
- 42% chose increased sales to existing customers.

Smaller, minority proportions chose fewer issues reported and/or service complaints, improved new customer acquisition, lower average cost of sales, new and existing customers.

Regarding ROI targets achieved by organizations that prioritized them, success rates reported were:

- 100% improved new-customer acquisition
- 63% increased sales to existing customers
- 60% fewer issues reported and/or service complaints
- 57% qualitative observation

# Higher customer retention/lower churn and higher customer satisfaction rates top the list of planned ROI measurements.

The following is chart shows all responses:





#### **SOLUTION PROVIDERS**

We offer response statistics in tabular form for the multi-select question: What should a prospect look for in a solution provider?

Response	Percent
Deep sentiment/opinion extraction	84%
Ability to use specialized dictionaries or taxonomies	74%
Broad information extraction capability	63%
Adaptation for particular sectors (e.g., hospitality, retail, health care, communications)	53%
BI (business intelligence) integration	53%
Predictive-analytics integration	53%
Support for multiple languages	48%
Ability to create custom workflows	37%
Low cost	32%
Hosted or "as a service" option	32%
Specialized VoC analysis interface	21%



## **ADVICE TO END USERS**

As befits a survey on application of text analytics, this one included a question that allows for free-form response: What three bits of advice do you have for someone embarking on VoC text analytics? This was also the central question in the analyst, end-user, and vendor that formed the other component of the study.

Advice responses lend themselves to clustering in four categories:

- Business Goals
- Evaluation/Pilot
- Implementation/Startup
- Operational Principles

In the Business Goals category, we have the always useful advice,

• Clearly define initiatives based on strategic objectives.

All the rest of the advice in this category presupposed that strategic objectives do, in fact, justify a VoC text analytics program. A number of items address making the case, that:

- Unless there is a taste for innovation in the organization, it is hard to fund a project.
- You should understand your business. Are they ready?
- Is your business ready to take action on the data?
- Understand that you're going to run into opposition.
- Try to impress on the people you're talking to that there's urgency for change.
- You need an in-house team championing the effort.

A number of advice points fall into the Evaluation/Pilot category:

- Pilot with open source tools to learn.
- Study the TA industry. Develop detailed documentation of your needs.
- Do the homework. Go through benchmarking 3-4 tools in the domain. Look at extra services: initiation, end-user training.
- Do a proof of concept evaluation.



- Understand the problem you're solving and the audience you're solving it for.
- You want to look for a tool that you can use across different constituencies.
- Adherence to UIMA the Unstructured Information Management Architecture is important.

The **Implementation/Start-up** category formed the largest cluster of bestpractices/advice responses. Addressing that category, several current users and analysts advise a progressive approach:

- Think big, start small, use an incremental approach.
- Start in an area where value can be created quickly and expand from there.
- Win fast and win often.
- Those quick wins mean a lot. It took us a while to understand the technology.
- Don't go after the biggest problem first.
- Collect requirements from all potential internal customers and ensure they are included when building out reporting.
- Provide simple access to data and analysis.

One advisor suggests looking for services while two emphasize proper training and one stresses "Have the right tool in place." Two focus on staff:

- Dedicate someone to the job.
- I'll emphasize getting a core team together that knows the technology and can support it in the organization.

Two focus on startup time:

- Don't underestimate the time/effort needed to build libraries.
- It takes time to develop the screening rules to make text analytics really useful.

And one user advises:

• Keep the project off executive radar until you know if it is meaningful in your business.

While another explains:

• I didn't tell anybody what I did. I just did it.



The Operational Principles category included three accuracy points:

- Users should demand sensitivity of analysis.
- Don't expect the high levels of accuracy we experience in the quant side.
- Accuracy is less critical to VoC than most uses of Text Analytics (80%+ good enough).

These points are inconsistent. The first says that accuracy is important, the third that it is not critical for VoC work, and the second that (critical or not) you're not going to experience the levels you've come to expect in numbers-rooted analyses.

Otherwise, experience users advise, in the Operational Principles category:

- Text analytics is not push button. It takes work.
- Be open to possibilities. VoC text analytics can be transformative in a whole variety of ways.
- Don't try to mix initiatives, look for synergies and relations but keep scope within reach.
- Integrate both structured and unstructured data.
- [Forget ROI.] We need to get a return on time [spent on manual processes and on implementing automated text analytics].



# BUSINESS OBJECTS VOICE OF THE CUSTOMER SOLUTION PROFILE

BusinessObjects XI 3.0 with Text Analysis and Intelligent Search is a softwareplatform solution providing a way to perform buzz analysis, Voice of the Customer (VoC) analysis, root-cause analysis and other comprehensive business intelligence (BI) tasks. The solution operates on information extracted from the gamut of textual sources in conjunction with data from conventional BI sources, in particular from transactional and operational systems. The solution is designed for end-user usability by business directors, managers and analysts via comprehensive integration with Business Objects' suite of market-leading business intelligence (BI) tools.

#### **COMPANY OVERVIEW**

The mission of Business Objects, an SAP company, is to transform the way the world works. The company characterizes itself as not just a business intelligence company, but additionally as a company that is "in the business of helping companies become more intelligent." The company counts more than 43,000 customers worldwide – including over 80% of the Fortune 500 – and a network of more than 3,000 partners and resellers. Business Objects acquired deep linguistic, text analysis and visualization capabilities with the 2007 acquisition of Inxight Software. Inxight was itself founded in 1997 as a spin-off of Xerox's PARC and was widely recognized as a leading provider of enterprise software solutions for information discovery. The companies and their technologies are now fully integrated. Business Objects explains that "this combination enables full spectrum business intelligence, enabling customers to see at-a-glance, explore and discover hidden relationships, trends and patterns, turning large volumes of information into actionable intelligence for applications in Voice of the Customer analysis, claims analysis, counter-terrorism and beyond."

## **PRODUCT SET**

**Business Objects'** Voice of the Customer (VoC) solution relies on the company's end-user BI tools and interfaces – Crystal Reports, Xcelsius and BusinessObjects XI 3.0 – interoperating with the company's linguistic and text analysis technologies. Those technologies "complement BI with the wisdom in text sources by 'reading' text documents in a variety of file formats and in more than 30 major languages."



Voice of the Customer		Competitor Mentions- Corvette		
VOICE OF	the customer		Survey ID Comment	
$\frown$			10187	Great car but seriously considering Corvette instead/
			10518	Wish handling was more dynamic like Corvette.
	Neutral Sentiment Positive Sentiment Negative Sentiment		10968	We will be switching over our entire company fleet to Corvette - I hate you
			11437	Swiftmobile's tires aren't as good as a Corvette's.
			11942	Cost of Supercar's maintenance is too high - selling it and buying a Corvet
			1198	Corvette has more power than the Supercar.
6	Chartegia Customera		12027 Wish you offered free driver's training like Corvette does.	
-			12184 First of all your car does not even offer the bare minimum features that Co	
<b>D</b>			14195	Of course, I expected the Wondertruck to have more practical cargo capac
Products	Competitors	Region	14527	I should never have sold my Corvette to buy this piece of junk.
			2480	Your service is nowhere near as good as when I used to have a Corvette.
USA Good Service Select a Category Service Problems		2824 I was going to buy a Corvette, but my husband liked th	I was going to buy a Corvette, but my husband liked the orange color of th	
		3230	It's interesting to note that the Corvette usually gets better ratings than th	
		3559	I would have chosen to buy a Corvette, but it cost 10K more.	
MM	Produc	t Problems	3588	Needs better seats - like those in a Corvette.
(TL F	Good S	Service	200	I like the Supercar, but to compare it to a Convette is ridiculous
			Business Objects an SAP company	

BusinessObjects Text Analysis software accesses over 200 different document types and text in languages such as French, German, Spanish, Traditional and Simplified Chinese and a variety of other Asian, European and Middle Eastern languages.

Extraction, categorization and summarization capabilities identify the concepts, sentiments, people, organizations, places and other information in the text sources. The software converts information from those sources into structured data for unified analysis of text and data.



Business Objects' language-processing components, deployed in VoC and other applications as part of BusinessObjects XI 3.0, include:

- LinguistX Platform: A multi-lingual natural language processing engine that allows for the integration of text analysis into search engines, data mining applications, security applications and legal discovery, storage and routing tools.
- ThingFinder: Extract, tag, and index more than 35 types of named entities

Copyright © 2008 Powell Media and Alta Plana Corporation All Rights Reserved

**b**•eye Business Intelligence Network™ The Global Vision for BI and Beyond

(such as people, organizations, places, addresses and dates), relations and events in all major languages for information discovery applications like publishing, legal discovery, categorization, link analysis, data and text mining, smart storage, security, business intelligence and more. The 35 outof-the box entities can be supplemented with application-specific, list-based or pattern-based entities, relationships and events.

• **Categorizer:** Identifies what topic or topics a document is about, in relation to a business taxonomy so that information can be routed to the right department.

The company also provides a federated search tool as part of BusinessObjects XI 3.0:

• **Business Objects Intelligent Search** allows you to search ("crawl") all sources inside and outside an organization, including public websites, deep websites (e.g., the publicly accessible Patent Databases or SEC DBs) and internal systems, as well as secure and password-protected subscription sources from a single secure search box (federated search). Results are re-ranked for relevance and clustered by people, companies and other concepts in full text of results. Automated alerts watch for new items containing terms of interest or web page changes.





The company describes characteristics of organizations that can benefit from a BI solution that incorporates text analytics. These organizations, according to Business Objects, share the challenges that:

- Information critical to decision making is hidden in text documents.
- Utilizing that information would require someone to read the documents a lengthy, expensive and error-prone process.
- The volume of text content is overwhelming for available resources.
- Accessing text information in a timely fashion would give a competitive advantage.
- They can describe what they are looking to understand from within their content and they can provide sample documents.

The next version of BusinessObjects XI 3.0 with Text Analysis and Intelligent Search is due out in the second half of 2008. It is expected to feature expanded outof-the-box capabilities for VoC analysis and more sample dashboards. Editing and validation capabilities will also be expanded.

## **VOICE OF THE CUSTOMER APPLICATION**

Business Objects lays claim to the number one spot in text analytics, presumably on the basis of the text analytics products' Inxight heritage. The Voice of the Customer (VoC) product is new, however, it brings well established technology to bear to answer a variety of critical business questions in a number of frequently encountered scenarios enumerated by the company:

#### Brand Management (Marketing Scenario)

- What are people saying online about my brand, product or service?
- What are people saying about my competition?
- What do consumers want that's not available today?
- Where must we do damage control?

#### Marketing Campaign Direction (Marketing Scenario)

- How well are we meeting the needs of particular market segments?
- Should we change our marketing or product offering strategy for a segment?
- How successful is our current marketing campaign that is, how much buzz has it generated?
- What do existing customers like about us and what do potential customers like about our competition?



• Why didn't the campaign generate the expected revenue targets?

#### **Business Process Improvement (Customer Assurance Scenario)**

- How easy is it to do business with us?
- What are our customers' pain points in doing business with us?
- Where can we improve our process?

#### Customer Retention (Marketing Scenario)

- How satisfied are my customers?
- What are customers most unhappy about?
- What corrective action would make up for delay or disruption of service?
- How can we improve our service and product to avoid these issues?
- Why are we losing customers?

#### Performance Evaluation and Improvement (Operations, Support, Assurance)

- What percentage of my customers had issues that were (not) satisfactorily resolved by a customer representative?
- When were we able to make an irate customer happy?
- How can we improve the performance of our customer representatives?

#### Product Offering Enhancements (Customer Support, Product Groups, Marketing)

- What defects are customers most upset about?
- What immediate corrective action is required to handle serious defects?
- What enhancements do customers want most?
- Are there any paradigm-shifting opportunities waiting to be seized upon? That is, what do potential customers want that currently is not available?

#### Up-Sell, Cross-Sell and Account Management (Sales)

- What existing products should I tell this customer about?
- What issues are my strategic accounts facing?
- Which representatives are performing best as perceived by customers?



The company envisions the BusinessObjects Text Analysis VoC solution as being used by staff charged with meeting a variety of business needs:

- The **Director of Public Relations and Marketing Analysts** would use the solution to better understand online content regarding consumer perception of brand, competition and the market in general such as what consumers like and dislike about your organization's brand, product or service, what they like about the competition, and what consumers want that's not currently available. This data can help the director and analysts both in measuring the buzz generated by marketing campaigns and in monitoring change over time to see trends such as how the market reacts to campaigns and market events.
- A **Marketing Director** would analyze customer feedback captured by corporate CRM systems or call-center logs, as well as in customer survey logs and the .com feedback channel, to evaluate marketing campaigns and to plan future campaigns. For this user, VoC illuminates how well the needs of particular market segments are being met, what existing customers like about the company and the competition, and why a campaign did or didn't generate the expected revenue targets. Over time, these results can be used for trending and monitoring.
- A **Customer Assurance Director** would analyze customer feedback from various sources to identify problems with ease of doing business and areas for business process improvement. Such information is useful in defining strategic customer service improvement programs for various internal groups.
- The **Customer Loyalty Director** uses BusinessObjects VoC text analysis to analyze freeform text comments from surveys explaining customer satisfaction statistics and answering questions that reveal what customers are most unhappy about and why customers are being lost.
- The VP of Quality, VP of Products and Product Managers get reports of product usage, when products fail, most frequent customer enhancement requests and high-priority suggestions for quality improvements. Tracking such detail over time gives insight into the evolution of an organization's product issues and quality.
- Sales Managers use BusinessObjects Text Analysis VoC to mine CRM systems for up-sell and cross-sell opportunities such as when an existing customer mentions wanting an additional service, determining whether individuals at strategic accounts are having issues that may stand in the way of important deals, and understanding who the top performers are.

Business Objects elaborates that several application areas are related to VoC, but are different enough that they require a separate solution. Some examples include a "know your customer" solution for fraud analytics, credit companies and insurance, where analysts need to assess the risk related to customers. The goal here is similar to an intelligence solution – know your enemy – but differs in domain. Voice of the

**b**•eye Business Intelligence Network™ The Global Vision for BI and Beyond

employee, as expressed in employee satisfaction surveys, overlaps more with the VoC capabilities. Text analysis is also deployed by customers for case management, claims analysis and buzz analysis applications.



The company explains that outside the VoC arena, customers are using Business Objects' text analytics software for a diverse array of applications that include counter-terrorism, root-cause analysis, editorial automation, information visualization, advanced research portals, expertise identification, investigative applications and email risk assessment. Most of these applications are deployed across different industries.

#### DEPLOYMENT

Business Objects licenses BusinessObjects XI 3.0 with Text Analysis for in-house installation and use. User applications can additionally access the system through a web-services interface.

Copyright © 2008 Powell Media and Alta Plana Corporation All Rights Reserved



Buzz Analysis capabilities, in particular, are presented in partnership with Newstin.

The company also offers on-demand media monitoring services.

The company offers a workflow for implementing a BusinessObjects Voice of the Customer (VoC) text analytics solution as illustrated in Table 1.

Table 1. Workflow Outline for Implementing BusinessObjects Text Analysis for VoC Needs						
Step	Component	Status				
1. Ingest source data	Source adapters, crawler, BusinessObjects Data Integrator	Required				
2. Filter data to relevant set or sort by business unit	Text Analysis categorization	Optional				
3. Extract relevant data	Text Analysis extraction, leveraging predefined language modules, solutions packs, and customer dictionaries	Required				
4. Output to database	Oracle, Microsoft SQL Server, and others	Required				
5. Integrate with structured data	BusinessObjects Enterprise and Universes	Optional but recommended				
6. Display and analyze	BusinessObjects Web Intelligence, Xcelsius, Crystal Reports, or other dashboards	Required				

## **SUMMARY**

Business Objects is the market-leading business intelligence vendor with an exemplary track record meeting enterprise business-information requirements. The BusinessObjects XI 3.0 with Text Analysis and Intelligent Search software platform, via a solution set tailored to corporate Voice of the Customer (VoC) analysis needs, is an excellent choice for current Business Objects customers and partners and for other organizations that wish to implement a BI-focused VoC solution.



# **IBM VOICE OF THE CUSTOMER SOLUTION PROFILE**

IBM OmniFind Analytics Edition is an advanced solution for analysis and mining of unstructured information. The software is designed to help users find, analyze, mine and ultimately maximize the value of unstructured information, regardless of storage location or format. Visualization tools allow organizations to uncover new insights, patterns and trends in text-sourced information that can be used to manage risk, reduce costs and maximize customer service opportunities.

OmniFind Analytics Edition recently garnered the SSPA's "Innovation in Voice of Customer Award."

#### **COMPANY OVERVIEW**

IBM characterizes itself as "a proven leader in delivering on demand information technology and services, and has extensive experience in designing and implementing end-to-end solutions to solve real business problems – around the world and across many industries."

IBM traces its history to products invented the 1880s. The company has worked continuously since then to deliver hardware, software and service solutions that both automate and optimize everyday business processes, the latter through the application of leading-edge analytical methods. Referencing Hans Peter Luhn's seminal <u>1958 *IBM Journal* paper</u>, "A Business Intelligence System," – Luhn writes of an "intelligence system [that] will utilize data-processing machines for auto-abstracting and auto-encoding of documents. Both incoming and internally generated documents are automatically abstracted, characterized by a word pattern and sent automatically to appropriate action points." – IBM could lay a claim to having laid the systematic foundations for both text analytics and business intelligence. OmniFind Analytics Edition should be evaluated in light of IBM's very long involvement in application of these and other enterprise technologies to the gamut of real-world business problems.

IBM has made significant investments in text analytics research and services in recent decades with hundreds of staff engaged in developing new capabilities and over 200 patents in the space. The company has been able to transform research innovations into technologies that meet special challenges that arise in dealing with unstructured text. The company provides as an illustration its ability to leverage work by IBM Research to create analytics packages and user interface (UI) interaction metaphors that allow business users to extract actionable insights from text "out-of-the-box," with minimal upfront configuration or domain adaptation required.

Analyst firm Gartner named IBM a Leader in its Magic Quadrant for Information Access Technology, stating "IBM offers extremely rich content analytics considering its comparative newness to the market, a reflection of its depth in information access technology."



## **PRODUCT SET**

Today, IBM offers a complete information access portfolio, IBM Content Discovery, including the OmniFind family of products for search-based content access, the IBM Classification Module for auto-content-classification, and the award-winning IBM Content Analyzer (IBM OmniFind Analytics Edition) for delivering content-centric business intelligence. These offerings can be deployed stand-alone or combined with other products to enable solutions.

The company states that "enabling clients to tap their information assets is a key part of IBM's vision for information on demand: obtaining faster access to information across the enterprise helps enable people to make better decisions and optimize their businesses."

OmniFind Analytics Edition provides a horizontal text analytics and content-mining platform and provides application programming interfaces and a data model that enable straightforward integration with specific line-of-business applications. The integrated solution provides correlation analysis, trend analysis, delta analysis, automated alerts, drill-down navigation, semantic search and keyword search. IBM designed these advanced capabilities to reveal situations such as causes of customer dissatisfaction or detection of products that may require frequent repairs or a recall and track changes over time. Figure 1 depicts the OmniFind Analytics Edition user interface.

OmniFind Analytics Edition is based on the Unstructured Information Management Architecture (UIMA). UIMA provides an open framework and standard interfaces for creating and composing analytics capable of identifying and extracting the facts and relationships expressed in unstructured information. UIMA was created by IBM and released to open source in 2006 with the goal of accelerating adoption, usage and value-added innovation around the framework and interfaces. Development is managed as an Apache Incubator project. <u>OmniFind Analytics Edition harnesses</u> <u>UIMA to support building full-text and semantic search indexes</u> while <u>OmniFind</u> <u>Analytics Edition deploys UIMA for information extraction and text analysis</u>.

OmniFind Analytics Edition combines text analysis with associated structured detail, via optional integration with the IBM DB2 Warehouse, for richer insights. It is designed to be highly scalable, with end-user response times similar to keyword search. It provides a rich, out-of-the-box user interface for interactive analysis that is accessible and understandable by business users.

The current release of IBM OmniFind Analytics Edition is v8.4.1. With the planned August 2008 release of v8.4.2, the product will be renamed IBM Content Analyzer.





## **VOICE OF THE CUSTOMER APPLICATION**

IBM's Voice of the Customer (VOC) application is targeted to organizations that collect and/or analyze customer information. It is designed for individuals who manage and conduct customer-care, call-center, business-strategy, marketing, quality-control and business intelligence functions. The company explains that OmniFind Analytics Edition "provides an out-of-the box user interface (UI) enabling business users to explore and analyze customer information, combining the ability to select and drill down on structured and extracted unstructured dimensions; use search to define subsets of data based on unstructured content; and visualize data in list, time-series, distribution, multidimensional correlation, and document views."

"The typical user for OmniFind Analytics Edition is a line-of-business analyst who uses the solution to analyze business performance, locate support bottlenecks or service problems, and identify churn issues and/or new marketing opportunities. Insights gained from the use of OmniFind Analytics Edition are fed back into the business through the user organization's usual business analysis and optimization processes. OmniFind Analytics Edition also offers alerting capabilities that enable automatic detection of emerging problems and triggering of business processes to mitigate them – for example, automatically identifying customer churn candidates from VOC analysis."

The solution streamlines and advances the collection, review and analysis of

*Copyright* © 2008 *Powell Media and Alta Plana Corporation All Rights Reserved* 



information housed in customer-care applications. It enables comprehensive review and analysis of both structured and unstructured content – recorded customer calls, customer email, text from chat sessions, contracts, survey comments, etc. – to reveal valuable customer insights. OmniFind Analytics Edition is the first tool that provides this combined view.

IBM sees text analytics based VoC solutions as an emerging market, albeit one where business hurdles – for instance, helping prospective clients understand the business benefits of analytics-based VoC solutions and the projected return on investment (ROI) – are diminishing as organizations look for differentiation in customer service and customer insight. The solution was launched in the second quarter of 2007.

The solution additionally answers the demands of diverse functions such as customer churn alerting and online media analysis for corporate brand reputation monitoring. Outside the VoC realm, OmniFind Analytics Edition supports quality/warranty insight and quality early warning; compliance investigations and eDiscovery; contracts insight; and healthcare-payer coverage optimization. More generally, it supports delivery of content-centric business intelligence and content dashboards in enterprise content management deployments.

## DEPLOYMENT

IBM licenses OmniFind Analytics Edition for in-house installation and use. The company provides optional business and technical consulting and integration services for text analytics-reliant solutions for Voice of the Customer and the gamut of enterprise challenges. IBM Global Business Services has been the primary services partner for VoC solutions; however, the technology arm of the company is expanding its relationships with other system integrators as well.

Solution deployments may be complemented by optional use of the IBM Classification Module, which uses a more basic set of text analytics capabilities in conjunction with statistical modeling and traditional business rules to provide automatic, adaptive classification of content (including VOC-related content, such as customer surveys). IBM Classification Module and OmniFind Analytics Edition are often deployed together, with Classification Module performing a first-pass organization or prioritization of customer content (e.g., based on problem area, customer type, even sentiment level) to focus second-phase detailed analysis with OmniFind Analytics Edition.

Additionally, IBM content-analytics solutions can leverage the full power of IBM's information management infrastructures, including the company's market-leading enterprise content management platforms such as IBM FileNet P8 and DB2 data warehouse capabilities, to provide end-to-end information storage, management and insight solutions.

IBM recommends that users "focus on a solution that's easily usable by the key business stakeholders (e.g., analysts in the customer service organization) as it will enable rapid adoption and easier justification of ROI benefits. Ensure that any solution you choose integrates both structured and unstructured data, providing the



360° view that only the combination of the two can provide. Ensure as well that the solution does not require you to pre-identify or define issues to be identified in unstructured content, like call logs, as that will limit your insight to what you're expecting and prohibit serendipitous discovery of business optimization opportunities. And finally, deploy VOC in the context of an overall information management and insight strategy, avoiding siloed technology and point solutions that will not address future information management and insight needs across the organization."

IBM is currently piloting a managed (outsourced) services offering for VOC, but has not announced a timeline for general availability.

## SUMMARY

IBM's long history developing and deploying enterprise technologies and solutions for core business functions, including customer care, marketing, product design and quality needs that are part of Voice of the Customer initiatives, positions the company as an excellent choice as a VOC solution provider. The OmniFind platform provides essential, highly scalable integration and analytical infrastructure as well as bindings to data warehousing technologies that facilitate unified analysis of structured and unstructured information. In Analytics Edition, the company has designed a user interface for business analyst usability, free of the complications that often accompany interfaces designed for complex business intelligence, data warehouse or statistical analysis needs. Instead, the Analytics Edition UI provides for sophisticated linguistic analysis, information extraction and trend analysis linked specifically to the need to deliver insights into customer issues and business operations. It constitutes a strong solution for Voice of the Customer and other text analytics that require both high-performance, scalable analysis of unstructured information and integration of structured-unstructured analyses.



# **SPSS VOICE OF THE CUSTOMER SOLUTION PROFILE**

SPSS software collects and analyzes data about people's attitudes and behaviors to improve business outcomes. The company's Voice of the Customer (VoC) solution is an important application, designed for business end users, of the company's market-defining predictive analytics software solutions.

#### **COMPANY OVERVIEW**

SPSS is a predictive-analytics software company, the only company focused exclusively on predictive analytics needs, on the ability to understand, predict and act. The company's mission is to "drive the widespread use of data in decision making."

SPSS is expert in people data, in actions, attributes and attitudes. The company numbers more than 250,000 public sector, academic and commercial customers including the vast majority of the Fortune 1000: top global commercial banks, telecommunication services companies, retailers and market-research firms. In the public sector, SPSS counts as clients all major national governments, every branch of the U.S. military, and every U.S. state government; and they claim title to being the #1 statistical software on university campuses.

The company was founded in 1968 and operates internationally. The company maintains partnerships to deliver predictive capabilities to business intelligence (BI) providers including Business Objects, IBM, MicroStrategy and Cognos.

## **PRODUCT SET**

SPSS sells predictive analytic tools and solutions that help analysts and organizations find and acquire new customers, boost retention of profitable customers, expand sales, detect noncompliance and reduce risk. The technology harnesses the company's advanced mathematical and statistical expertise to extract predictive knowledge that makes business processes smarter and more adaptive. The company's predictive analytics capabilities, whether applied to text or numerical data or both, are delivered both through the SPSS's own user-facing applications and as technology that complements and extends business intelligence (BI) offerings from the company's partners.

SPSS Predictive Text Analytics extends the company's predictive analytics solutions to text, providing for integrated analytics of structured and unstructured information in the same models. The aim is to enable more reliable conclusions and more effective action, to:

• Uncover concepts and relationships that would otherwise be hidden in textual source.



- Boost the accuracy of predictive models, providing "lift" that enhances predictive effectiveness.
- Support both interactive and automated analysis and decision making.

SPSS's solution combines the natural language processing (NLP) capabilities of the company's LexiQuest text mining products with the advanced analytical capabilities of the Clementine data mining workbench. While Voice of the Customer (VoC) users will be most interested in Clementine and in SPSS Text Analysis for Surveys, they should be aware of the full set of SPSS text analytics product, namely:

- **Text Mining for Clementine**, which adds NLP-based information extraction to the Clementine data mining workbench. The extension allows Clementine to offer users unified classification, clustering and predictive modeling of numerical and text-sourced data with model deployment anywhere a Clementine model can be deployed.
- LexiQuest Mine, a tool for concept extraction, lexical analysis and relationship visualization that focuses on textual sources.
- LexiQuest Categorize, which provides classification capabilities that automate accurate text processing including sorting and routing, handling sources that include email, notes fields from call centers or other applications, and the spectrum of corporate reports and documents.
- SPSS Text Analysis for Surveys, a desktop product that is designed specifically to analyze text responses to open-ended survey questions. SPSS Text Analysis for Surveys provides built-in concept extraction and categorization and allows the user to integrate results with quantitative survey data for further analysis using either SPSS or Microsoft Excel.

Market researchers and analysts look to SPSS's Dimensions product line, designed to allow users to "capture and incorporate data about people's attitudes and preferences into their analytical decision making." SPSS explains that Dimensions "manages the entire research life cycle by easily authoring and conducting surveys quickly" and "consolidates research efforts, securing sensitive information and increasing organizational impact."

Clementine, including text-analytics capabilities, constitutes a key component of all SPSS predictive applications including those that support VoC needs. These applications include:

- <u>**PredictiveCallCenter**</u>, designed to improve customer intimacy and satisfaction.
- <u>PredictiveClaims</u>, designed to improve insurance-claim and fraud detection.
- <u>**PredictiveMarketing**</u>, designed to improve the effectiveness of marketing campaigns.

**b**•eye Business Intelligence Network™ The Global Vision for BI and Beyond

These products are designed to allow users to capture a complete market perspective, collecting data from the broad universe of customers and constituents, accessing data in multiple formats and languages. They apply Text Mining for Clementine which, as SPSS describes the software:

- Offers a broader understanding of customers' perspectives, taking into account qualitative feedback across multiple channels.
- Clarifies customer feedback, exploring the associations between human opinions and related topics.
- Improves predictions by combining the insights from unstructured sources with traditional business data.

SPSS does emphasize the applicability of its predictive analytics solution set for analysis of data, both numerical and text-sourced, targeting a broad set of enterprise goals. The company explains that users apply SPSS predictive analytics to, among other tasks:

- Increase revenues
- Identify fraud/minimize risk
- Bring products to market
- Claim handling (insurance)
- Find cures for diseases; improve treatment outcomes
- Fight crime and terrorism
- Improve manufacturing quality
- Improve inventory management
- Predict equipment failures
- Identify cyber-security threats
- Market research

Enterprise deployment of the SPSS predictive-analytics software is facilitated by Predictive Enterprise Services, "a unified platform providing a systematic approach to adopting and deploying Predictive Analytics consisting of Predictive Enterprise Repository and Predictive Enterprise Manager that, respectively, centralize and secure analytical assets (repository) and automate and manage analytical processes (manager). Figure 1 depicts the Predictive Enterprise Services visual-workflow interface, which enables a process to be built from a number of different components.





Figure 1: The Predictive Enterprise Services visual-workflow interface

## **VOICE OF THE CUSTOMER APPLICATION**

SPSS places a premium on serving organizations that want to improve customer satisfaction, whether as an end in itself or as a way to boost profitability or both. According to SPSS CEO Jack Noonan, "It doesn't matter whether you're dealing with a customer through an ATM, a Web store, a call center or at point of sale in your store; you want to make the interaction as positive as possible. Five years ago everything was about cutting cost. You would be focusing predictive analytics on changing business process but typically to save money. Now the focus is on changing business process to increase revenue profitably."

SPSS's Voice of the Customer solution is designed for business end users, for corporate roles that include:

- Marketing VP or Director
- CRM/Customer Insight Manager
- Quality Assurance Manager
- Product Manager/Marketer
- Survey and Market Research Director

The solution supports work under the rubrics Enterprise Feedback Management and Customer Experience Management in addition to the Voice of the Customer label, supporting any initiative that seeks to analyze data about people's attitudes and behaviors to improve business outcomes.

SPSS explains that its VoC solutions are not only about BI-style reporting, but also "greatly contribute in enhancing predictive models to better anticipate business

**b**•eye Business Intelligence Network™ The Global Vision for BI and Beyond

outcomes. They are interoperable with existing technology and approaches." The company provides connectors to most information systems that enable easy access to customer data. "Unstructured data can be collected from surveys, blogs, call-center notes and email and then analyzed seamlessly."

SPSS Text Analytics VoC solution can be extended to support audio data and a spectrum of languages via company partnerships and product integration with:

- **CallMiner** for call mining and advanced analysis of acoustic and speech-to-text data.
- LanguageWeaver for automated translation. A connector with Language Weaver servers is available through Clementine and Text Mining for Clementine to translate 14 different languages into English.

The company has developed VoC text-analytics best practices. SPSS advises customers to seek advanced capabilities that allow them (drawing from and quoting company sources) to:

- Understand dirty and noisy data: VoC text analytics must be able to deal with typos, miswriting and abbreviations which are the rule while analyzing emails, blogs or call-center data.
- Extend efforts to sentiment analysis: By not measuring only topics, but also emotional responses linked to topics, organizations will improve the accuracy of predictions based on trends found within data.
- Avoid black boxes, applications that are hardly customizable or that only advanced experts can really tune.
- Seriously consider support for multiple languages as more and more companies are getting global and need to better understand specific markets.
- Seek an integrated VoC platform for the predictive analytics journey, going beyond BI technologies that focus on what's happening now or what's happened in the past by primarily using financial or product data. For organizations to take the most effective action, they need to know and plan for what may happen in the future by using people data and predictive analytics.
- Avoid data silos and properly collect customer data through different channels.
- Integrate structured and unstructured data, which alone cannot provide a global and accurate understanding of customer behavior.

## DEPLOYMENT

SPSS licenses text analytics and its full set of predictive analytics technologies for in-house installation and use. In addition, SPSS Online Services can analyze



customer needs and provide a hosted solution for collecting and analyzing data.

The company offers consulting and customization services to assist in new implementations or configure existing applications to vertical industry and customer requirements, for Voice of the Customer and the gamut of enterprise challenges.

SPSS has found that capturing Voice of the Customer data as expressed in surveys and social media and other sources of enterprise feedback, coupled with sentiment analysis and analysis of structured data such as demographics and transactional history, permits more accurate results, better predictive modeling and customer understanding, and significant return on investment (ROI). <u>Nucleus Research determined</u> that "94 percent of SPSS customers achieved a positive ROI with an average payback period of 10.7 months." Over 90% of users attributed an increase in productivity to SPSS. 81% of projects were deployed on time, 75% on or under budget.

<u>Nucleus Research, in a December 2007 Guidebook on SPSS Text Mining</u>, further found that the SPSS technology can reduce customer churn by more than 50% when properly implemented as part of an overall customer-satisfaction strategy. They also found that text-mining functionality increases analyst productivity by up to 50%.6

Analyst firm Gartner positioned SPSS in the leaders quadrant of its "Magic Quadrant for Customer Data Mining, 2Q07 (May 2007)." According to Gartner, "SPSS has the broadest vision of the analysis of all types of data (behavioral, demographic, survey and unstructured). Enterprises seeking to make full use of their data (particularly Web traffic, text and survey data) should evaluate SPSS."

## SUMMARY

SPSS has unparalleled experience delivering predictive analytics solutions – for data generated by enterprise marketing, customer-care, product development, quality and feedback management needs – for both numerical and textual information. The company supports Voice of the Customer and a spectrum of related initiatives. It has generated an extensive list of satisfied customers internationally and has won very positive analyst and market reaction for all elements of its predictive analytics solutions set including Text Mining for Clementine and other software targeting enterprise VoC needs.



# IBM CUSTOMER CASE STUDY: NTT DoCoMo

## **COMPANY OVERVIEW**

NTT DoCoMo is Japan's leading mobile communications operator and also an influential force in advancing mobile communications technology on a global scale. In addition to providing voice and data communications to millions in Japan, DoCoMo creates global industry standards and groundbreaking mobile services such as i-mode and, most recently, "Osaifu-Keitai," mobile phones with wallet functions. Powered by research and development and a "customer first" philosophy, DoCoMo is constantly reinventing the concept of mobile communications and pioneering cutting-edge, cost-effective services that make life richer and more convenient for customers.

With over 53 million subscribers currently, DoCoMo accounts for more than half of Japan's mobile phone market and has one of the largest subscriber bases of any mobile phone company in the world. Established in July 1992 to take over the mobile communications business of Nippon Telegraph and Telephone Corporation (NTT), NTT DoCoMo launched its first digital cellular phone service in the next year.

## BUSINESS DRIVERS: UNDERSTANDING THE CUSTOMER, SERVING THE CUSTOMER

NTT DoCoMo has built on its initial use of Voice of the Customer text analytics, a study of market reaction to the introduction of new products, to create a comprehensive program that assesses customer needs and satisfaction and allows the company to better serve customers. The company wishes to know how closely DoCoMo products and services meet customer and enterprise requirements and to manage sources of customer dissatisfaction. The company wishes to understand customer attitudes and how it can systematize responses to customer issues.

Makoto Ichise, a manager responsible for text-analytics initiatives in NTT DoCoMo's information systems department, wonders, "I see many companies emphasizing customers first, listen to the customers, but...how many customers are actually listened to?" DoCoMo numbers its customers at above 53 million, generating over 20 million call-center contacts each year, and they try to listen to as many customers as possible, a goal facilitated by implementation of a Voice of the Customer text analytics program.

## **REQUIREMENTS AND METHODOLOGY**

Ichise reports that NTT DoCoMo started its Voice of the Customer program over five years ago, in April 2003. The company's first Voice of the Customer initiative was an investigation of customer reaction to a new multimedia service for i-mode mobile telephones, called "Freedom of Mobile Access," that it had launched the year



before. At that point in time, the early, steep growth curve had begun to level off. While substantial growth still lay ahead, nonetheless this introduction came at a point of transition to slower growth rates for i-mode mobile phones.

DoCoMo started with observation rather than with an active feedback management program, initially using information stored in Excel. More broadly, the company studies the spectrum of contact center information, derived from customer-staff interactions captured in transcripts. Analyses allow the company to understand not only customer needs, but also employee awareness and, thereby, employee ability to respond to customer needs.

The Voice of the Customer system is integrated with DoCoMo's Aladin customer database. Work is carried out in house via the installed CS-Aladin analytical system, which integrates IBM's OmniFind Analytics Edition system. Analyses allow DoCoMo to generate business rules that, when put into execution, enable DoCoMo to identify common needs and wants and improve responsiveness.

DoCoMo offers two scenarios of text analysis, one focused on knowledge discovery and the second on classification for processing. The Discovery scenario aims to identify valuable information without imposing preconceived models. DoCoMo wishes to discover significant topics – whether complaints, concerns, requests or questions – and to identify trends. Manual analysis can provide these types of insight, but is not an effective approach to this large-scale program.

DoCoMo's Voice of the Customer Classification scenario seeks to assign interactions to categories along multiple axes. These include the contact type (for instance: praise, inquiry, concern and opinion) and subject area (for example, quality, function or performance, design, accessibility and support). DoCoMo additionally considers urgency in its classification model – whether a direct, real-time response is required or whether the need for response is less immediate – and also includes case treatment, for instance, whether a document needs to be kept or may be disposed of. Overall, DoCoMo classifies into around 1,000 categories, from which it has defined about 25,000 mining rules.

DoCoMo plans to expand the reach of its Voice of the Customer program beyond call center transcripts to encompass blogs.

#### **BUSINESS INTEGRATION**

NTT DoCoMo's CS-Aladin analytical solution, which relies on IBM's OmniFind Analytics Edition system for Voice of the Customer analyses, is integrated with its implementation of the Aladin and Aladin CTI solutions for customer-facing call center operations. Those call center systems feed approximately 200,000 records each month to analytical databases for Voice of the Customer quantitative and qualitative analyses.

On the output side, reports are delivered to a variety of DoCoMo business divisions, which, in turn, issue requests for further analyses.



## **BUSINESS BENEFITS**

NTT DoCoMo is clear about the benefits of sharing analytical insights across the entire organization, enabling them to:

- Use the insights to improve products and services.
- Use the insights to boost customer satisfaction.
- Share insights with customers when appropriate.

DoCoMo derives customer insight from over 20 million contact-center calls per year. The company's OmniFind Analytics Edition solution affords employees the opportunity to understand intent and think about how to apply findings in practice. The company applies analysis for:

- Alert generation.
- Trend analysis.
- Compliant monitoring.

DoCoMo has experienced demonstrable business benefit from text mining. In the area of service transformation, DoCoMo offers the example of qualitative analysis that identified unusual customer requests regarding use of cellular phones at international airports. Customers wish to use their cellular phones until take-off for overseas trips. And customers wish to use their same number abroad despite having been too busy to make prior adaptive arrangements. These requests, mined from customer interactions, led DoCoMo to open shops at major airports in Japan. The popularity of these airport shops led the company to open multiple shops at some airports to handle customer demand.

DoCoMo offers an example of quantitative analysis that identified a widespread complaint that its fee system was too complicated to understand. As a result, the company simplified its fee system, making it both easier to understand and more competitive.

In summary, Voice of the Customer text analytics enables a deeper understanding of the customer – to improve sensitivity – and in the words of IT Manager Makoto Ichise, to "convert negative messages into opportunities and requests into solutions." Ichise remarks that "textual records usually require contextual understanding, including the author's [customer's] emotions" – a difficult problem – but text analytics "often leads to rewarding results."



# SPSS CUSTOMER CASE STUDY: CABLECOM GMBH

#### **COMPANY OVERVIEW**

Cablecom is the largest cable network operator in Switzerland. The company provides its customers with services for cable television, broadband Internet access, and mobile and fixed network telephony.

As of March 31, 2008, Cablecom had over 1.5 million television customers, of which 300,000 made use of the digital service, 468,000 were Internet customers (high speed) and 298,000 were telephony customers (digital phone). The company also provides voice, data and value-added services for business customers. And Cablecom supplies cable network operators with facility construction as well as with application and transmission services for television, telephony and Internet. The company's own network connects around 1.9 million households and supplies all the larger towns in Switzerland. At the end of December the company had around 1,500 employees and its turnover was over 1 billion Swiss francs.

<u>Cablecom</u> is a national subsidiary of UPC Broadband, the European cable network group of Liberty Global Inc.

## **BUSINESS DRIVERS: A FOCUS ON SATISFACTION**

Cablecom has maintained its market position as Switzerland's largest cable network operator via an intent focus on customer satisfaction. But while superior service and happy customers are ends in themselves, they are also the key to profitability in competitive, mature markets such as Cablecom's.

Cablecom had gone through a period of rapid expansion, with acquisition of new customers driving revenue growth and profitability. However, as the company's market matured – as customer-acquisition opportunities diminished – staff turned their attention to customer retention. According to Federico Cesconi, Cablecom's director of business intelligence (BI), the company has found that retention is linked directly to service quality and satisfaction; hence, the IT mandate became detecting quality issues and those customers who were not completely happy.

## **REQUIREMENTS AND METHODOLOGY**

Cablecom had six years of experience working with analytics vendor SPSS as documented in published case studies. Nonetheless, when it came time to extend customer-retention efforts to textual data sources, Cesconi and team surveyed the vendor marketplace to find the solution that would best help them meet business needs. They identified other potential suppliers but chose to proceed with an SPSS solution based on the Cablecom's strong, satisfied business relationship with SPSS – in Cesconi's words, SPSS is "not really a supplier, but a partner" – and the SPSS products' technical capabilities. Further, Cablecom found that adding text mining to an existing SPSS Clementine solution is not so expensive. Expense, in this case, is

*Copyright* © 2008 *Powell Media and Alta Plana Corporation All Rights Reserved* 



measured not only in software licensing cost; expense calculations should include the cost of staff training and ongoing operations. Cablecom faced modest additional costs due to the integration of text analytics capabilities into the familiar Clementine interface.

Cablecom's Voice of the Customer (VoC) work, aimed at boosting customer retention, is carried out primarily by an internal staff of 15. It is part of an Enterprise Feedback Management (EFM) effort that encompasses text-analytics and also use of the NetPromoter Score, which provides a mechanism for measuring customer attitudes. The effort allows the company to generate categories of drivers of customer satisfaction and dissatisfaction, with segmentation according to customer lifetime value and NetPromoter-measured satisfaction.

Cablecom has several years experience with data mining. Staff had achieved good predictive accuracy with a variety of churn models, but they began to see the limitations of data mining – the need to go beyond numbers to understand "customer emotions."

Cablecom staff had found particularly useful the unstructured information – freetext, verbatim responses – collected in surveys, but non-automated analysis was very cost ineffective, an obstacle overcome via Voice of the Customer text analytics. The company now initiates online and interactive voice-response surveys. Survey outreach is triggered by events such as a call-center inquiry and the customer's reaching a certain point in the product life cycle. Automating analyses with SPSS text analytics has enabled Cablecom to extend EFM scope to accommodate approximately 20,000 feedback items per month.

Plans include extending VoC text analytics to blogs and online forums.

### **BUSINESS INTEGRATION**

Voice of the Customer-driven retention is a high-priority, high-visibility initiative for Cablecom. NetPromoter adoption was at the CEO's instigation, and consolidated VoC findings are reported monthly to the company's Board of Directors.

BI Director Federico Cesconi had in the past been responsible for customer relationship management (CRM) analytics. Cablecom's Chief Financial and Marketing Officers decided to create an independent BI unit, led by Cesconi, tasked to work with a variety of business units. The BI unit has two groups: one focused on "business insights" linked to operations and finance; and the second, a "customer insights" group that measures business key performance indicators (KPIs) from the customer's point of view.

Executive visibility is important, but it's action that produces tangible results. Cablecom's Voice of the Customer initiative produces *actionable* information. There are two streams:

1. Cablecom's VoC findings help identify operational gaps and directly support process improvement.



2. Cablecom's VoC findings are part of a one-to-one support initiative where the company reaches out to customers with issues – customers who are dissatisfied or likely to terminate their service – in a personalized effort to fix problems. This form of interaction is expensive, but the cost is justified by retention results.

## **BUSINESS BENEFITS**

Cablecom has seen a tremendous improvement in churn and customer satisfaction according to Director of Business Intelligence Federico Cesconi. The company quantified results. It assessed the effect of its personalized call-back effort by studying 1,200 cases two months after call-back. Findings showed that this VoC-driven step would reverse customer-satisfaction problems in 55% of cases. Customers who rated their satisfaction at 0-6 on a ten-point scale could be boosted to an 8-10 point score.

## **LESSONS LEARNED**

Cablecom's Federico Cesconi reports that the company's Voice of the Customer work generated "a lot of surprises" – findings and results that were not anticipated – including that "the customer has a totally different point of view about your business" from staff's view. He cites an example that he and colleagues had always thought it was important to solve customer problems "very, very fast." Customer feedback showed, however, that for certain problems, the customer can wait, that what's most important is that "when we solve problems, we keep our promise." Cesconi offers the lesson that organizations should account for both the business and the customer points of view.

When Cesconi discusses Cablecom VoC initiatives with contacts in other industries, he recommends careful structuring of Enterprise Feedback Management efforts. For instance, some organizations think that use of an online tool such as Survey Monkey is enough. Cesconi, by contrast, finds immense value in Cablecom's comprehensive EFM approach. "The goal is not simply to collect feedback. You need to be able to enhance the customer experience."



# **AUTHOR INFORMATION**

## **SETH GRIMES**

Seth Grimes is a business intelligence, data warehousing, and decision systems expert. He founded Washington D.C.-based Alta Plana Corporation in 1997 and consults on business intelligence and analytics strategy and implementation for clients in the U.S. and internationally. In addition to consulting, Seth writes and speaks on data management and analysis systems, industry trends and emerging analytical technologies.

Seth is Text Analytics Channel Expert for the *Business Intelligence Network* and a *Data Warehousing Institute* instructor. He is founding chair of the *Text Analytics Summit* series, dating to 2005.

Seth can be reached at grimes@altaplana.com, +1 301-270-0795.

## **ALTA PLANA CORPORATION**

7300 Willow Avenue Takoma Park, MD 20912 +1 301-270-0795 *altaplana.com*