European Text Analytics

A white paper prepared for

European Text Analytics Summit 2007

Amsterdam, April 26-27, 2007

(www.TextAnalyticsNews.com/europeo7)



Seth Grimes

Alta Plana Corporation

Alta Plana

TABLE OF CONTENTS

I	ln'	TRODUCTION: EUROPE 2007	3
2,	TE	CHNOLOGY BASICS	4
		CHNOLOGY ADOPTION	-
-		JROPEAN INNOVATION	-
•		RESEARCH CENTERS	
	•	Innovative Projects and Products	
		Technology and Service Vendors	
		JROPEAN OUTLOOK	-
_	5.1	THE MARKET FOR SOFTWARE AND SERVICES	
	•		
	5.4	Market Estimation	, 1C

INTRODUCTION: EUROPE 2007

The interplay of languages and cultures drives European social and business interactions to an extent not found in any other region. For Europe, written and spoken words both unify and divide. They are key both to maintenance of national identities and to erasure of artificial barriers to social and economic integration, whether within official bodies such as the European Union, in cross-border business transactions, or in personal encounters.

Europe has a deep and rich scientific, literary, and legal heritage and a tradition of strong governmental support for national and cross-cutting pure and applied research. Given this background and given the role of multilingualism in everyday life – the European Union alone counts 27 member states and 23 official languages – it is no surprise that Europe has nurtured innovation in computational linguistics and similar language technologies.

European innovation has emerged as state-of-the-art text-analytics software solutions. These solutions compete internationally. Europe's contribution to text analytics is quite out of proportion to the continent's size and population, driven by national and international imperatives keenly felt by institutions and individuals.

This paper's author has in the past characterized text analytics as "technology and process both, a mechanism for knowledge discovery applied to documents, a means of finding value in text." Text analytics can be further described as facilitating machine processing of human communications. It enables a more natural human-computer interface. Now that computers are ubiquitous, text analytics offers enormous potential to transform interactions among individuals, business, and governments.

Text analytics opens new vistas for search, the Web's "killer app," and creates new categories of software for fields as diverse as biomedical research, intelligence and law enforcement, corporate operations and communications, and legal and intellectual-property analysis. Europe has led the way, not only in creation of basic text technologies, but also in their application to these and many other important, real-world problems.

This paper will describe text analytics applications of particular European interest. It will survey a number of technologies that originated in Europe. It will present examples of innovative European research projects, programs, and centers and will forecast coming technical and market developments.

As will be seen, the 2007 European text analytics scene is dynamic and robust. European text analytics is holding its own in the international research community and the global marketplace and is evolving to meet rapidly growing European and international demand.

2 TECHNOLOGY BASICS¹

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.

Multilingual intake presents special challenges due to the wide variety of character sets, word morphologies, grammars and syntaxes, vocabularies and idioms, and cultural contexts from which the spectrum of human languages are built. Challenge, however, creates opportunity.

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. 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 mining from those latter technologies.

Because text analytics looks at document sets and identifies interdocument relationships, it supplies context that enables far greater relevance to 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-mining 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.

Material in this section is drawn from <u>The Developing Text Mining Market</u> by Seth Grimes, Alta Plana, (altaplana.com/TheDevelopingTextMiningMarket.pdf), 2005.

3 TECHNOLOGY ADOPTION

Text, as written language, is closely linked with culture. This linkage creates a unique challenge for software that would automate text processing. The software must meet general, global business needs while simultaneously handling narrow, localized concerns particular to given languages and application domains.

Text analytics technical requirements include both internationalization and sufficient specificity to accept input in any human language (whether Western, Middle Eastern, African, or Asian) and input from both specialized domains such as biology, law, and manufacturing and from the Web. This need is typically met with modularized software architectures. Such architectures provide a generalized computing framework that accepts domain- and language-specific *plug-in* components and workflow capabilities that can be matched to application needs.

User adoption of text analytics, as of any specialized computing technology, varies greatly by region, business domain, and organizational profile. Uptake depends on perceptions of need, on availability of solutions, and on the existence of research and business climates that encourage and embrace innovation. For text analytics, Europe meets the three uptake criteria in varying degrees.

European government and industry have long recognized the role that technology can play in bridging language divides to promote social and economic integration and to create opportunities for regional and international business. The European Union sponsors many research programs, as do national and local economic-development agencies. These programs typically involve academia, private-sector commercialization partners, and government adopters.

Given European societal goals, the availability of research-program funding, and Europe's many strong academic institutions, text-analytics research in European academia is on a general par with academic research elsewhere. The Internet has only widened strong collaboration among European and international academics.

Within the government sector, Europe has aggressively adopted text analytics to meet critical national-security and law-enforcement goals. Criminal and terrorist networks are as likely to exploit national borders as they are to find them an impediment, but they do leave an electronic trail. The information they generate, while subject to privacy laws, is easy to exchange, and text analytics is uniquely capable of the data intake, integration, and analysis needed to fight these threats.

Government-prompted adoption of text analytics in business settings – for intellectual-property management, corporate governance, legal discovery, and the like – does appear to lag uptake outside Europe. European business is not subject to requirements directly comparable to U.S. Sarbanes-Oxley corporate governance mandates (which, n.b., do apply to all corporations with 300 or more U.S. shareholders) and eDiscovery information-preservation rules. Such needs have sped U.S. adoption of solutions that automate text processing but have not applied as widely in Europe.

Lastly, while Europe boasts world-class industrial computing-research centers, as well as start-ups that play in the text-rich, exploding social Web, corporate line-of-business adoption in Europe also appears to lag uptake elsewhere. The level of European interest appears very high however, particularly among media companies, publishers, and organizations conducting and consuming applied scientific research. Adoption rates should soon match international levels.

4 EUROPEAN INNOVATION

Europe boasts diverse and vibrant text-analytics research programs and centers, both governmental and academic. The European Commission sponsors noteworthy consortium and university initiatives throughout the continent. Other efforts are supported nationally and by industry. Many of these academic and industrial projects have been built out to create market-leading commercial software tools and services.

4.1 RESEARCH CENTERS

We will provide examples of notable European research centers that together hint at the breadth and orientation of efforts distributed around the continent.

Start with two Italian organizations: the Interuniversity Supercomputing Consortium, CINECA; and the Institute for Computational Linguistics of the Center for National Research, ILC-CNR. CINECA unites national-government and academic resources to further an applied-computing agenda in cooperation with industry, including IBM. It is comprised of 28 Italian universities and sponsors programs in many areas, not just text mining², serving as a "high-technology bridge between the academic world, research, and the world of industry and public administration." ILC-CNR³ in Pisa has a national and pan-European focus, having participated in dozens of European Commission projects. They consider language to be "the privileged medium for the interactions through which most of social, economical and cultural activities take place and its applications have a fundamental impact in many fields." The institute's work is in design and construction of corpora and lexicons and related ontologies, and in development of tools for Italian-language text processing, information extraction, and domain application.

The European Commission's Joint Research Centre (JRC) is also located in Italy and similarly sponsors text analytics activities. The **JRC** states that they "have been using Language Technology since 1998 to fight the *information overflow* and to overcome the language barrier." This work is performed under the auspices of the Information Society Directorate General within the JRC's Institute for the Protection and Security of the Citizen (IPSC). The JRC reports⁴ development of a language-technology toolkit with components for:

- Multilingual and cross-lingual retrieval of potentially user-relevant documents.
- Analysis of documents and information extraction with language-neutral representation of information.
- Visualization of documents and document-collection contents.

Their quite-interesting applications work includes the Europe Media Monitor NewsExplorer, which uses JRC-developed technology to automatically generate daily news summaries (press.jrc.it/NewsExplorer/home/en/latest.html).

Heading north and west, the National Centre for Text Mining, NaCTeM⁵, located

_

² CINECA text mining, see: www.cineca.it/gai/area/textmining.htm

³ ILC-CNR, see: www.ilc.cnr.it/indexflash.html

⁴ JRC Language Technology Activities, see: langtech.jrc.it/index.html

⁵ An overview of NaCTeM work may be found at www.ariadne.ac.uk/issue42/ananiadou/

at the Universities of Manchester and Liverpool, claims title to being the first publicly funded text mining centre in the world, supporting the UK academic community and participating in international research initiatives. NaCTeM was founded in the summer of 2004 and has focused initially on bioscience and mining of biomedical texts. They deliver analytical services via the TerMine term management system, AcroMine for expansion of Medline acronyms, and a TerMine integration with the Cheshire XML search engine; they provide and support software that includes toolkits for the University of Tokyo's GENIA software for information extraction from scientific literature.

Dr. Sophia Ananiadou of the University of Manchester explains that text mining is invaluable to systems biology researchers who need integrated approaches to generate hypotheses, and that the use of text mining technology is a must for facilitating scientific discovery given the amount of textual data generated every day⁶. She quotes an associate, Prof. Douglas Kell, who states that NaCTeM has tapped into this potential with great success:

Of particular interest to system-biology researchers is synonym detection and linking textual knowledge with existing databases. This problem has been addressed by the TerMine system, which automatically discovers important concepts in text. One of the most impressive outcomes of the work of NaCTeM are the systems MEDIE and InfoPubMed, based on the GENIA tools, which perform semantic text mining based on full parsing. These outcomes are important for the discovery and classification of novel protein-protein interactions, which cannot be detected by high-throughput methods.

As a last example of an academically centered initiative, we cite PASCAL⁷, Pattern Analysis, Statistical modeling and ComputAtional Learning, a Network of Excellence established under the European Commission's Sixth Framework Programme. (Framework Programme 7 started in 2007 and itself has quite interesting elements, for example, 4.2 Intelligent Content and Semantics⁸.) The PASCAL objective is "to build a Europe-wide Distributed Institute" to pioneer "core enabling technologies for multimodal interfaces that are capable of natural and seamless interaction with and among individual human users."

4.2 INNOVATIVE PROJECTS AND PRODUCTS

Europe is home to projects that have had very broad impact in the text-analytics world. Some of these projects are collaborative efforts that have produced and support software tools and resources that are used worldwide. Others are research efforts that have had considerable success as commercial products. Consider, in the latter category, Xerox's XeLDA⁹.

XeLDA, variously the Xerox Engine for Linguistic Dependent Applications or the Xerox Linguistic Development Architecture, is a multilingual engine for transforming, normalizing, and extracting information from text. XeLDA is the product of natural-language processing research at the Xerox Research Centre Europe (XRCE) in Grenoble, done in cooperation with the Palo Alto Research Center (PARC). XeLDA is the linguistic engine within software vendor TEMIS's Insight suite of text-analytics products, which integrates the engine

⁶ Ananiadou, S., Kell, D.B. and Tsujii, J. (2006) Text Mining and its Potential Applications in Systems Biology, in *Trends in Biotechnology (TIBTECH)*, vol.24(12), pp.571-579.

⁷ For PASCAL information, see: www.pascal-network.org/

FP7 information is at ec.europa.eu/information_society/istevent/koln2007/cf/network-detail.cfm?id=1081

⁹ For XeLDA background, visit: www.xrce.xerox.com/competencies/past-projects/platforms/xelda.html

within a framework that includes information-extraction, clustering, and classification tools. TEMIS further offers a product line called Luxid for annotation, integration, and analysis of information extracted from text. A variety of industry vertical applications take advantage of the products' modularized architecture, which supports use of plug-in "skill cartridges." TEMIS's presence is strongest in Europe, reflecting the company's origins, but is growing rapidly worldwide.

As stated, Europe is home to noteworthy collaborative projects. **NooJ** is a quite interesting even though it targets linguistics rather than broader text analytics¹⁰. According to author Max Silberztein, NooJ provides a development environment for "linguists who wish to formalize a set of linguistic phenomena (from the morphological level to the semantic level), corpus linguists who want to parse large corpora (e.g., for discourse analysis), or computer programmers who want to add NLP functionalities to their applications." Silberztein reports that NooJ, which is freeware, is used by a dozens of European research centers and companies.

With even broader capabilities and worldwide utilization, **GATE**^{II}, the General Architecture for Text Engineering from the Natural Language Processing Group at the University of Sheffield (UK), is made up of three elements:

- An *architecture* describing how language-processing systems are made up of components.
- A framework written in Java and tested on Linux, Windows and Solaris.
- A graphical development environment built on the framework.

A quotation provided by project lead Hamish Cunningham illustrates that free, open-source GATE is as much a philosophy as a technology. Cunningham states,

GATE has made it easier to do good science with repeatable experiments and standardized measurement. It also helps collaboration between researchers sharing software and data, and has shown how public money can have a beneficial impact on a research field and significant commercial interests by being open and, dare I say it, honest. One problem with the research landscape, particularly in fields related to AI, is that measurement and repeatability come a distant second to obscurantist marketing speak, and this is part of a general trend towards universities as sources of "product" to be sold to student or business "clients". We need to recognize the human beings are better at understanding what needs to be done than the blind blunt instrument of profitability, and restructure our scientific programs accordingly.

YALE – Yet Another Learning Environment – is another groundbreaking open source tool¹². The project is based in Dortmund, Germany, with contributors worldwide who have created a comprehensive knowledge-discovery environment supporting both data and text mining and a large variety of extensions that utilize a modularized, plug-in architecture. YALE project lead Ingo Mierswa and project initiator Ralf Klinkenberg have followed a path now familiar in the open-source world: They offer a commercially licensed version and related consulting services through a start-up company, Rapid-I.

¹⁰ For NooJ information, see: www.nooj4nlp.net/

II GATE pages: http://gate.ac.uk/

¹² YALE information is at: yale.sf.net/

4.3 TECHNOLOGY AND SERVICE VENDORS

A number of ground-breaking and market-leading text-analytics products have European roots. The previous section cites **TEMIS**, a French company that commercialized linguistic technology developed by Xerox and has built on a suite of analytical tools and industry plug-ins.

SPSS, a leader in the general analytics market with special strengths in market research, survey analysis, and predictive analytics, is another company whose text-analytics tools are of European origin. SPSS tools acquired those tools when it purchased a French company, LexiQuest, in 2002. LexiQuest's developed their natural language processing (NLP) technology with the support of the European Economic Community, now the European Community, and successfully licensed it to French agencies and prominent French and international companies.

SPSS's "predictive text analytics" technology folds LexiQuest tools into the company's Clementine data-mining workbench. These solutions allow users to analyze text-extracted concepts alongside data from other sources to create and solve integrated predictive models. The company states that the addition of attitudinal and interaction data discovered in text can provide *lift*, ranging from 10 to 50 percent, to results produced from only traditional sources. They claim over 2,000 text-analytics customers worldwide, many of them in Europe.

We next cite the two enterprise search market leaders, both European companies, Fast Search & Transfer (FAST) and Autonomy. Both companies go far beyond keyword search to support information extraction, content classification, document clustering, and creation and management of taxonomies: text-analytics functions. FAST was founded in 1997 in Oslo and has exhibited a very steep growth curve. Of late, FAST has been particularly aggressive in partnering with business intelligence (BI) vendors and in proposing a search-centric approach to BI, their Adaptive Information Warehouse Platform. Autonomy was founded just the year before and is headquartered in Cambridge. Their growth has also been rapid, fueled in part by acquisitions such as their 2005 takeover of enterprise-search rival Verity. They have a broad product line that includes tools for search of diverse types of media.

Lastly, we cite two European companies that analyze online social media. Both offer services rooted in text mining, delivering results to analyst users in packaged form. Brussels-based **Attentio** describes itself as providing European companies with market monitoring and analysis tools that continuously track a broad range of media for market intelligence about their brands, products, and fiercest competitors. Attentio services track newsgroups, blogs, Web sites that host reviews, and syndicated sources. They detect and report trends, clusters, and sentiments. And **Market Sentinel**, a London company, similarly monitors blogs and the Web for customers in a variety of sectors. One interesting point about these two companies is that they are selling solutions to end users – they are not selling technology.

European technology and service vendors have clearly found a ready market for products rooted in text analytics both at home in Europe and internationally. They are joined by foreign text-analytics software publishers who have established European offices and partner networks – Insightful, Inxight, Megaputer, Nstein, and SAS are among them – because they perceive the value and potential of the European market.

5 EUROPEAN OUTLOOK

5.1 THE MARKET FOR SOFTWARE AND SERVICES

Technology drives economic development, and involvement with emerging technologies like text analytics can have a multifaceted effect of boosting both the academic and industrial sectors and of benefiting both society and business adopters. Witness already-cited research initiatives as well as regional and localized programs. ITI Life Sciences' text-mining R&D program provides an example of the latter, a £5.3 million, three-year investment in construction of software for information extraction from scientific and medical literature. The University of Edinburgh (UK) and a subsidiary of U.S.-based biological- and chemical-information management company Cognia are conducting the work.

Biomedicine, manufacturing (e.g., quality monitoring), and governmental applications for intelligence, counterterrorism, and law enforcement derived the earliest significant benefits from text-analytics technology. The last two years, 2005 and 2006, saw increased uptake in areas such as survey analysis, media monitoring, patents, and publishing. Publishers are interested in generating content databases for data mining and in delivering increasingly sophisticated information-retrieval capabilities. Current growth areas include monitoring and analysis of networked social media for applications such as reputation management. We should also expect significant growth in Europe as elsewhere in the integrated analysis of text and numerical data, the embedding of advanced text-processing capabilities in line-of-business applications, and the practical emergence of question-answering as a high-end search product.

Introduction of text analytics for users in most segments will continue to involve vendor professional-services and consulting engagements as a high proportion of overall implementation cost.

5.2 MARKET ESTIMATION

The market for text-analytics software and services is substantial and growing rapidly. TEMIS cofounder Alessandro Zanasi of the Wessex of Institute of Technology (UK) offered in mid-2006 that "data and text mining is an expanding field and constitutes a market estimated to be more than US\$12 billion."

The estimate of the author of this paper is that the worldwide market for text-analytics software from pure-play vendors for licenses, support, and professional services plays out to about US\$200 million for 2005. Pure-plays include ClearForest, Inxight, Nstein, TEMIS, and Teragram. Annual growth is in the 25 to 40 percent range. We further allocate to text analytics a portion of overall revenue of UK-based Autonomy, SPSS, SAS, and other companies that sit partway in the text-analytics market, applying a conservative multiplier of estimated license revenue to compute user/contractor labor, and we add the assumed value of in-house text analytics at companies such as Factiva, Thomson, and Reed Elsevier. The sum is a US\$2 billion worldwide text-analytics market.

Europe is likely 30 to 40 percent of the worldwide text-analytics market. Interest is widespread in Asia but adoption lags that of North America and Europe. Therefore, the European demand-side (user) market proportion may dip in coming years, but Europe's share of the supply side of the market, or software sales and service delivery, should grow.

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 government statistical agencies, marketing firms, software publishers, and analytics users.

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, is available online at altaplana.com/TheDevelopingTextMiningMarket.pdf, and his 2006 summit paper, Finding Value in Text Analytics, is posted at altaplana.com/FindingValueInTextAnalytics.pdf.

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

EUROPEAN TEXT ANALYTICS SUMMIT 2007

The European Text Analytics Summit 2007 (www.textanalyticsnews.com/europeo7), slated for April 26-27, 2007 in Amsterdam, is a mindshare event for the leading developers, researchers, vendors, tech-savvy users, and newcomers to the textanalytics space.

It is the first European Text Analytics Summit, following on the heels of two highly successful Text Analytics Summits held in Boston in June 2005 and June 2006. Analyst Curt Monash wrote in *ComputerWorld* 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 2007 Text Analytics Summit is scheduled for mid-June 2007, again in Boston.

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 technologist 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



Real-world case studies: Hear from Telecom Italia, Merck KGaA, Philips, Pfizer, HP, the Dutch Police, the European Commission, University of Nebraska-Omaha, AstraZeneca and more!



THE European TEXT ANALYTICS SUMMIT

26th and 27th April 2007 - The Marriott Hotel, Amsterdam

Deliver the most effective text-based solutions, understand the needs of your customers and secure ROI in the exploding European market

The European text analytics market is exploding... Prepare your business to take maximum advantage Learn from more

- HEAR REAL LIFE CASE STUDIES from market leaders successfully using text technologies including Telecom Italia, Merck KGaA, the Dutch Police, Pfizer, Philips, AstraZeneca, Cisco Systems, University of Nebraska, Deloitte, the European Commission + many more!
- FIND OUT HOW TO TRANSFORM THE NEW EUROPEAN MARKET INTO NEW REVENUE. Identify which vertical industries and horizontal applications are ready for your business
- NETWORK WITH THE LEADING USERS and discuss the potential of text mining with key industry players from across Europe. Find out which technologies they require and how you can deliver them!
- EUROPEAN TECHNOLOGY UPDATE: Discover the latest developments in text analytics and assess the real needs of your text focused clients
- COMMUNICATION IS THE KEY! Find out where the booming markets and opportunities are and how to get your message to the right end users!
- LOOK INTO THE FUTURE Get the very latest thinking on what's next for text technologies moving forward and how your business can take the lead!

than 30 industry leaders including



























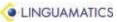












Summit Sponsors









Silver Sponsor



Official Media Sponsors





























"Text analytics technologies, more specifically enterprise search. have improved my work in ways I never thought possible. In fact, what used to take months, now takes iust minutes!"

> Dr. Burch T. Kealey **University of Nebraska**

Look inside for the complete 2007 European Summit agenda

THE European TEXT ANALYTICS SUMMIT

Be part of the world class event where Europe's most innovative businesses and leading edge solution providers join forces to focus on profit from text

Text analytics is rapidly becoming a powerful tool in the corporate armoury of market-leading companies

How so?

Reputation management. CRM.
Competitive intelligence. Fraud prevention.
Patent searching. Accelerated R&D.
These are all prime examples of business processes where text technologies are currently delivering tangible –and significant – ROI for savvy, early-adopter companies.

Companies have been swift to see how successfully they can maximise profits and help make major strategic decisions.

Small wonder then, that text analytics is zooming to the top of the corporate agenda in 2007. While equally, it is no surprise to discover that identifying the text analytics tools, applications and solutions that are potentially of greatest benefit can be a tricky, time-consuming process.

Happily, there's an excellent alternative:

Accept this invitation to take part in the inaugural European Text Analytics Summit to gain the knowledge and know-how you need to develop and implement an unstructured data strategy.

A BOLD PROMISE? IN JUST TWO FACT-FILLED DAYS YOU WILL PROFIT FROM:

- REAL LIFE CASE STUDIES FROM PACE-SETTING COMPANIES ACROSS EUROPE. Learn valuable lessons from their successes – and failures – the knowledge you need for your own benefit.
- DISCUSS COMMON CONCERNS, swap market intelligence and benchmark your own business with senior executives involved text analytics projects.
- ADDRESS YOUR CHALLENGES AND EXPLORE POTENTIAL SOLUTIONS with a variety of Europe's most advanced solution providers
- NETWORK WITH THE LEADING END USERS and discuss the potential of text mining with the key industry players across Europe. Find out which technologies they require and how you can deliver them.



THE EVENT IN THE US HAS SOLD OUT FOR TWO YEARS IN A ROW: Now fully updated and greatly expanded to meet the needs of businesses across Europe

North America has been leading the way in exploiting unstructured text mining. Two Summits – attended by over 400 senior executives – have been held already, and a third is planned.

The European Text Analytics Summit is Europe's only commercially-focused text analytics event. Take a look at the Summit Agenda, to see how we have focussed in on issues that are current, hot and relevant to your business needs.

We have spent months talking to everyone involved in the unstructured textual data space: businesses, solution providers, consultants and academics. Every presentation, case study and panel discussion has been specifically tailored to give you a thorough understanding of the systems you need in place to feature amongst the very best in your particular industry.

LEARN FROM THE BIGGEST AND THE BEST – and meet the people that really matter

Big businesses that are already making healthy profits from leveraging their data are already confirmed to speak at The European Text Analytics Summit– including Merck KGaA, Telecom Italia, Pfizer, Philips and Cisco Systems. Come and find out about their real life deployments – and the results they are achieving.

Additionally, you will hear from the University of Nebraska, and the Dutch National Police will discuss their 'digital washing machine' concept, that enables them to fight crime more easily by quickly extracting pertinent information from the millions of digital documents they hold.

Leading solution providers such as SPSS,Temis, Nstein, SAS, Basis Technology, ISYS Search, Linguamatics, Vivísimo and Infolutions will walk you through their new applications and technologies to make sure you stay ahead of the game. Furthermore, you will hear from NaCTeM, the UK National Centre for Text Mining, and the European Commission Security Research, which is dedicated to funding text analytics security projects.

It all adds up to a treasure trove of current, first-hand business and market intelligence that is simply not available from any other source.

APPLICATIONS AND TEXT TECHNOLOGIES IN ACTION:

What lesson have been learned, and what are the implications for your own business?

We know you want real world, real life information about the practicalities of text analytics. So that's what we've delivered:

Security Applications: Learn from the European Committee Security Research and the Dutch Police how your company can design effective technologies to combat fraud prevention, terrorist threat prevention and other types of corporate crime.

Multilingual Technologies: Examine the latest multilingual technologies – essential for Pan European business – including crossscript matching, cross- lingual referencing and language identification.

Pharmaceuticals: Find out first hand from **Merck KGaA, Pfizer** and **Astra-Zeneca** how indispensable text technologies are across the entire corporation – all the way from R&D to manufacturing, distribution and beyond.

Publishing: With advertisers allocating more and more of their budgets to online media, text technologies are helping publishers make a successful transition to the digital world. Find out how to attract more customers and generate new revenue streams.

Research: Text analytics plays a fundamental role in the facilitation of quick, relevant research. Discover how the University of Nebraska-Omaha reduced three months of manual search to just two manageable minutes using text analytics.

CRM: Discover how **Telecom Italia** is successfully deploying text analytics to its advantage, monitoring customer requests and needs to produce a more profitable sales and marketing strategy. You'll also discover how the company selected and implemented its text analytics tools.

Reputation management: Find out how Cisco Systems is using text analytics to monitor its reputation – and how mining the blogs, call centre notes, emails and internet can enhance your own business.

Human resources: How can text technologies be used to assist your HR needs? Leading solution providers will outline the massive gains to be made – provided you implement the correct tools.

Step by step, you will learn how to identify the most promising applications, size up the technical challenges, connect with tech-savvy trailblazers, and tap into the rich market potential

THE EUROPEAN TEXT ANALYTICS SUMMIT

will make a definite difference to your business' bottom line

This Summit will bring together tech-savvy businesses looking to increase their ROI from text analytics, organisations that are keen to get their text analytics projects off to a flying start, and Europe's most innovative solution providers.

END USERS

Register today

and save

REGISTER TODAY AND:

- Participate in three lively and interactive Panel Debates focused on business needs, technology and marketing.
- Get expert answers to your text focused questions from the industry leaders
- Take part in six industry-specific Roundtables that discuss challenges and opportunities
- Learn about the latest technological advances and emerging markets
- Network and make valuable new contacts during lunches, coffee breaks and the Summit Cocktail Party; we have built more than 12 hours of specific networking into the programme

By the time the Summit draws to a close, you will be armed with a detailed road map of ideas for your business to fine-tune and implement in the months ahead.

Remember: This world class event is the first and only one of its kind in Europe. Come and discover how to transform unstructured textual data into lucrative information and profits. Reserve your place today, and we look forward to seeing you in Amsterdam, 26th – 27th April '07

"Text Analytics is the key detection technology to guarantee European citizen security... safeguarding their privacy and civil liberties!"

> Alessandro Zanasi, Independent Advisor to the European Commission Security Research

Solution Providers

The European Text Analytics Summit 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 TextAnalyticsNews now to take advantage of this unique opportunity.

- · Brand visibility
- Exhibition stands
- · Networking party sponsorship
- Pre-and post conference publicity opportunities
- Advertising
- Delegate pack inserts and seat drops
- Conference stationary sponsorship
- Report and whitepaper sponsorship

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

Contact us now to discuss your sponsorship package.

Call: 1800 814 3459 ext. 326 (freephone US) or +44 (0) 20 7375 7555 (worldwide) to speak to the TextAnalyticsNews team.

Email: belen@textanalyticsnews.com

Exhibition Opportunities

Which strategic alliances will allow me to leverage my capability? Who can make my in-house R&D efforts go the extra mile? What partnerships will allow me to take my solution forward in Europe?

The unstructured data space is gearing up to corner emerging markets. The flurry of acquisitions and start-ups attracting venture capital say it all.

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

Your exhibition area gives you an unrivalled opportunity to:

- Provide exposure for your market-leading solutions
- · Network with potential partners
- Form relationships with tech–savvy end users

Exhibition space is selling out fast, and will be allocated on a first come first served basis.

Contact us now to secure your exhibition space.

Call: 1800 814 3459 ext. 326 (freephone US) or +44 (0) 20 7375 7555 (worldwide) to speak to the TextAnalyticsNews team.

Email: belen@textanalyticsnews.com

Who will be there?

This event has sold out 2 years in a row! The European Text Analytics Summit sister events, the Text Analytics Summit, attracted over 400 professionals from the unstructured data space.

European Text Analytics Summit is the only annual event that focuses purely on the business of creating value from text. It is the perfect forum for meeting, networking with and learning from the major players in the unstructured data value chain.

The key decision makers and senior executives from the leading solution providers and tech-savvy end users will be in attendance, making it imperative for your business to take advantage of this opportunity to build crucial strategic alliances and client partnerships.

Tech-savvy end users from all industries from legal experts to patent counsels, from corporate communication professionals to HR analysts, will be giving top tips for leveraging the technology and setting their challenges for innovation for the coming months.

- 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

CREATE GROUNDBREAKING T

DAY ONE, 26 APRIL 2007

Chairman's welcome & opening speech: Text analytics: The European scene

Find out what the pressing needs and special challenges of the text analytics market in Europe are and see the hottest emerging trends so that you can stay ahead of your competition!

Gain an expert appraisal on the unique potential of text analytics within enterprise knowledge discovery. Hear an overview of real current applications to ensure you are getting the most from your text analytics deployment

Chairman: **Seth Grimes,** President, **Alta Plana**

A European Issue: Securing a prosperous future thanks to text analytics technology

- European security: understand the threats we face, the objectives that need to be met and the cutting edge solutions that are providing the answers
- Obtain the very latest updates on the European Commissions text analytics funding programme – discover what it will mean for both developers – and – end users!
- Analyse the key driving technologies behind European security applications and how you can harness the power of them today

Alessandro Zanasi, Advisor, European Commission Security Research

Visual analytics, text mining and automated knowledge discovery in Pharma



- Hear how Merck KGaA uses advanced visual analytics to make faster decisions which deliver the required information in a record time
- Evaluate how quantitative assessment is obtained using automated target discovery to save time and resources...
 Make sure you do the same!
- Find out how automatic mapping and advanced visual analytics are applied for analysis of intellectual assets putting Merck KGaA at the forefront of their competitors.

Luca Toldo, *Bioinformatics and Scientific Information Services***, Merck KGaA**

The challenge of text: setting up a new customer experience management



- Hear how Telecom Italia improved their CRM strategy by mining customer service notes, blogs and articles to significantly improve their customer satisfaction
- Discover how the information derived from unstructured data has provided invaluable intelligence giving Telecom Italia a competitive advantage
- Telecom Italia will guide you through their journey of how they have successfully deployed text technologies... and profited!

Steffano M. de' Rosi, CRM Manager, Telecom Italia

Unlock your company's hidden knowledge with next generation text analytics

- Discover how innovative text analytics solutions can serve industry-specific knowledge discovery needs for your text focused clients
- Understand how a single entry-point solution brings a comprehensive set of advanced knowledge discovery capabilities to analysts and information professionals
- Learn how text analytics supports demanding activities such as Competitive Intelligence, Customer Sentiment Analysis, and Reputation Management by tracking and analysing websites and other sources of unstructured data.

Guillaume Mazières, Vice President Sales & Marketing, **Temis**

Empower your company's pro-activity with state-of-the-art Text Analytics

- Evaluate how a Dutch Industry leader is benchmarking their reputation from unstructured open-sources online
- Understand why text analytics is becoming an information one-stop shopping solution in a company
- Discover how Information Professionals lift their productivity by editing more reports with a collaborative focus on analysing filtered documents
- Implement cross-lingual text analytics with sentiment measurement in 9 languages including Chinese and Mandarin

Bertrand Dussauge, Vice-President Marketing Analytics, LexisNexis Client to be confirmed

Classifying customer needs-based segmentation using call centre notes



- Hear how HP is able to segment their customer base using the latest text analytics technologies
- See how HP performs segmentation using free form text notes obtained from call centre notes/comments and how to apply this to your business
- Apply effective customer scoring from the needs-based segmentation to ensure you are placing the right customer in the correct needs-based segmentation for maximum returns in direct marketing activities

Randy Collica, Sr. Business/Data Mining Analyst, HP, on behalf of SAS

Getting value out of text mining in Pharma Biotech



- See how AstraZeneca uses text analytics to deal with the vast quantities of information generated by life sciences and understand the key benefits for your business.
- Understand how relevant knowledge is found to inform key decisions by using text mining, and take away tips on how to minimise your costs and maximise your profits.
- Find out how existing knowledge contained in ontologies can be used to gain the best results from your unstructured data!

David Milward, CTO, Linguamatics Ian Dix, Programme Manager, AstraZeneca

"We expect the European Text
Analytics Summit to be a key
meeting place for those interested in
Text Mining, Knowledge Discovery
and Information Management.
With so many related industries
combined, and a program packed
with case studies, networking and
exposition, the Summit is sure to
deliver incredible value on all fronts."

Dave Haucke, VP Global Marketing, ISYS Search Software

Super Discounts! Call now +44 (0)20 7375 7575 and save €200

EXT ANALYTICS SOLUTIONS AND DELIVER MAXIM

DAY TWO, 27 APRIL 2007

PANEL DEBATE: Future technologies and new applications

What does the future hold for Text Analytics? What are the newest technologies? Hear about the next big thing set to hit the European Market!

- Technology landscape: How is the current text mining industry shaped in Europe? What has happened in the last 12 months?
- Technology exchange: Which technologies are most widely used in Europe? Which technologies will provide you the highest return?
- Technology forecast: Discover the technologies under development which will shape the landscape of future markets + tips on how to design your tools of tomorrow... TODAY!

Dr Andrew Salway

Jaap van der Meer, Director, Translation Automation User Society Moderator: Seth Grimes, President, Alta Plana

If you are interested in participating in this panel, call +44 207 3757 555 - NOW!

NETWORKING COCKTAIL PARTY

Join your peers for an evening of networking at the Text Analytics Summit 2007 Cocktail Party



FREE!

Free reports and whitepapers: www.textanalyticsnews.com

Download yours

NOW!

Going beyond search: You found it, now what can you do with it?



- Discover how the University of Nebraska uses text mining to analyse results and group data to aid them in their critical research
- See how the University of Nebraska leveraged search to go beyond retrieval to include analytical and mining functions and get an insight into the results they have seen
- Find out how an entire 3 months of work was reduced to just 2 manageable minutes using text analytics technologies

Dr Burch T. Kealey, *Professor and Researcher*, University of Nebraska, speaking on behalf of ISYS Search

Embedding text mining solutions in pharmaceutical research



- Discover how Pfizer deploys text mining solutions which are accessible for research scientists and easy to integrate into their everyday work
- Hear Pfizer strategies from simple searches to advance co-occurrence matrices, to know how to stay ahead of the game
- Understand how Pfizer continuously re-compute searches across full literature databases and examine how they scaled the volume through a grid computing approach to get an insight into the deployment of your products!

lan Harrow, Senior Principal Scientist, Pfizer

Rob Hernandez, Senior Scientist, Pfizer

Lee Harland, Senior Principal Scientist, Pfizer

Pfizer Full Team

Text mining for Biomedicine

- Get a valuable insight into the role of the UK National Centre for Text Mining and the latest technologies they are developing
- Identify the tools and services for biomedicine that will have the biggest impact, including automatic term management, semantic retrieval, relation mining and interaction network retrieval

Sophia Ananiadou, Co-director, NaCTeM

How Cisco Systems uses computers to manage its reputation



- Examine how Cisco Systems
 measures and analyses written
 information to optimise reaction during
 acquisitions, earnings, launches, and
 other events to stay ahead of their
 competition
- Analyse how opinions and experiences are balanced with data when positioning an issue, or defining corporate communications messaging
- Learn how Cisco manages approximately 300,000 articles coming in per day from various sources that cascade through multi-stage filtering and analysis

Seth Redmore, *Manager of Marketing Operations,* **Cisco Systems**

Search results that make sense: Are you doing search wrong?

- Search engine insight: learn what constitutes a good search and how to develop your products accordingly
- Discover how a friendly search engine can help your company find the data they are looking for quickly and reliably, and how this will drive your business ahead of the competition
- Find out how a search engine that understands your client's business will ensure your product is fast, scalable, reliable and useable

Raul Valdes, CEO, Vivísimo, Client to be confirmed

Automated Patent Classification for Philips IP&S using text mining and 3D-visualisation

- Analyse how high precision and recall in IP classification is used at Philips and the results obtained
- See Philips' latest achievements in automated text mining and visualization, and how you can apply the techniques to your business
- Discover Philips' experiences with automated patent classification solution, and gain a valuable insight into their use of text analytics

Dr. Anton Heijs, Chief Technology & Innovation Officer,

Treparel Information Solutions

Mr. Piet van Zanten, Senior Patent Information Specialist, Philips

UM VALUE TO YOUR END USER

Understand, predict and act using text analytics **◄**



- Uncover hidden patterns, relationships and trends using text analytics, to gain greater insight into articles, reports, surveys, call centre notes, emails and chats
- Understand how commercial organisations are using predictive text analytics to support their CRM processes and improve profits by analyzing customer preferences
- Hear how public sector bodies deploy text analytics to protect public safety and security by predicting what types of fraud, waste and abuse are likely to occur, when and where
- Learn how the Dutch National Police uses the 'digital washing machine' they created, together with feature extraction, to get faster insight in huge amounts of unstructured data

Olivier Jouve, *Vice-president of Marketing Strategy,* **SPSS**

Jochen van der Wal, Technical Engineer, The Dutch National Police

PANEL DEBATE: Let the end users have their say!

Introduction by Henk Alles: Make room for what matters

- How much do you know about European end users? Do you know what solutions they need?
- What do know what European end users expect from you? How can you establish yourself as the leader in the market?
- Hear the importance of text mining to leading end users and understand their results and be prepared to produce the perfect text technology product!

Followed by an in-depth discussion with the end users:

Deloitte

Chris Bowman, Lafourche Parish Public Schools

Introduction and moderator:
Henk Alles, President, Infolutions

SPEAKER UPDATES

Find out who's just joined the agenda at: www.textanalyticsnews.com/eu07/ speakers.shtml

ROUND TABLES

TextAnalyticsNews are proud to bring you interactive roundtables to ensure you leave Amsterdam with an in-depth knowledge of how you and your competitors address the critical issues facing the text analytics space

WHAT ARE ROUNDTABLES?

Roundtables are small, topic focused interactive discussionseach comprising between 10 and 12 participants- that are designed to enhance your learning and give you practical know-how to apply to your business

WHAT WILL I LEARN?

- Through active discussion rather than passive listening, you will be able to benchmark you own strategies against those of your competitors
- You will be able to question colleagues from within the industry and get expert answers to your specific questions
- Roundtables are an extremely effective method of identifying who you would like to spend more time speaking with, both during and after the summit

ROUNDTABLE TOPICS INCLUDE:

- Effective marketing in Europe
- Switching potential end users to adopters
- Multilingual technologies
- Text Analytics in Pharma
- Voice and video Mining
- Potential verticals for Text Analytics in Europe
- Security and Text Analysis in Europe
- Text Analytics for CRM

5 NETWORKING OPPORTUNITIES!

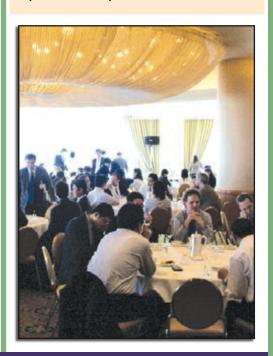
Themed Mindshare
Roundtables are a chance to
discuss important business issues
with people who share the same
problems and challenges as you.
An ideal forum to share experiences
and find solutions

Cocktail Party Gives you time to relax, unwind and network with the people that matter the most to you!

12+ Hours of Networking
Meet dozens of new people and
potential business partners

Exhibition Test drive the latest text technologies, learn about killer applications, ask questions, offer feedback and gain first-hand knowledge of cutting edge technology

Panel Sessions offer you a unique opportunity to interact with the leading businesses and solution providers of the unstructured data space in Europe.





The European Text Analytics Summit, following two blockbusting summits in the US!

SAVE €200

If you register before the 9th of March!

END USERS

Register today and save 40%!!

V! USER

26th and 27th April 2007
The Marriott Hotel in Amsterdam

Date & Venue

Hotel Discounts SAVE €€€

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

Group Discounts SAVE €€€

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

Contact the TextAnalyticsNews team on +44 (0)20 7375 7575 or email europe@textanalyticsnews.com for more details

THE EUROPEAN TEXT ANALYTICS SUMMIT

3 EASY STEPS TO REGISTER NOW!

1. Your Choice of Registration Package

Please tick the package price box you require below:

	DISCOUNT PRICE Register by March 9th	STANDARD PRICE
GOLD PASS: 2 Day Pass Full access to conference and exhibition Full Summit Pass (solution providers) Access to presentations post show	€1295 +VAT	€1495 +VAT
END USERS PASS: 2 Day Pass Full access to conference and exhibition End Users Pass (end users only) Access to presentations post show Users of Text Analytics Technologies	€895 +VAT	€995 +VAT

2. De	legat	te Det	tails
-------	-------	--------	-------

Please photocopy this form fo	r muliple registrations	Customer Priority Code:
Mr / Mrs / Ms / Dr:	First name:	
Last name:		
Company:	Position/Title:	
Telephone:	Fax:	
Email:		
Address:		
Postcode:	Country:	

5 EASY WAYS TO REGISTER TODAY!

- Call TextAnalyticsNews on +44 (0) 20 7375 7575
- Fax This form to +44 (0) 20 7375 7576
- Email The TextAnalyticsNews registration team on europe@textanalyticsnews.com
- Mail This form to TextAnalyticsNews 7-9 Fashion Street, London, E16PX, UK
- Online Register your details securely at http://www.textanalyticsnews.com/eu07/register.shtml

3.	P	av	m	en	t
	_		ши		

•
Choose one of the following payment options:
☐ I enclose a check/draft for: €
☐ 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

Cancellation Policy

Cancellation Policy: Places are transferable without any charge. Cancellations before the 16th of March 2007 incur and administrative charge of 25%. If you cancel your registration after the 16th of March 2007 we will be obliged to charge the full fee. Please note- you must justify TextAnalyticsNews in writing of a cancellation, or we will be obliged to charge the full fee. The organisers 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 of the invoice date and relevant details will appear on the invoice. Please see terms & conditions on http://www.textanalyticsnews.com/eu07/terms.shtml

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



Real-world case studies: Hear from Telecom Italia, Merck KGaA, Philips, Pfizer, HP, the Dutch Police, the European Commission, University of Nebraska-Omaha, AstraZeneca and more!



THE European TEXT ANALYTICS SUMMIT

26th and 27th April 2007 - The Marriott Hotel, Amsterdam

Deliver the most effective text-based solutions, understand the needs of your customers and secure ROI in the exploding European market

Your roadmap to text analytics success: Increase your ROI by delivering key text solutions your European customers want!

7 REASONS TO ATTEND

Real-life case studies

Hear from tech-savvy and end users who really know what they are buying. They will give us their overview of the product landscape, discuss their evolving needs and set their challenges for innovation.

New markets, new players, new revenue!

New applications for text analysis are being created every day. Discover which horizontal and vertical markets are ready and waiting for your next solution.

Leverage your technology: Whether you are a solution provider wanting to corner emerging markets or an end user wanting to make your data work harder, you'll get the information you need for building new applications with bottom line impact.

Get a roadmap for commercial success:

The greatest minds in the industry will tell you know the massive potential of text analytics will be transformed into revenue.

Text Analysis ROI:

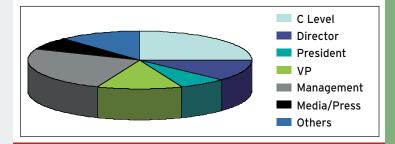
End users are investing more and more in text analysis. Need a place to begin? Our session will give you a comprehensive model for ROI.

Cross-industry mindshare: Find out how text mining users in a broad range of vertical industries are making their data work harder and delivering business-critical actionable data.

Business model: Every day, start-ups providing text-driven solutions for all industries are making a big splash in the market place, but which business models will deliver long term success?

Take a look at our past attendees!

The European Text Analytics Summit sister events, the Text Analytics Summit, attracted over 400 professionals from the unstructured data space. Now we are opening up new markets ensuring we drive the industry forward!



5 Easy Ways to Register

www.textanalyticsnews.com

(worldwide)

- @ europe@textanalyticsnews.com1800 814 3459 or +44 (0) 20 7375 7575
- 1800 814 3460 or +44 (0) 20 7375 7576 (worldwide)
- TextAnalyticsNews, 7-9 Fashion St, London, E1 6PX, UK

Look inside for the complete 2007 European Summit agenda