

What's Next for Text

Text Analytics Today and Tomorrow: Market, Technology, and Trends

A white paper prepared for

Text Analytics Summit 2007

Boston, June 12-13, 2007

www.TextAnalyticsNews.com/usa



Seth Grimes

Alta Plana Corporation

Alta Plana

TABLE OF CONTENTS

| | | |
|-------|--|----|
| 1 | INTRODUCTION..... | 3 |
| 2 | MARKET SUCCESS | 4 |
| 2.1 | FROM MINING TO ANALYTICS..... | 4 |
| 2.2 | FROM TOOLS TO SOLUTIONS | 4 |
| 2.3 | EMERGING APPLICATIONS..... | 5 |
| 2.3.1 | SENTIMENT EXTRACTION, MARKET RESEARCH, AND SURVEY ANALYSIS | 6 |
| 2.3.2 | SEARCH ENRICHMENT..... | 6 |
| 2.3.3 | MEDIA MINING..... | 6 |
| 3 | CONTINUING CHALLENGES..... | 7 |
| 3.1 | QUESTION ANSWERING: SEARCH DONE RIGHT | 7 |
| 3.2 | MULTILINGUALISM | 7 |
| 3.3 | SEMANTIC WEB SERVICES..... | 7 |
| 3.4 | A GRAND CHALLENGE..... | 8 |
| 4 | MARKET OUTLOOK | 9 |
| 4.1 | VENDOR TRENDS..... | 9 |
| 4.1.1 | CONSOLIDATION | 9 |
| 4.1.2 | CLARIFICATION..... | 9 |
| 4.1.3 | ALLIANCE | 10 |
| 4.1.4 | EMERGENCE..... | 10 |
| 4.2 | MARKET VALIDATION | 10 |
| 5 | APPENDIX: TECHNOLOGY BASICS | 11 |

I INTRODUCTION

Text analytics has entered the IT mainstream:

- The technology has matured in ways that reduce complexity, boost accessibility and usability, and widen appeal for end users.
- Usage has broadened to encompass new types of applications and new business domains.
- The field is supported by a vibrant, international research community with strong academic, government, and industry participation.
- Sales are strong and growing at a rapid pace with available options ranging from a wealth of specialized, free, open-source software tools to enterprise scale, integrated analytical systems.
- The market outlook is creating opportunity for new entrants and prompting a healthy level of merger and acquisition activity.
- The value of text analytics and its potential enterprise role are newly understood by the wider world of analysts, trade journalists, and end users.

Text analytics' successes date back up to ten years, achieved in fields such as life sciences and counterterrorism. The technology has notably revolutionized the pharmaceutical drug-discovery process. Scientists can mine biomedical literature for protein interactions that suggest therapies, reserving time-consuming and expensive laboratory work and clinical trials for the few, most promising leads. Intelligence analysts can similarly use automated methods to discover telling patterns hidden in large volumes of textual materials, materials that were previously inaccessible due to source language, volume, and resource constraints.

Yet investigative analyses have been pushed from the text-analytics center stage. While many users still seek needles in haystacks, more-recent text-analytics adopters are most interested in statistically described characteristics of larger populations. They're looking for trends and big picture rules rather than for the few exceptional cases. They want a matching style of analysis, one that has served them well in working with conventional, numerical data sources. They want familiar business intelligence (BI) interfaces and techniques, also adaptation of concepts such as Extract, Transform, Load (ETL) to their work with textual sources. And they want software that deals automatically with linguistic complexity and delivers information in the structures and interfaces they're using for line-of-business applications and analytics.

The shift from a knowledge-discovery to an analytics paradigm is part of text mining's expansion to support a larger world of enterprise analytics. This shift allows and encourages a majority of organizations to investigate and adopt the technology. The questions of those concerned with enterprise analytics are no longer *if, what, or how*, but rather *where and when* – questions this paper will address in examining *What's Next for Text*.

2 MARKET SUCCESS

“Organizations embracing text analytics all report having an epiphany moment when they suddenly knew more than before.”

– Philip Russom, the Data Warehousing Institute

2.1 FROM MINING TO ANALYTICS

Text mining derives from data mining, a statistically rooted approach to classification, clustering, and derivation of association rules, with data mining applied to fielded, numerical data and text mining extending processes to encompass textual data sources. The aim is description and prediction. The technologies are typically embedded in larger knowledge-discovery efforts that add data management and visualization to the mix. But where semantics in data mining – the meanings of variables and values – are captured in variable names and descriptions and other metadata, when working with text, it's a series of analytical steps that infers machine-comprehensible semantic meaning in “unstructured” textual source materials.

Text is replete with linguistic and grammatical and compositional structure – the form and arrangement of the words and sentences from which documents are built. People exploit this inherent structure without conscious thought when they read. Machines must similarly exploit it to extend automated knowledge discovery to text. They use analytical software that annotates and restructures text into mineable form via small but important steps that discern the morphology of words, the boundaries of sentences and other syntactical units, and the contextual meaning of terms.

Search is one of the Web's (and the enterprise's) killer apps, and ironically search's short-comings suggest the value text analytics can deliver. Users want to work with natural language and they want more than the information retrieval – essentially, document retrieval – that search (currently) offers. Text analytics add semantics possibilities both in the processing of search queries and in the use of search findings. Secondly, search is poised to transform the business intelligence world, again aided by a dose of text analytics. Vendors have struggled for years to make BI accessible to a greater variety of end users and for a greater variety of applications. Operational BI, dashboards and portals, visualization, and spreadsheet integration have all advanced BI technology toward meeting this goal, but it is the promise of cutting through the clutter of menus and catalogs that is the best candidate to deliver.

2.2 FROM TOOLS TO SOLUTIONS

A shift in market focus from tools to solutions is a hallmark of market maturation. Tools work well for technically oriented researchers – traditional text miners – who expect to do significant configuration and customization in order to accomplish highly specialized tasks. This class of users works with software components and workbench interfaces to design processes that meet individual needs. Such tool-centric approaches serve them well but are unsuitable for the broader world of enterprise data and business analysts whose needs are solution oriented.

One can build solutions quickly when software components are easy to configure to

¹ www.tdwi.org/Research/display.aspx?id=8355

meet domain- or application-specific needs while providing a high level of capability and performance for generalized problems. The market has evolved in recent years to provide such solutions. Text analytics is now enterprise-ready:

- Entity extraction accuracy – precision and recall – is high enough to meet the needs of a broad variety of applications in disparate business domains.
- Usable solutions are now available to address the challenge of tapping the axiomatic 80 percent of enterprise information that is locked in textual form.
- Packages are available for mainstream business needs – survey processing, competitive intelligence, CRM, reputation management – that integrate with line-of-business applications and conventional analytical software.

Therefore:

- Text analytics is now accessible to business analysts, not just linguists and statisticians and scientific researchers.

And enterprise integration is facilitated by widening Unstructured Information Management Architecture (UIMA) compliance. The specification for this open-source, platform-independent, software-integration framework is being developed for “interoperability, discovery, and composition of analytics across modalities, domain models, and frameworks”² by an OASIS technical committee, and creation of a reference implementation is now an Apache incubator project.

2.3 EMERGING APPLICATIONS

Growth is largely driven by the wealth of unstructured information found on the external web, in corporate intranets, document repositories, call-centers, and in customer and employee business communications.³

- UIMA Chief Architect David Ferrucci and fellow IBM researchers

The earliest text analytics successes have been in investigative applications in the life sciences and intelligence. Breck Baldwin of Alias-I goes so far as to state that “the Nobel Prize in Medicine will [soon] go to a researcher primarily focused on text analytics because of the discoveries made possible by the technology.” Such segments will remain important, but the majority of market growth will occur elsewhere.

Justin Langseth, president and cofounder of text-mining vendor Clarabridge, offers that “the biggest changes over the last year have been squarely associated with extremely rapid growth of the text-mining market within the commercial sector, primarily for marketing uses – product positioning before launches, competitive intelligence, quantifying sentiment, early issue detection – and for customer service and product management uses – listening to the voice of the customer to better understand their needs, frustrations, suggestions, etc.”

Emerging applications include market research and survey analysis, reputation management via monitoring of networked social media, and semantic enrichment of Web and enterprise search. We will also look at extension of analytical approaches to media beyond text. These will be important areas for text analytics in the near term.

² xml.coverpages.org/UIMA-TC-Announce.html

³ “Towards an Interoperability Standard for Text and Multi-Modal Analytics,” [domino.research.ibm.com/library/cyberdig.nsf/papers/1898F3F640FEF47E8525723C00551250/\\$File/rc24122.pdf](http://domino.research.ibm.com/library/cyberdig.nsf/papers/1898F3F640FEF47E8525723C00551250/$File/rc24122.pdf)

2.3.1 SENTIMENT EXTRACTION, MARKET RESEARCH, AND SURVEY ANALYSIS

For market researchers and survey analysts, text analytics is a compelling replacement for slow, expensive, and cumbersome old-school approaches to dealing with free-form information. One human-resources manager interviewed reported that text-analytics software allows him to accomplish in half a day work that formerly took eight analysts one week. Further, automated processes may be integrated with line-of-business software, for instance for Customer Relationship Management. This integration creates a new market category, Enterprise Feedback Management.

These applications rely on sentiment-discovery capabilities that go beyond basic entity extraction. Sentiment can be hard to discern and quantify accurately, so Anne Milley, a SAS Institute marketing executive, advocates a hybrid approach that joins fielded ratings with text analysis. Milley states that “model accuracy is good – and we believe most likely better than models that use non-supervised methods.”

Reputation management further applies sentiment extraction and related techniques to social media: blogs, message boards, and other online sites where opinion can propagate quickly. Organizations wish to detect postings that cite them and their products, competitors, and customers. They wish to handle incidents expeditiously to maintain a positive image, and they rely ever increasingly on text analytics.

2.3.2 SEARCH ENRICHMENT

Users tolerate often-low search accuracy because consumer search is free and far better than inability to find information. One tactic, manual tagging based on *folksonomies*, improves findability and retrieval relevance but can be only a partial remedy due to the Internet's volume and volatility.

Where established and informal search and tagging methods fall short is where semantic technologies can play a leading role. Semantic indexing and search provide the possibility of going beyond keywords to exploit intent, context, and concepts inferred from both search queries and target documents.

Enhanced search is an excellent resource for investigative analyses and may obviate the need, in many cases, to extract information to databases. Alias-I's Baldwin explains that “database filling will no longer be seen as the end goal of high-end text analytics. Reversing a long standing trend, enhanced search portals are the new focus of technologies ranging from brand sentiment to entity and relation extraction.”

2.3.3 MEDIA MINING

User organizations and vendors recognize that while there's a lot of text out there to tap, there's more to “unstructured” information than just text. In terms of technology application, free text is simply a nearest next step away from fielded data into a world of diverse media. Media miners automate information extraction from speech, sound, still images, and video and the analysis and processing of those forms. Important early applications are the same as those text analysts tackle: automated handling of customer communications, security applications, and the like, but with a need to apply native methods, for instance the use of phonology to mine speech rather than just blanket transcription to text. There are dissimilarities, for instance, media mining may involve a real-time angle that is not imperative in text processing. Nonetheless, integrated data-text-media analytics is another area that is *next for text*.

3 CONTINUING CHALLENGES

Went finding and seeking,
Finding less than sought
Seeking more than found,
Every detail minding
Of the seeking or the finding.

– Langston Hughes, Old Walt

Text analytics meets many challenges; nonetheless, we seek ever more powerful technology to respond to evolving demand. Many challenges are unmet or partly met. They suggest *what's next for text* and are well worth exploring.

3.1 QUESTION ANSWERING: SEARCH DONE RIGHT

Few search users are looking for the hit lists of dubious relevance returned by the major search engines. Rather, most searches are really questions – *Who is the president of Peru? What's the best price in the Washington, DC area for a Novara Safari bike?* – and most searchers are seeking facts rather than undigested documents. It is text analytics that will deliver the capabilities needed to provide answers, for search done right.

Any attempt at question-answering involves understanding the searcher's intended meaning, finding suitable information sources, extracting facts, and composing responses. Matthew Glotzbach, head of products for Google Enterprise, has said that "question answering is the future of enterprise search," a statement that applies equally to public Web search. Advanced text technologies are an essential ingredient.

3.2 MULTILINGUALISM

Multilingual and cross-lingual text analytics – going beyond *multiply monolingual* processing – is perhaps more talked about than usable. We wish to exploit text without regard to language, which requires more than ability to operate in a single language or to translate from foreign languages into a canonical language, whether English or some other. Software must understand, equivalence, and jointly process vocabulary, grammar, and concepts for disparate languages. These functions are nominally within the capability of many existing text-analytics tools, yet for most users, real-world constraints come into play. In the words of an executive of one multilingual text-analytics vendor, "Everyone seems to ask about it, and they often have a checkbox on their RFPs for it, but in practice, dealing with the massive amount of data in English first keeps foreign language handling off the table."

3.3 SEMANTIC WEB SERVICES

The term *semantic Web services* could refer both to services for the still-elusive Semantic Web and to Web services that provide semantic analysis on-demand to invoking applications. The first referent is a vision whose realization remains elusive a decade after it was first articulated by Web inventor Tim Berners-Lee: it is difficult to get information publishers to mark up their content. The second would provide an automated complement to *folksonomies* that overlay semantics on the existing, chaotic computing world: think semantic *mash-ups*. Initial attempts at this second type of semantic Web services, albeit short on accuracy, are available. They are an important step in another direction that text analytics is heading.

3.4 A GRAND CHALLENGE

Ronen Feldman, professor of computer science at Bar-Ilan University and cofounder of text-analytics vendor ClearForest, proposes a “grand challenge” for text mining⁴. Feldman’s grand challenge is expressed in testable terms:

Text mining systems that will be able to pass standard reading comprehension tests such as SAT, GRE, GMAT, etc.

Meeting this grand challenge entails improved Entity Extraction,

We are seeking domain independent and language independent [named entity recognition and] relation extraction systems that will be able to reach precision of 98-100% and recall of 95-100%. Since the systems should work in any domain, they must be totally autonomous and require no human intervention.

Those Autonomous Text Analysis systems

will analyze huge corpuses and come up with truly interesting findings that are not captured by any single document in the corpus and are not known before... Such systems can then be used for alerting purposes in the financial domain, the anti-terror domain, the biomedical domain and many other commercial domains.

Feldman believes that within five years researchers will be able to demonstrate systems that meet the Extraction and Autonomicity targets, using the Web (and presumably other corpora) for above-average scores on grand-challenge test questions.

Yet the best-of-five answer to a multiple-choice question, based on a few paragraphs of clean text, could conceivably be picked by moderately sophisticated pattern-matching software. Developers are more ambitious, working to boost accuracy when dealing with real-world information sources and conditions, to deliver the ability to mine noisy materials such as call-center notes, survey responses, e-mail, and the like. Here, syntax may be fractured and ungrammatical, spelling is irregular with abbreviations, and a given source document may contain externalities, what a linguist would call exophora, or references that are not resolved in examining a single source document. And, noting that much information found on the Web (and in the enterprise) is of dubious accuracy and authority, a next-generation, high accuracy, real-world, autonomous text miner would assess and weigh the correctness of identified responses in order to formulate a single, contextually best answer.

Feldman cites the Turing Test as, essentially, a generalization of his text-mining grand challenge. The Turing Test is a conversation: Can a person tell that an interlocutor is a machine? Conversations take place over time. They flow and meander and are sometimes discontinuous. The language usage is fractured and meaning may depend on context and externalities. And responses are contextual rather than selected from among a small set of prepared choices in an artificial test. Passing the 57-year-old Turing Test, suitably updated, may be the best real-world grand challenge, guiding researchers toward a comprehensive treatment of text.

Text analytics is already part-way there: witness visualizations of mined information ranging from networks of protein-protein interactions to the rendering of search-engine results in hierarchical clusters. Yet many essential tasks remain undone. Accomplishing those tasks in the coming years is *What's Next for Text*.

⁴ “What Are The Grand Challenges for Data Mining? KDD-2006 Panel Report,” www.acm.org/sigs/sigkdd/explorations/issues/8-2-2006-12/9-is-there-grand-challenge-for-dm-explorations-final.pdf

4 MARKET OUTLOOK

The market boasts a diversity of text-analytics products that accommodate a spectrum of application types and business domains. The market includes pure-play text analytics and computational-linguistics vendors. It includes solution providers that target particular applications and domains and it includes companies offering integrated analytics. Search vendors loom large, and business-intelligence companies are looking to add text-analytics to their product sets. Content-management vendors are involved although their interest is largely limited to information retrieval.

The author estimates a worldwide market for text-analytics software licenses, support, and professional services of about \$200 million for 2006 with 25 percent annual growth through 2010.

This figure covers pure-plays, targeted solutions, and an allotment of the portion of other-analytics revenue that is attributable to text analytics. By contrast, Gartner estimates worldwide business-intelligence software licensing, exclusive of services, at around \$2.5 billion with growth under 10 percent. Applying a conservative multiplier to compute user/contractor labor, and assessing the value of academic research and of content and database products produced with technology developed in-house by publishers, we arrive at a \$2 billion valuation of worldwide text-analytics research, software, services, and applications.

4.1 VENDOR TRENDS

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

4.1.1 CONSOLIDATION

Text analytics is part of a larger data- and content-management analytics market. Vendor consolidation must be considered in context of the larger market. The deals are significant even if smaller than, say, Oracle's \$3 billion purchase of business-intelligence vendor Hyperion. The largest text-technologies transaction was the late-2005 Autonomy acquisition of enterprise-search rival Verity, Inc. for approximately \$500 million. The combined company has annual revenue of over \$200 million. More recently, BI/integration vendor Informatica paid \$55 million for Itemfield to add the ability to tap "unstructured" sources to its ETL suite. The deal closed in December, 2006. And enterprise-search giant Fast Search & Transfer has announced a \$23 million purchase of Convera's RetrievalWare business unit that is expected to close in the second quarter of 2007. RetrievalWare provides context-aware information access using a variety of text-analytics technologies that will complement current FAST capabilities.

Expect a quickened pace of merger and acquisition activity as database, BI, and enterprise-applications vendors seek to add text technologies to their product lines. Smaller companies, including some that are struggling, are particularly inviting targets.

4.1.2 CLARIFICATION

There is significant room for market clarification due to the variety of technologies in

use and the ways vendors have positioned themselves. On the technology front, we have traditional inverted keyword indexes for search, varieties of semantic indexing and analysis, Bayesian supervised classifiers, and relational-linguistic entity extraction. As vendors consolidate – as in the Autonomy-Verity merger – and as they add capabilities, the market will require technology clarification.

That clarification may reinforce product categories or it may blur boundaries and reduce them. Products already sit in multiple categories, such as computational linguistics, text-analytics companies, general/integrated or industry analytics, text-BI, Web and social-media mining, contextual advertising, content database publishers, and search vendors. These categories are useful for product differentiation but may hinder maximum market growth

4.1.3 ALLIANCE

The text-analytics market has many inter-vendor alliances. Many solution providers, both within the text-analytics world and in nearby content-management and BI segments, license entity-extraction and linguistic technologies for resale in “OEM” partnerships. Expect such cross-licensing arrangements to continue due to the specialized nature of certain general-application technologies.

4.1.4 EMERGENCE

Text analytics is fertile ground for new offerings. Most often, new products and solutions emerge as the commercialization of academic research or as the product-realization of in-house industrial research. Much basic research is government funded in the U.S. and by the European Commission, particularly when it addresses scientific, linguistic, and homeland-security challenges. Other, more immediately commercializable research is venture funded.

In other cases, such as SAIC's January 2005 sale of its Content Analyst division and intellectual property, new products emerge as spin-offs.

4.2 MARKET VALIDATION

Consider text analytics as an industry and not just a technology. Analyst attention and willingness of funders to make new and repeat investments are strong validation of text analytics' vitality and potential.

Analyst firms have newly added text analytics to their research portfolios, a success indicator that validates the importance and commercial value of the market. Whereas in past years only IDC and the 451 Group tracked and reported on text analytics, analysts from Forrester, Gartner, Hurwitz, and the Data Warehousing Institute have joined the fray. Look for more in-depth and frequent analyst coverage as the firms respond to a growing volume of client requests.

And in the last year, venture funders have placed multiple, multi-million dollar investments in companies including Clarabridge (\$7.2 million), InforSense (\$10 million), and Nstein (CAN\$10 million). Growth of companies including Attensity, Basis Technology, ClearForest, Inxight, SRA (NetOwl), and TEMIS has been fueled by strategic infusions of venture capital. Investors range from conventional venture firms to organizations such as IBM and In-Q-Tel that actively promote leading edge technologies to economic-development agencies such as ITI Life Sciences. Look for continued capital investments to support new entrants and growing firms.

5 APPENDIX: TECHNOLOGY BASICS

Thus the orb he roamed
With narrow search, and with inspection deep
Considered every creature.

– John Milton, Paradise Lost

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 e-mail, 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 extract, tag, 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, and extracted data may be used for other forms of analysis. 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.

SETH GRIMES

Seth Grimes is a business intelligence, data warehousing, and decision systems expert, a consultant and contributing editor for *Intelligent Enterprise*. He writes the magazine's Breakthrough Analysis column and contributes feature articles.

Seth founded Washington DC-based Alta Plana Corporation in 1997 and consults on information systems strategy for clients that include primarily government statistical agencies and marketing firms and also select software publishers.

Seth writes and speaks on data management and analysis systems, industry trends, and emerging analytical technologies. He is chair of the European Text Analytics Summit and chaired the 2005 and 2006 summits in Boston and will chair the 2007 summit. His white paper for the 2005 summit, [The Developing Text Mining Market](http://altaplana.com/TheDevelopingTextMiningMarket.pdf), is available online at altaplana.com/TheDevelopingTextMiningMarket.pdf. His 2006 summit paper, [Finding Value in Text Analytics](http://altaplana.com/FindingValueInTextAnalytics.pdf), is posted at altaplana.com/FindingValueInTextAnalytics.pdf and his 2007 European Text Analytics Summit, [European Text Analytics](http://altaplana.com/EuropeanTextAnalytics.pdf), from which Section 5 of this paper is excerpted, is posted at altaplana.com/EuropeanTextAnalytics.pdf.

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

TEXT ANALYTICS SUMMIT 2007

The Text Analytics Summit 2007 (www.textanalyticsnews.com/usa), slated for June 12-13, 2007 in Boston, is a mindshare event for the leading developers, researchers, vendors, tech-savvy users, and newcomers to the text-analytics space.

This third annual Boston summit follows on the heels of a first European summit, held in Amsterdam in April 2007. Analyst Curt Monash wrote in *Computer World* that “the [2005] Text Mining Summit ... was one of the best conferences I’ve been to in a long time.” SPSS Vice President Olivier Jouve called the 2006 summit “the best conference I attended last year.”

The European and North American Text Analytics Summits both provide an opportunity for researchers and vendors to identify promising applications, size up technical challenges, and connect with users eager to keep up with market developments. Text-analytics users and prospective users in any application or industry find an unmissable opportunity to learn from peers and understand the bottom-line impact of the latest deployments. Developers and marketers benefit from the opportunity to engage end users and technologists to better understand market requirements, technology developments, and product directions.

ALTA PLANA CORPORATION

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

TEXTANALYTICS
NEWS.COM

Back by popular demand for its third blockbusting year!

SAVE \$200
If you register before April 27

3rd ANNUAL
TEXT ANALYTICS SUMMIT



12 - 13 June 2007 • Boston Marriott Newton • USA

Discover lucrative new Text Mining markets, hear from the biggest commercial end users and analyze the leading applications that will increase your profits

Take advantage of new opportunities in the exploding US text mining market and meet cutting edge end users!

IN-DEPTH INSIGHT FROM THE BIGGEST NAMES IN GLOBAL TEXT ANALYTICS!



Hadar Shemtov, Director of Research, Natural Language Technologies, Yahoo, Inc.



Seth Grimes, President, Alta Plana



John Joyner, Director of Indexing Automation & Data Management, Thomson Learning



Ramana Rao, ex CTO, iCurrent Inc. (Founder, Inxight Software, Inc.)



Dr David Bean, CTO, Attensity

- **Brand New Case Studies:** Hear tech-savvy end users - including Dow Jones, Jet Blue Airways, Thomson Learning, Whirlpool and more - explain how they use text technologies to convert unstructured data into real, actionable intelligence
- **Business Opportunities:** Understand which horizontal and vertical markets are ready to be exploited: How can you capitalize on multi-lingual mining, search, email archiving, converged data, text and speech & more?
- **Technology Update:** Evaluate the very latest solutions for text analytics: Converging technologies, complete solutions, demonstration sessions, interactive workshops and live mining. Discover the latest tools impacting the bottom line for end users?
- **Sneak Preview:** Get up-to-speed on tomorrow's text capabilities and new markets and discover who's next in line to deploy text analytics in a BIG way

How many reasons do you need to attend the summit this year?

- ★ 250+ senior level attendees
- ★ 30+ Industry leading expert speakers
- ★ 15+ Stand exhibition
- ★ 10 Industry specific breakout roundtables
- Plus
- ★ Workshops
- ★ Text Mining Demonstrations
- ★ Extensive Networking
- ★ & Business Card Swaps



Official Bronze Sponsor



Official Speaker Recruitment Sponsor



Official Joint Sponsors

Summit Sponsors

Official Media Sponsors



"With text analytics booming, this year's Summit will be the best ever!"

Seth Grimes, President, Alta Plana

Look inside now to see what's new for 2007!

As text analytics technologies prepare to go mainstream, join your peers at this proven, top level event that shows you how to capitalize on all the opportunities – and maximize your rewards

In this age of information overload, it's no wonder that text analytics technology – and its ability to dramatically accelerate business decision making – is poised to hit the mass market. With an increasing range of tools that can spin your raw information into actionable intelligence, more and more North American organizations are joining the race to turn their data into dollars.

"85% of business information exists in text form" Merrill Lynch

Text analytics technologies and applications are spreading across all sorts of diverse industries like wildfire.

But like any innovation, it's not all plain sailing. The management of unstructured data is a large – and growing – problem. And the need to unlock the hidden potential of their data assets is recognized as a major business challenge by many of today's leading organizations.

What's *your* current position?

Looking for an effective fast-track solution? Want to harness additional profit locked away in your data? Eager to discover if your direct competitors are already stealing a competitive advantage? Open to the opportunity of picking the brains of cutting edge, real world experts who are already profiting handsomely from analysis technologies?

If so you cannot miss the **3rd Annual Text Analytics Summit**. An event so popular it had sold out in the last 2 years!

This is your chance to learn from the people who are currently deploying the latest, most powerful text-based solutions in all sorts of different commercial businesses including Pharmaceuticals, Life Sciences, Finance, Air Travel, and Media & Publishing as well as Government. Plus case studies from Yahoo Inc., Whirlpool, Dow Jones, Anderson Analytics, SPSS and more. Come and track their implementations.... understand their results.... and harness their knowledge to improve your own business development strategies.

"If you want to come up-to-speed quickly on what's available, who are the vendors, and who are the successful users of Text Analytics, this is the only place to be"

Dr David Bean, CTO, Attensity

Bringing you even more in 2007

- ★ **MORE** Case Studies and insights into how businesses are applying text analytics technologies – and the results being achieved
- ★ **MORE** lively panel sessions with discussion & debate
- ★ **MORE** education sessions, live demonstrations and user workshops
- ★ **MORE** industry-focused roundtable discussions
- ★ **MORE** live onsite text mining – with results!
- ★ **MORE** networking, exhibition space and dynamic group sessions

Organized by **Text Analytics News**, this is North America's only recognized annual meeting place focused entirely on text mining applications, solutions, support and guidance. Already, over 500 text focused business executives have attended our past Summits, where they were able to identify the most promising applications, size up technical challenges, speak up on what they required from solutions and connect with the key text analytics experts.

Now we are back by popular demand for a third sell-out event. Which means that – if you're quick – you'll be able to find out how far the technologies have come... where they're about to go, and what's next for text... plus how to solve the challenges that remain for the unstructured data market. From technical hurdles such as striking the ideal balance between analytical convergence and usability, to market barriers like customer resistance and vertical variations, the **3rd Annual Text Analytics Summit** will equip you with the know-how and contacts to deliver a real difference to your bottom line.

"The text analytics summit in Boston was the best conference I attended last year"

Olivier Jouve, VP Text mining, SPSS



If you've attended our previous Summits, you already know that this is where America's pace-setting businesses put their heads together with today's leading developers to discuss how the technologies can be leveraged for maximum value... and this year we are adding even more:

- **Commercial Businesses:** This is a one-of-a-kind opportunity to examine the latest capabilities, compare and contrast the technologies and meet with global market leaders who will show you how to extract maximum value from your organization's idle textual data. If you already operate within the unstructured data market, this really is an opportunity to secure your lead in the space. If you're still weighing up the options, it's a chance to help you catch up on the ground you're already losing.
- **Solution providers:** Come and deepen your understanding of market push and pull factors, what is driving customers towards particular products and what they want from your solutions. You'll leave the Summit with the insights and market intelligence you need to market your fantastic technologies and their capabilities to an even wider audience. Plus you'll meet a whole stack of commercial end-users at the show as well.

"Past Summits brought together vendors & end users to talk text technologies & how to apply them. Now with text analytics booming, this year's Summit will be the best ever!"

Seth Grimes, President, Alta Plana

With a special focus on the industries that are **REALLY** testing solutions – come and hear from world leading organizations, and apply what they've learned to your own professional concerns

★ **Government:** Discover the very latest deployments in intelligence, knowledge management and discovery, predictive analytics, security and more as used by the US Government. Walk away with inside knowledge, new techniques and insight to apply to your business

★ **Pharmaceuticals & Bioinformatics:** Get a full update on how pace-setting pharma outfits are using text mining to boost their bottom line. Hear from leading pharmaceuticals companies on how they deploy text mining applications that extend way beyond drug discovery and patent mining

★ **Media & Publishing:** From data storage and retrieval to competitive intelligence and CRM, find out how leading media companies such as **Dow Jones** are using text technologies to strengthen their bottom line

★ **Enterprise:** How are **Yahoo! Inc., Thomson Learning, Whirlpool** and **Jet Blue Airways** deploying text mining solutions to business processes such as CRM, knowledge management, business intelligence, content management, reputation management, human resources, staffing & more?

Plus the impact on **financial services, research, education, air services, legal** and others. New case studies... do's and don'ts... plenty of time to ask questions... this is knowledge and information you simply won't get from any other source.

Hot topics that address the issues that are right at the top of your corporate agenda, and enable you walk away with tomorrow's text analytics knowledge – before your competitors even know what's available!

We spend months interviewing dozens of key people to ensure we create a summit agenda that absolutely reflects your professional concerns. This means the summit addresses your biggest challenges, including;

- Integrated text analytics with BI
- Survey Analysis
- Web 2.0 / the Semantic Web
- The role of Text Mining in search
- Email archive mining
- New verticals, new applications
- The future of Text Analytics

Take a look at the full Agenda and you'll see exactly how useful this event will be.

Over two exciting days you will:

- **Evaluate** the real world front line results from the biggest names deploying text analytics technologies today. No academic jargon, or tired old PowerPoints. Instead, knowledge you can put to good use, as soon as you're back in the office

**EARLY BIRD DISCOUNTS! CALL NOW
1 800 814 3459 AND SAVE \$200**

**Spaces sold out
in 2005 & 2006!
Make sure you
don't miss this
must attend event
in 2007!**

Register now at www.textanalyticsnews.com

"The text mining market now has up to \$100m annual product revenue - & is growing at 40% to 60% a year"

Curt Monash, leading analyst of and strategic advisor to the software industry

- **Explore** the successes and the pitfalls of text mining – expert speakers will guide you through the capabilities and restraints of text analytics technologies
- **Network** with senior executives from a host of organizations that already deploy text analytics, find out how they use the technologies, the results they're enjoying, and the ROI's they currently experience
- **Assess** the very latest solutions from a host of leading solution providers and take a sneak preview into the future of the technologies.

New for 2007!

- ★ **14+ hours** of excellent networking opportunities
- ★ **Live testing & demonstration forum:**
A key addition to 2007 – take notes on live mining techniques and applications that will deliver the tools to save your company money
- ★ **Live on-site text mining exhibition:**
Get hands on experience of how the tools work for your business
- ★ **Industry specific interactive roundtables:** Exclusive roundtable discussion sessions concentrating on the challenges that matter to you most
- ★ **Top level keynote sessions:**
Led by Ramana Rao, Hadar Shemtov & more
- ★ **250+ senior level attendees:**
Our biggest Text Analytics Summit so far!
- ★ **New customer case studies:**
New users, new industries, new opportunities...

Go to the event website at
www.textanalyticsnews.com/usa
to get the latest updates on this dynamic event!

Whether text analytics is already key to your business strategies, curious about the capabilities of emerging tools, or even a market leading vendor, there is no better forum in North America to learn more and maximize your potential gains.

Already, we have signed up the biggest names in the industry, and organizations including **Whirpool, Dow Jones, Jet Blue Airways** and **Thomson Learning** will guide you through case studies, lead thought-provoking discussions and answer all your questions.

Equally, cutting edge solution providers will take you through the advancements in technology and the capabilities of tomorrow. You will hear from **IBM, Clarabridge, Anderson Analytics, SPSS, Attensity** and more as they reveal new capabilities and uncover new technologies.

The 3rd Annual Text Analytics Summit 2007 will deliver exclusive keynote presentations, mindshare panel sessions, real life case studies, new industry breakout roundtables and interactive workshops where we want you to make your voice heard.

Remember, the Text Mining 2005 and Text Analytics 2006 summits were both sell out events, this year will be no exception! Make sure you're in the conference hall and not on left on the waiting list. We strongly recommend immediate booking.

Don't forget group bookings get big discounts... bring along your whole team and save big bucks – call us today to find out how! We look forward to seeing you in Boston, 12-13 June 2007.



Solution Providers STOP!

Do you offer text mining solutions to organizations across multiple industries to help them extract real value from unstructured data?

Do you crave the opportunity for meetings and doing business with key decision makers from the biggest end users deploying text analytics solutions today?

... then come to this event to win new customers from and secure sales from new industries!

The 3rd Annual Text Analytics Summit 2007 offers you the unique opportunity to maximize ROI for your 2007 marketing spend. At an event packed with senior level executives in the unstructured data space, you can consolidate your position as a market leader through our bespoke sponsorship packages. Contact Text Analytics News now to take advantage of this unique opportunity.

- Brand visibility
- Exhibition stands
- Networking party
- Branded seat covers
- Advertising
- Delegate pack inserts and seat drops
- Pre- and post conference publicity opportunities
- Conference stationary sponsorship
- Report and whitepaper sponsorship
- All new live text mining demonstration sessions

We will tailor a sponsorship package to your budget and requirements!

Contact us now to discuss your sponsorship package.

Call: 1 800 814 3459 (freephone US)
or +44 (0) 20 7375 7575 (worldwide) to speak to the Text Analytics News team.

Email: info@textanalyticsnews.com

**Sponsorship
has SOLD OUT
2 years running...
OPPORTUNITIES
ARE LIMITED...
Call Today!**

Exhibition Zone Get your technologies seen!

Where can I find new customers for my text mining solutions? Which emerging industries will offer my business real lucrative opportunities? Which partnerships will allow me to take my solutions forward and add significantly to my bottom line?

In 2005 we came together to discuss the technological landscape, size up the capabilities and assess text mining applications. In 2006 we gathered to hear from the real early adopting end users deploying text analytics solutions across various industries. 2007 will bring you even more end users from a wider spectrum of industries. There really is only one place to capture the attention of your potential customers!

A select number of vendors, consultants and solution providers will have the opportunity to create a "business base" for themselves at the 3rd Annual Text Analytics Summit 2007.

Your exhibition area gives you an unrivalled opportunity to:

- Provide exposure for your market-leading solutions
- Demonstrate your technology to attentive and curious end users
- Network with potential partners
- Form relationships with senior level stakeholders from right across the text mining business chain

Exhibition space is already selling out fast and will be allocated on a first come first served basis... 2005 and 2006 sold out – make sure you don't miss out in 2007!

For more information and to view our Sales and Marketing option pack contact us today.

Call: 1 800 814 3459 (freephone US & Canada) or

+44 (0) 20 7375 7575 (worldwide) to speak to the Text Analytics News team.

Email: info@textanalyticsnews.com

**Exhibition
spaces are
limited and will
sell out fast...
Make sure you get
your technologies
seen!**

The biggest names in text analytics attending...?

The 3rd Annual Text Analytics Summit 2007 is the premier forum for meeting the leading minds in global text mining strategy. At no other time and in no other place can you find such a high concentration of text analytics know-how and innovative ideas in one venue!

Solution providers – will be giving you practical hints for creating bottom line impacts within various markets.

Tech-savvy end users from newly emerging verticals such as Air Travel, Entertainment & Media, News & Research, Finance, Customer relations, Call centers and more!

**See A List Of Last
Years Attendees
Online Right Now:
[www.
textanalyticsnews.
com/usa/
attendees.shtml](http://www.textanalyticsnews.com/usa/attendees.shtml)**

- Software developers and users in content management, search, BI, competitive intelligence, security, CRM, KM, KD, BPM, patent, warranty, predictive analytics, reputation management, HR, bioinformatics, cheminformatics and matinformatics
- Technology and ontology providers
- CKO
- CIO
- CTO
- Enterprise Information Architects
- Data Storage and Management Professionals
- Data Storage and Management Solution Providers
- Knowledge Management Professionals
- Data Engineers
- Industry Analysts
- Consultants
- Venture Capitalists

"Most of your business data is stored as text. This is the text analytics conference of the year!" Olivier Jouve - SPSS

Visit www.textanalyticsnews.com for the latest agenda and speakers

SAVE \$200

If you register before April 27

NEW FOR TEXT ANALYTICS IN 2007!

LIVE DEMONSTRATION SESSIONS

ALL NEW!

Make the most of the exclusive live onsite demonstrations taking place throughout both summit days. This is your only opportunity to see a myriad of live mining suites at work... Saving you time & money!



PRE SUMMIT INTERACTIVE WORKSHOP

ALL NEW!

Get a masterclass on the latest Text Technologies at our pre summit workshop. For more information on who will be leading you through their solutions **go to www.textanalyticsnews.com/usa NOW**



INDUSTRY SPECIFIC ROUNDTABLE DISCUSSIONS

UP-DATED!

A selection of small, topic focused interactive discussions designed to enhance your practical know how, learning and contacts in the text analytics space.

Industries to be discussed include: **Government; Enterprise; Biotech; Pharma; Life Sciences; Survey**

Analysis; Media; Publishing; Call Centers; CRM

NEW AGENDA, NEW SPEAKERS, NEW SESSIONS, NEW FOR 2007!

Hadar Shemtov, *Director of Research, NLP, Yahoo! Inc.*

Ramana Rao, *CEO, iCurrent Inc. (Founder Inxight Software, Inc.)*

Bryan Jeppsen, *Research Analyst, Jet Blue Airways*

Mark Frazee, *Manager Performance Measures, Consumer & BI, Whirlpool*

John Joyner, *Director of Indexing Automation & Data Management, Thompson Learning*

Simon Bradstock, *VP Corporate Products, Dow Jones*

Dr. David Bean, *Founder & CTO, Attensity*

Chris Jones, *Manager of Analytics, Intuit Inc.*

Dave Kellogg, *President & CEO, Mark Logic*
Michelle de Haaff, *VP Products & Marketing, Attensity*

Seth Grimes, *President, Alta Plana*

Breck Baldwin, *President, Alias-i*

Tom Anderson, *Founder & Managing Partner, Anderson Analytics*

Justin Langseth, *President, Clarabridge Inc.*

Chalapathy Neti, *Sr. Manager, IBM Research*

Olivier Jouve, *VP Market Strategy, SPSS*

Ponani Gopalakrishnan, *Director, IBM SWG Discovery Solutions*

Chris Bowman, *Supervisor of Federal Programs, Lafourche Parish Public Schools*

Curt Monash, *President, Monash Information Systems*

Dr. Nicholas Kushmerick, *Chief Scientist, QL2 Software*

Jonathan Stern, *CEO, Zoom Information Inc.*

This list is growing... for a full update on who has just joined the program see: www.textanalyticsnews.com/usa/speakers.shtml

EXHIBITION ZONE



Meet today's leading solution providers as they take you through the latest cutting edge technologies.

- Find out which technologies are best for your business
- Speak with the experts who will make sure you get it right – first time!
- View the very best offerings from the very best solution providers...

Attensity, Anderson Analytics, Clarabridge, Inxight, Nstein, SPSS and many more...

MEET THE BIGGEST NAMES IN TEXT ANALYTICS

Go to www.textanalyticsnews.com/usa/attendees.shtml to see a list of last years attendees!

From C-level decision makers, Directors, Presidents, and Management – there is no other place with such a high concentration of text analytics know-how and spending power in one room!

REGISTER NOW!

NETWORKING COCKTAIL PARTY

Text Analytics News invites you to join all attendees as they wind down from a day of insightful discussion, strategic debate and amazing networking for an evening of reflection.



"The text analytics summit in Boston was the best conference I attended last year!"

Olivier Jouve,
VP Text mining,
SPSS



GET MORE OUT OF YOUR TEXT, DRIVE UP YOUR PROFITS AND ADD TO YOUR BOTTOM LINE IN 2007!

DAY 1: Tuesday 12th June

Chairman's Welcome & Opening Speech

What's next for text?

Summit Chair Seth Grimes will present a brief history of text analytics, assess the state of the technology and the marketplace, and forecast development directions to understand "What's Next for Text?" Seth will peer into the text-analytics crystal ball to answer key questions:

- What are users looking for... and what will they find?
- Why isn't everyone using text analytics? They should be, right?
- What are the tech wizards cooking up?
- Who's in and who's on the way?

Seth Grimes, *President, Alta Plana Corporation*

The Next 20 years: Reaping the benefits of reaping text

keynote session

- Discover the role of text in large complex systems that leverage human and machine capabilities
- Find out the importance that Text Analytics plays in systems that Retrieve, Extract, Arrange and Present. What will such REAP systems allow you to do with your data?
- Take away a clear understanding on the applications of REAP systems - Discovery, Derivation and Delivery - and make sure you are ahead of the game.

Ramana Rao, *CEO, iCurrent Inc.*
(Founder, Inxight Software Inc.)

Industry Technology Power Panel

The world's leading minds from the unstructured text analytics space will discuss the very latest issues, the forthcoming challenges and what is driving the ongoing need for text-focused products.

★ **Technology Update:** Get up to date on the very latest technological capabilities and ensure you're ahead of your competitors as this industry leading panel covers the latest from text mining solutions

★ **Technology Debate:** Which technologies are the best? Why is it so difficult to produce a converged solution? Where will the most practical applications be for future breakthroughs in the technology?

★ **Technology Forecast:** Where will the capabilities be next year? How quickly are solutions being out-dated? What do I need to know about tomorrow's technology to make my decisions more effective today?

Olivier Jouve, *VP Text Mining, SPSS*

Ramana Rao, *CEO, iCurrent Inc.* Hadar Shemtov, *Director of Research, Natural Language Technologies, Yahoo! Inc.*

David Bean, *Founder & CEO, Attensity*

Moderator: Seth Grimes

How text mining transformed Intuit's voice of the customer analytics

joint case study

- Find out how Clarabridge's Content Mining Platform™ has helped Intuit move from labor intensive manual categorization to automatic categorization of a larger set of data
- Discover techniques that will help you go from processing a small percentage of internal data to analyzing 100% of internal and external data
- Understand the advantages identifying and responding to customer needs and wants and easily incorporate all VOC data into customer service and product development

Justin Langseth, *President and CTO, Clarabridge, Inc.*

Chris Jones, *Manager of Analytics, Intuit, Inc*

Nuance in voice of the customer: Little words make all the difference

Enterprises have spent millions of dollars capturing, storing and maintaining customer data.

- Realize the greatest threats to customer satisfaction and the tools that go far beyond simply leveraging structured information
- Learn the key steps in the customer feedback and analysis process from innovators in the field of Text Analytics and apply the lessons to your business

Dr. David Bean, *Founder & CTO, Attensity*

Unified Analytics and IBM's Omnifind Analytics Edition

- Take an exclusive look at IBM's work on "Unified Analytics" and how this remarkable project aims to add a solution to the ongoing textual, image and voice convergence challenge
- Find out more about IBM's brand new Omnifind Analytics Edition based on the work carried out on the open UIMA standard and what this technology means for you
- Discover the capabilities of converged solutions and what they mean for the future of text analytics technologies

Chalopathy Neti, *Sr. Manager, IBM Research*

Ponani Gopalakrishnan, *Director,*

IBM SWG Discovery solutions

Whirlpool - Ahead of the curve

end user case study

Are you taking advantage of the millions of touches you have every year with your customer? Text analytics is becoming more than an early warning system for manufacturers such as Whirlpool.

- Discover how Whirlpool extracts insights from millions of unstructured documents to learn the expectations and desires of customers that would otherwise go unnoticed
- Learn Whirlpool's process to parse and process information and get inside the minds of customers few other companies can truly comprehend

Mark Frazee, *Manager Performance Measures, Consumer and Business Intelligence, Whirlpool*

Transforming the future of search with visualization and navigation

end user case study

- Find out how search can help build a growth-oriented business simply by decreasing the time employees need to find information to make quality business decisions
- Discover how vital information embedded in users search results can be used to take them beyond basic search and see the emerging trends and issues that can affect your business decision making process
- The future of search in the workplace and how text mining, visualization and discovery will become a basic expectation of future searchers to generate relevant and meaningful search results

Simon Bradstock, *Vice President Corporate Products, Dow Jones*

Commercial End User Panel

Gain exclusive insight from the all important end user perspective as they discuss how text mining technologies are being deployed across their industries, the benefits they have seen and what they want from solutions over the next 12 months

★ **Your customer:** What are the most important functions of text mining solutions today? What could be mined more effectively? How are text analytics solutions selected?

★ **Exclusive customer mindshare:** How do applications differ across various industries? How can current solutions better meet the needs of the end user? When will vendors offer converged solutions? Where else can end users apply the technology?

★ **New industry customers:** Who is using text analytics technologies today? Who are the "intrigued" customers and which vertical markets do they represent? Who are the newest text analytics end users and what do they need from you?

Bryan Jeppsen, *Research Analyst, JetBlue Airways*

Mark Frazee, *Manager Performance Measures, Consumer and Business Intelligence, Whirlpool*

Chris Bowman, *Text Analytics Evangelist, Lafourche Parish Public Schools*

John Joyner, *Director, Indexing Automation and Data Management, Thomson Learning*

Moderator: Seth Grimes

Rescuing "lost" data - using text mining to apply thesaurus-based indexing to digitized print material

end user case study

Thomson Scientific is one of the world's largest providers of scientific information.

- Discover how Thomson Scientific have deployed TEMIS Text Mining solutions to index the BIOSIS Digital Archive (over 2 million scientific documents!)
- View the wide range of term types from organisms, chemicals, diseases, geographical locations and more Thomson have mined to uncover jewels of information
- Learn how Thomson Scientific benefit from harnessing the ability to process over 2 million terms of complex thesauri through the very latest TEMIS Luxid software

Mirko Minnich, *Vice President & Chief Architect, Thomson Scientific*

The only roadmap you need to increase profits from Text Analytics

DAY 2: Wednesday 13th June

Chairman's Opening Remarks

Seth Grimes, *President, Alta Plana Corporation*

What Can We Learn from Terabytes of Text: A View from a Search Engine

keynote session

- Learn about the role of text mining techniques in search applications
- Discover how information extraction and acquisition of lexical knowledge benefit the operations of a search engine
- Appreciate the potential and challenges of applying text mining on a very large scale!

Hadar Shemtov, *PhD, Director of Research, Natural Language Technologies, Yahoo! Inc.*

JetBlue's flight into Text Analytics

end user case study

- Discover how JetBlue Airways manage customer feedback and how Text Analytics fits into the plan to maximize returns on their data
- Follow the steps JetBlue Airways take to evaluate vendors, select a partner and implement a solution
- See the results for yourself and find out how Jet Blue will be using the latest technologies to get what they want from their unstructured data

Bryan Jeppsen, *Research Analyst, JetBlue Airways*

Topical categorization tools and services vendors... The publisher's perspective

end user case study

See what is at the top of the agenda for publishers when they deploy text mining technologies.

- Discover how current text mining technologies assist Thomson Learning in the support of controlled vocabulary, the tweak-ability of tools and the difficulty of managing training sets vs. difficulty of maintaining rules
- Find out how Thomson Learning has worked with rules-based vendors and view the pros and cons from their perspective.
- What are the pros and cons of current statistical tools and the service vendors that use various tools?

John Joyner, *Director, Indexing Automation and Data Management, Thomson Learning*

Reserve your place today!
call 1 800 814 3459

Expert Analyst Panel

Join leading analysts as they discuss the most pressing issues affecting the text analytics space today. This is your unrivalled opportunity to hear from the people who sit on the fence.

- ★ **Convergence of data & text:** What benefits will it really bring the industry? Who will be the first users? Who will offer the technology? How soon will it become available?
- ★ **Search & Mining:** How much of an importance does search play in text mining? Can search technologies be used to increase the capabilities of text mining?
- ★ **Applications:** Which applications are being created for text analytics technologies? How will businesses benefit? What is next for the future of text technologies?

Philip Howard, *Research Director, Bloor-Research*

Moderator: Seth Grimes

If you are an analyst and would like to join this panel please call: 1 800 814 3459 ext 220 – NOW!

Turning online voc into actionable insights: using text mining for web content analysis. (A travel industry case study)

case study

- Learn how powerful software can be leveraged to gain immediate insight for your marketing department
- Discover how Text Analytics can extract opinions and sentiments from Blogs and ensure you can do the same
- Understand the process of a web project from web-scraping, data preparation, coding, and data mining through to analysis
- Understand how web content analysis may help you gain the information advantage

Tom H. C. Anderson, *Founder & Managing Partner, Anderson Analytics, LLC*

Olivier Jouve, *Vice President Market Strategy, SPSS, Inc.*

Text Analytics: The Next Killer Application Enabler

case study

This session will highlight three real-world case study applications that utilize text analytics

- Discover how text analytics, when combined with federated search and visualization techniques, provides real-world benefits to enterprises of all kinds, from governments to financial services to manufacturing
- Understand how text analytics turns commonly used business intelligence, collaboration and search software into killer applications
- Hear from an industry leader how Inxight's text analytics solutions have improved its responsiveness and competitiveness

Christopher Riopel, *Director of Product Management, Inxight Software*

Understanding not indexing the internet

ZoomInfo, a semantic search engine, uses its patented Natural Language Processing algorithms to understand and organize the business web

- Discover how ZoomInfo's patented technology works and how it is able to understand the web
- See how this technology successfully combines search and text mining technologies to offer a unique search engine
- Find out how search is changing based on the needs of the users and technology

Yonatan Stern, *CEO, ZoomInfo*

Industry Marketing Panel

Drive your key marketing messages to end users across a multitude of industries. Our expert marketing panel will share their thoughts on how to achieve the best communication for your target audience

- ★ **Marketing via application:** Which application segments are the biggest? Which are growing the fastest?
- ★ **Know your product:** How is the "whole product" shaping up in each market sector you pursue? What combination of your technology, your services, partner technology, and partner services is best meeting the customer's needs?
- ★ **Understand your customer:** Which marketing pitches do customers, prospects, and influencers agree with? Which true ones can you make that you feel customers would not buy into? And what do IT departments think about text analytics?

Dave Kellog, *President & CEO, Mark Logic Corporation*

Michelle de Haaff, *VP Products & Marketing, Attensity*

Moderator: Curt Monash, *President, Monash Information Services*



Register now at www.textanalyticsnews.com



3RD ANNUAL TEXT ANALYTICS SUMMIT

3 EASY STEPS TO REGISTER NOW!

1. Your Choice of Registration Package

Please tick the package price box you require below:

| | DISCOUNTED BOOKINGS BEFORE | | STANDARD PRICE |
|---|--|--|--|
| | 16 MARCH 2007 | 27 APRIL 2007 | |
| 2-DAY PLATINUM PASS Full access to all sessions, workshops & exhibition, cocktail party, full audio & visual CD, post-conference access to all data & presentations | \$1,495 <input type="checkbox"/> | \$1,595 <input type="checkbox"/> | \$1,795 <input type="checkbox"/> |
| 2-DAY GOLD PASS Full access to all sessions, workshops and exhibition, cocktail party, post-conference access to online presentations | \$1,395 <input type="checkbox"/> | \$1,495 <input type="checkbox"/> | \$1,695 <input type="checkbox"/> |
| END USERS PASS AVAILABLE TO END USERS ONLY - ALL END USERS WILL BE VERIFIED. Full access to all sessions, workshops and exhibition, cocktail party | \$795 <input type="checkbox"/> | \$895 <input type="checkbox"/> | \$1095 <input type="checkbox"/> |
| CONFERENCE CD | <input type="checkbox"/> \$300 | | |

Date & Venue

June 12 -13th, 2007
The Boston Marriott Newton, USA

Hotel Discounts

We have negotiated a special room rate at the Boston Marriott Newton. Reservation and price details will be sent to you when you register.

Group Discounts

Take advantage of Text Analytics News' unique group discounts. The more people you bring, the more money your company saves!

Contact the Text Analytics News team on 1 800 814 3459 or +44 (0)20 7375 7575 or email info@textanalyticsnews.com for more details

2. Delegate Details

Please photocopy this form for multiple registrations

Customer Priority Code:

Mr / Mrs / Ms / Dr: _____ First name: _____

Last name: _____

Company: _____ Position/Title: _____

Telephone: _____ Fax: _____

Email: _____

Address: _____

Zip: _____ Country: _____

3. Payment

Choose one of the following payment options:

I enclose a check/draft for: \$ _____
(payable to First Conferences Ltd)

Please invoice my company: \$ _____
Purchase Order No.: _____

Please charge my credit card: \$ _____
 Amex Visa Mastercard

Credit card number: _____

Expiry date: _____

Name on card: _____

Signature: _____

NB: FULL PAYMENT MUST BE RECEIVED BEFORE ACCESS TO THE SUMMIT

5 Easy Ways to Register today!

- **Call** Text Analytics News on 1 800 814 3459 or +44 (0)20 7375 7575
- **Fax** This form to 1 800 814 3460 +44 (0) 20 7375 7576
- **Emai** The TextAnalyticsNews registration team on register@textanalyticsnews.com
- **Mail** This form to TextAnalyticsNews 7-9 Fashion Street, London, E16PX, UK
- **Online** Register your details securely at <http://www.textanalyticsnews.com/USA/>

Cancellation Policy

Places are transferable without any charge. Cancellations before 25 May 2007 incur an administrative charge of 25%. If you cancel your registration after 25 May 2007 we will be obliged to charge the full fee. Please note – you must notify Text Analytics News in writing of a cancellation, or we will be obliged to charge the full fee. The organizers reserve the right to make changes to the programme without notice. All prices displayed are exclusive of vat (value added tax) unless otherwise stated but, vat will be charged, where applicable, at the prevailing rate on the invoice date and the relevant details will appear on the invoice. Please see terms & conditions on <https://secure.firstconf.com/textminingsummit/usa/terms.htm> for more information

Designed by The No Nonsense Group - www.tnng.co.uk

Fax this form to 1 800 814 3460

TEXTANALYTICS
NEWS.COM

Back by popular demand for its
third blockbusting year!

SAVE
\$200

If you register
before April 27

3rd ANNUAL

TEXT ANALYTICS SUMMIT



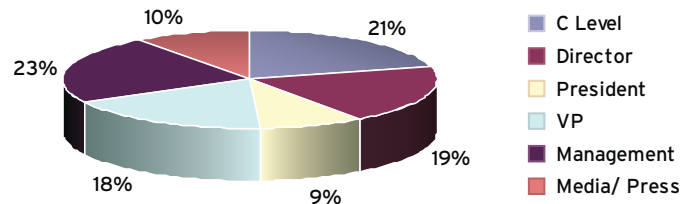
12 - 13 June 2007 • Boston Marriott Newton • USA

Discover lucrative new Text Mining markets, hear from the biggest commercial end users and analyze the leading applications that will increase your profits

Take advantage of new opportunities in the exploding US text mining market and meet cutting edge end users

Who will you meet at the 3rd Annual Text Analytics Summit 2007?

Text Analytics Summit 2007 is the annual think tank for C-level decision makers, Directors, Presidents and Management in the unstructured data space. Take a look at the make up of our past summits and see who you will meet this year!



10 Great reasons for you to attend!

- 1) Independent forums:** Make sure you receive a well-balanced, innovative and informative briefing to enable you to make the best decisions for your business – without the sales pitches!
- 2) Profit from text:** This is your opportunity to listen to, network with and learn from the best. Where else would you discover the answers to the most important text focused questions today?
- 3) Exclusive Text Analytics case studies:** Learn from end user experience, take away best practice tips and ensure your business is a market leader.
- 4) 14+ Hours of unprecedented networking opportunities:** The perfect opportunity to network with your industry peers who are facing the same complex challenges as you.
- 5) In-depth two day briefing:** The 3rd Annual Text Analytics Summit has the major players from every sector within the industry ensuring your business gets the best, most timely and detailed analysis of the market place available
- 6) Group and super-early bird discounts:** Register early for the event and we'll give you a \$300 discount off your conference ticket. Come along with your team and make more substantial savings on your ticket prices.
- 7) High quality speakers:** Our dynamic speaker line up will share their wealth of experience, knowledge and passion with you and ensure that you profit from their vision and take away superb best practice strategies.

- 8) Industry break-out roundtables:** More focus, more attention and more intimate networking. Designed for the ultimate in industry specific mind-share...

- 9) Live mining demonstration sessions:** Watch the technologies at work to see how they can suit your needs and see the full range of capabilities available today... Our summit exhibition ensures you get to test out today's latest technologies!

- 10) Lets talk text over a drink...** The perfect environment for networking Text Analytics News is proud to bring you our annual Text Analytics Industry Cocktail Party.

ALL NEW
FEATURES
FOR 2007

5 Easy Ways to Register Today



www.textanalyticsnews.com



bookme@textanalyticsnews.com



1800 814 3459 or +44 (0) 20 7375 7575 (worldwide)



1800 814 3460 or +44 (0) 20 7375 7576 (worldwide)



TextAnalyticsNews, 7-9 Fashion St, London, E1 6PX, UK

Make a significant difference to your bottom line with Text Analytics... Open NOW!