

Where Text Analytics is at Today

Robert Dale
Centre for Language Technology
www.clt.mq.edu.au

The Aim of This Talk

- To give you an idea of where the commercial activity is in text analytics
- To give you an idea of where current research is focussed in text analytics

The Problem: Too Much Information

- 80% of data is 'locked up in text'
- How big is the web?
 - A recent frequently cited answer: 155,583,825 sites
 - Google indexes in excess of 8 billion pages
- FAST's enterprise search solution handles up to 10 billion documents
- My hard disk contains 29.7Gb of work-related data in 106k documents
- My Outlook file is 800Mb with no attachments included

A Solution: Text Analytics

- Text Analytics is
 - ... the process of automatically extracting ...
 - ... structured or semi-structured information
 - ... from unstructured machine-readable documents

Text Analytics Components

- Search
- Text Categorisation and Clustering
- Entity Recognition
- Cross-Document Name Matching
- Text Summarisation
- Event Recognition
- Sentiment Analysis
- Question Answering

Search

- ... aka Information Retrieval
- Basic techniques:
 - treat documents as bags of words, look for overlap with query
- Smarter ideas:
 - Page rank; link analysis; hubs and authorities
 - Using linguistic knowledge
- State of the art: Still really searching for words rather than searching for concepts

Linguistic Knowledge in Search

- High frequency ‘closed class’ words are not very informative:
 - ‘the’, ‘and’, ‘of’
- Automatic detection of multiword concepts:
 - New York City subway
- Stemming or morphological analysis can find pages otherwise missed:
 - ‘apples’ → ‘apple’, ‘application’ → ‘apply’
- Query expansion via synonyms:
 - ‘car’ → ‘automobile’

[Home](#)[Products](#)[Consulting](#)[Company](#)[Support](#)[Media Centre](#)[Contact us](#)

Funnelback Internet & Enterprise Search

Funnelback is an Internet and Enterprise search engine company offering a suite of search solutions, including a 'Software as a Service' or hosted solution for the web and a fully customisable whole-of-enterprise solution for searching behind the firewall.

Used by organizations every day to find information contained in public websites, intranets, shared drives, databases and library catalogues, Funnelback's key point of difference is its superior ranking quality and the ability to tune its core ranking algorithm.

Funnelback's [client list](#) boasts blue chip companies such as Westpac and the ASX, one third of all Australian Universities, state and federal governments including the Australian Government, as well as clients in the United Kingdom and Canada.

See Funnelback in action

[Australian Government](#)

News

Funnelback has won a two year contract to provide search for the Australian

Quicklinks

[Case Studies](#)



A Microsoft® Subsidiary

Solutions

Industries

Service & Support

Customers

Partners

Knowledge Center

Solutions for your business needs

Solutions for your industry

Search Advertising Personalization Merchandising

The business of search.

You need to make better, faster decisions. Reinvent your business model. Locate an expert. Convert browsers to buyers. Glean insight from data. Sell targeted advertisements. Track important topics. Build a loyal user community. Drive revenue and reduce costs. While the business challenges vary, the solution is the same: search.

From inside the enterprise and across the Web, to the desktop and the mobile device, FAST provides the most advanced search platform and add-on solutions to help you overcome all your information access challenges.

[Find out more](#)

NEWS Microsoft Completes Tender Offer for Fast Search & Transfer

FAST

Now

Microsoft

Subsidiary

[Read the press](#)

[Read the Customer](#)

Google™ Welcome to Google Enterprise

Google Mini Search Appliance



New! Google Mini gets 3 major features and 6 more languages! [Learn more.](#)

The Google Mini offers the simplicity and power of Google's universal search technology at a great price. An integrated hardware and software solution, the Mini offers true plug-and-play installation and can be purchased online. Whether you're looking to search your company's internal information or your public website, the Mini has the features for your business.

Search your corporate data and improve employee productivity

With the Google Mini, you can offer employees relevant and secure search across intranets, file servers, and business applications. Employees' familiarity with Google's user interface from searching on Google.com will facilitate adoption in your enterprise, and instant access to corporate information will make them more productive.

Add search to your public website for increased conversion rates and page views

If you operate a public website, you probably already know that your visitors rely on search to find products and information on your site. When visitors cannot find what they search for, they leave your site. With relevant search results, they'll be more likely to purchase from and return to your website.

The Google Mini's relevant search results and customizable user interface make it a perfect solution for website search. Integration with Google Sitemaps also makes it easy to submit your website for inclusion in Google.com search results.

The Google Mini works with over [220 different file formats](#) and its integration with existing security systems ensures that employees can only access information they are supposed to view.

Additional Resources

[Features](#)

[Tours & demos](#)

[Success stories](#)

[Enterprise Superstars](#)

[Mini Map](#)

[FAQ](#)

[Contact sales](#)

[Home](#)

[Search](#)

- [Website](#)

- [Universal](#)

- [Products](#)

- [Search Appliance](#)

- [Google Mini](#)

- [Google Site Search](#)

- [Desktop Search](#)

- [Enterprise Labs](#)

[Google Apps](#)

[Geo](#)

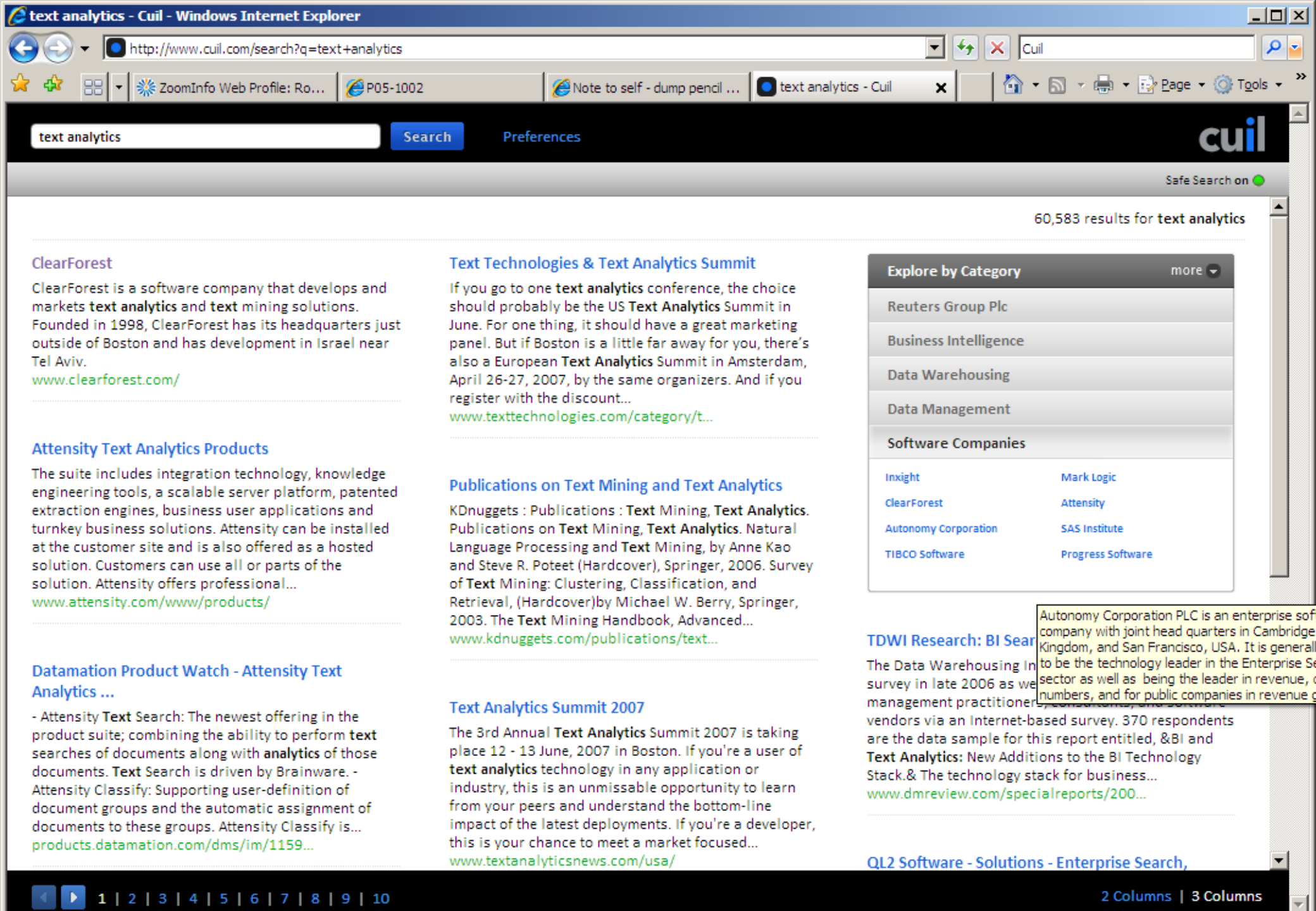
[Tours & demos](#)

[Partners & Developers](#)

[Solutions Marketplace](#)

[Customers](#)

[Enterprise Solutions](#)



text analytics

Search

Preferences



Safe Search on

60,583 results for text analytics

ClearForest

ClearForest is a software company that develops and markets text analytics and text mining solutions. Founded in 1998, ClearForest has its headquarters just outside of Boston and has development in Israel near Tel Aviv.
www.clearforest.com/

Attensity Text Analytics Products

The suite includes integration technology, knowledge engineering tools, a scalable server platform, patented extraction engines, business user applications and turnkey business solutions. Attensity can be installed at the customer site and is also offered as a hosted solution. Customers can use all or parts of the solution. Attensity offers professional...
www.attensity.com/www/products/

Datamation Product Watch - Attensity Text Analytics ...

- Attensity Text Search: The newest offering in the product suite; combining the ability to perform text searches of documents along with analytics of those documents. Text Search is driven by Brainware. - Attensity Classify: Supporting user-definition of document groups and the automatic assignment of documents to these groups. Attensity Classify is...
products.datamation.com/dms/im/1159...

Text Technologies & Text Analytics Summit

If you go to one text analytics conference, the choice should probably be the US Text Analytics Summit in June. For one thing, it should have a great marketing panel. But if Boston is a little far away for you, there's also a European Text Analytics Summit in Amsterdam, April 26-27, 2007, by the same organizers. And if you register with the discount...
www.texttechnologies.com/category/t...

Publications on Text Mining and Text Analytics

KDnuggets : Publications : Text Mining, Text Analytics. Publications on Text Mining, Text Analytics. Natural Language Processing and Text Mining, by Anne Kao and Steve R. Poteet (Hardcover), Springer, 2006. Survey of Text Mining: Clustering, Classification, and Retrieval, (Hardcover) by Michael W. Berry, Springer, 2003. The Text Mining Handbook, Advanced...
www.kdnuggets.com/publications/text...

Text Analytics Summit 2007

The 3rd Annual Text Analytics Summit 2007 is taking place 12 - 13 June, 2007 in Boston. If you're a user of text analytics technology in any application or industry, this is an unmissable opportunity to learn from your peers and understand the bottom-line impact of the latest deployments. If you're a developer, this is your chance to meet a market focused...
www.textanalyticsnews.com/usa/

Explore by Category

more

- Reuters Group Plc
- Business Intelligence
- Data Warehousing
- Data Management
- Software Companies
 - Inxight
 - ClearForest
 - Autonomy Corporation
 - TIBCO Software
 - Mark Logic
 - Attensity
 - SAS Institute
 - Progress Software

TDWI Research: BI Search

The Data Warehousing In survey in late 2006 as we management practitioners vendors via an Internet-based survey. 370 respondents are the data sample for this report entitled, &BI and Text Analytics: New Additions to the BI Technology Stack.& The technology stack for business...
www.dmreview.com/specialreports/200...

Autonomy Corporation PLC is an enterprise software company with joint head quarters in Cambridge Kingdom, and San Francisco, USA. It is generally to be the technology leader in the Enterprise S sector as well as being the leader in revenue, c numbers, and for public companies in revenue c

QL2 Software - Solutions - Enterprise Search,

Text Analytics Components

- Search
- Text Categorisation and Clustering
- Entity Recognition
- Cross-Document Name Matching
- Text Summarisation
- Event Recognition
- Sentiment Analysis
- Question Answering

Text Categorisation

- Given a set of categories, automatically determine which category a new item belongs to
- Machine learning identifies the characteristics of existing documents that cause them to belong to their categories

Uses of Text Categorisation

- Organising information into a structure
 - Email folders
 - Spam and junk mail detection
 - ASX company announcements – around 150 distinct document types
- Text categorisation can be rule-based or machine-learned

Text Clustering

- From a given set of uncategorized items, determine the most natural data-driven groupings
- Machine learning identifies the similarities and differences between existing documents



[advanced preferences](#)

[clusters](#) [sources](#) [sites](#)

remix

All Results (181)

- + Text Mining (27)
- + Customer, Voice (27)
- + Data warehousing (16)
- + Text Analytics Summit (13)
- + Attensity (15)
- + IBM (10)
- + ClearForest, Enabled (7)
 - Techniques, Extraction (5)
- + Anderson Analytics (5)
- + Analytic Solutions (5)

[more](#) | [all clusters](#)

find in clusters:

Top 180 results of at least 1,590,000 retrieved for the query **text analytics** ([details](#))

[Free SMS to your Friends in Australia](#)

Sponsored Results

Send 5 free SMS each day plus free Ringtones and Games for your Mobile. - [mobi.dada.net](#)

Search Results

1. [Text analytics - Wikipedia, the free encyclopedia](#)

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.

[en.wikipedia.org/wiki/Text_analytics](#) - [cache] - Live, Gigablast, Ask

2. [Noisy text analytics - Wikipedia, the free encyclopedia](#)

Noisy **Text Analytics** is a process of information extraction whose goal is to automatically extract structured or semistructured information from noisy unstructured **text** data. While **Text analytics** is a growing and mature field that has great value because of the huge amounts of data being produced, processing of noisy **text** is gaining in importance because a lot of common applications produce noisy **text** data.

[en.wikipedia.org/wiki/Noisy_text_analytics](#) - [cache] - Live, Gigablast, Ask

3. [The European Text Analytics Summit 2007](#)

The European **Text Analytics** Summit 2007 is the first commercial event in Europe to bring together tech savvy end users and solution providers to discuss ...

[www.textanalyticsnews.com/europe07](#) - [cache] - Gigablast, Ask



[about](#) | [products](#) | [solutions](#) | [press](#) | [partners](#) | [support](#)

[Advanced Search](#)

[Help](#)

NEW search the [Wikipedia](#) at [Clusty.com](#)

Clustered Results

Top 158 results of at least 1,114,000 retrieved for the query **text analytics** ([Details](#))

- ▶ [text analytics](#) (162)
- ⊕ ▶ [Attensity, Suite](#) (28)
- ⊕ ▶ [Text analytics solutions](#) (24)
- ⊕ ▶ [Mining](#) (20)
- ⊕ ▶ [Software](#) (21)
- ⊕ ▶ [Enterprise, Search And Text Analytics](#) (19)
- ⊕ ▶ [Text Analytics Summit](#) (18)
- ⊕ ▶ [ClearForest](#) (12)
- ⊕ ▶ [Business Intelligence](#) (12)
- ⊕ ▶ [Scientific](#) (8)
- ⊕ ▶ [Predictive Text Analytics](#) (6)
- ▼ [More](#)

[Text Analytics](#)

Overview of Companies and Products Contacts, News, Descriptions
[text-analytics.com](#) - Sponsored Listings 1

Sponsored Link

[Analytical Test](#)

Identify & Quantify Your Staffs Strengths - Profiles International
[www.profiles.net.au](#) - Sponsored Listings 2

Sponsored Link

1. [Attensity Text Analytics](#) [\[new window\]](#) [\[frame\]](#) [\[preview\]](#) [\[clusters\]](#)
 Attensity provides **text analytics** solutions that let companies detect and **analyze** critical events ... March 12, 2007 Attensity Law Enforcement **Analyst** Desktop Solution (LEADS) March 01, 2007
[www.attensity.com](#) - Ask 1, MSN 11
2. [Attensity Text Analytics Products](#) [\[new window\]](#) [\[frame\]](#) [\[cache\]](#) [\[preview\]](#) [\[clusters\]](#)
 Attensity Server, Attensity Discover, Attensity **Analytics** and Attensity Workstation products provide ... Attensity **Text Analytics** Products. Attensity's technology is the culmination of over a decade ...
[www.attensity.com/www/products](#) - Ask 2, MSN 10
3. [Content Analyst: Text Analytics: Overview](#) [\[new window\]](#) [\[frame\]](#) [\[cache\]](#) [\[preview\]](#) [\[clusters\]](#)
 Content **Analyst**™ Technology Applies **Text Analytics** To Make Sense of Unstructured Data. The sheer volume of data now available to companies and government organizations challenges their ability of to ...
[www.contentanalyst.com/text_analytics/overview.htm](#) - Wisenut 4, MSN 13
4. [Text Analytics Summit 2007](#) [\[new window\]](#) [\[frame\]](#) [\[cache\]](#) [\[preview\]](#) [\[clusters\]](#)
 First ever commercially focused **text** mining conference to teach you how to leverage unstructured ... The

Find in clusters:



See how Grokker can help your business

Selected Sources [3 of 3] [Add/Remove](#)

Yahoo!, Wikipedia, Amazon Books

text analytics

GROK

[Search Options](#)
 Outline View

 Map View

215 total results

[Expand Outline](#) | [Collapse Outline](#)

 Detail: [Less](#) [Medium](#) [More](#)
 Working List

0 items in your list

[View your list](#)
 Email Outline...

 Export Outline...

 Search within the outline:
 by keyword

 Exclude

by date

all most recent

by source

by domain

text analytics (215 results)

General (87)

 Text Analysis (30)

 Classical Mechanics (21)

 Analytic Methods (19)

 Analytic Geometry (16)

Topics Include (16)

 Analytic Functions (12)

 Text Mining (11)

 Applied Mathematics (10)

 Book Contains (10)

Business Intelligence (10)

Number Theory (9)

Text Analytics Summit (9)

Textual Analytics (9)

 Analytical Problems (8)

Differential Equations (8)

 Provides an Introduction (8)

text analytics

[Mathematical Methods of Classical Mechanics \(Graduate Texts in Mathematics\)](#)

[Add to Working List](#) | [Post to del.icio.us](#) | [Bookmark](#) | [Email](#)

In this text, the author constructs the mathematical apparatus of classical mechanics from the beginning, examining all the basic problems in dynamics, including the theory of oscillations, the theory of rigid body motion, and the Hamiltonian formalism. This modern approach, based on the theory of...

<http://www.amazon.com/Mathematical-Classical-Mechanics-Graduate-Mat...> - Friday, 5 September 1997

Source: Amazon Books

[Category:All articles to be expanded](#)

[Add to Working List](#) | [Post to del.icio.us](#) | [Bookmark](#) | [Email](#)

All articles to be expanded

http://en.wikipedia.org/wiki/Category:All_articles_to_be_expanded - Monday, 28 July 2008

Source: Wikipedia

[Text analytics - Wikipedia, the free encyclopedia](#)

[Add to Working List](#) | [Post to del.icio.us](#) | [Bookmark](#) | [Email](#)

Text analytics. From Wikipedia, the free encyclopedia. The introduction to this article provides insufficient context for those unfamiliar with the subject. Please ...

http://en.wikipedia.org/wiki/Text_analytics - 31k - Wednesday, 16 July 2008

See how Grokker can help your business

[News & Events](#) | [Blogs](#) | [Contact Us](#) | [Feedback](#) | [Help](#) | [Home](#)


Selected Sources [2 of 3] [Add/Remove](#)
 Yahoo!, Wikipedia

 [Search Options](#)
 Working List

0 items in your list

[View your list](#)
 [Email Map...](#)
 [Export Map...](#)

Search within the map:

by keyword

 Exclude

by date

all most recent

by source

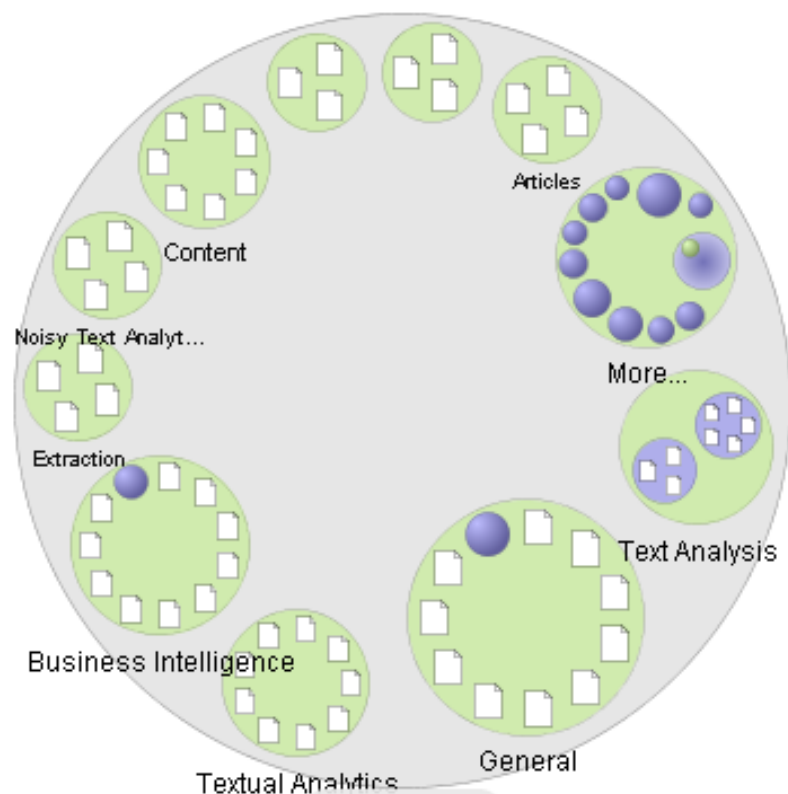
by domain

 Outline View

 Map View

173 total results

[Expand View](#)

 Detail: [Less](#) [Medium](#) [More](#)
[Expand Detail](#)

[Category: All articles to be expanded](#)
[Add to Working List](#) | [Post to del.icio.us](#) | [Bookmark](#) | [Email](#)

All articles to be expanded
[http://en.wikipedia.org/wiki/Category:All_articles_to_Monday, 28 July 2008](http://en.wikipedia.org/wiki/Category:All_articles_to_Monday,_28_July_2008)
 Source: Wikipedia

[Text analytics - Wikipedia, the free encyclopedia](#)
[Add to Working List](#) | [Post to del.icio.us](#) | [Bookmark](#) | [Email](#)

Text analytics. From Wikipedia, the free encyclopedia. The introduction to this article provides insufficient context for those unfamiliar with the subject. Please ...
[http://en.wikipedia.org/wiki/Text_analytics - 31k - Wednesday, 16 July 2008](http://en.wikipedia.org/wiki/Text_analytics_-_31k_-_Wednesday,_16_July_2008)
 Source: Yahoo!

[Noisy text analytics](#)
[Add to Working List](#) | [Post to del.icio.us](#) | [Bookmark](#) | [Email](#)

Noisy text analytics
[http://en.wikipedia.org/wiki/Noisy_text_analytics - Monday, 28 July 2008](http://en.wikipedia.org/wiki/Noisy_text_analytics_-_Monday,_28_July_2008)

Text Categorisation and Clustering

- Categorisation:
 - Taxonomies are alive and well
 - Vendor claim: deploying taxonomies can reduce the amount of time it takes to find information by 50%
- Clustering:
 - Available in some second-tier search engines
 - Why hasn't Google prime-timed it?

Text Analytics Components

- Search
- Text Categorisation and Clustering
- Entity Recognition
- Cross-Document Name Matching
- Text Summarisation
- Event Recognition
- Sentiment Analysis
- Question Answering

Named Entities

- The idea: entities are more important than words
- Traditionally:
 - People, organizations, geographical locations
 - Time expressions, currency amounts, weights and measures
- Now:
 - All of the above, plus key concepts or terms in any field

Approaches

- Gazetteers:
 - large lists of names and variants
- Rule-based systems:
 - Eg, "any initcapped string with 'Ltd' at the end is a company"
- Machine learning:
 - Based on the words that precede and follow initcapped strings, determine if they are likely to be named entities of a specific type

attensity

PRODUCTS

OVERVIEW

APPLICATIONS

ARCHITECTURE

ATTENSITY SERVER

ATTENSITY MARKET VOICE

INTEGRATION/SDK

KNOWLEDGE ENGINEERING

VOC ON-DEMAND

QUICK LINKS

- > [Attensity Discover](#)
- > [Attensity Analytics](#)
- > [Attensity Text Search](#)
- > [Attensity Classify](#)
- > [Attensity Alerts](#)
- > [Attensity Server](#)
- > [Attensity Integration](#)

ATTENSITY TEXT ANALYTICS PRODUCTS

Attensity's technology is the culmination of over a decade of research in computational linguistics at the University of Utah. This research led to breakthrough software that allows computers to understand and process free-form text, offering government and commercial organizations the opportunity to leverage the vast amounts of information contained in nonstructured formats.

The technology allows users to extract and analyze facts like who, what, where, when and why and then allows users to drill down to understand people, places and events and how they are related. It then creates output in XML and in a structured relational data format that is fused with existing structured data for analysis. Building on its award winning, patented text extraction technology, Attensity offers an integrated Text Analytics suite that enables users to seamlessly extract facts from text using Attensity's wide-range of statistical and linguistic extraction technologies, and then using a web-based application, further search, classify, discover and analyze text to rapidly and accurately identify issues, uncover problems and drive decisions.

Attensity offers a complete suite of products for Text Analytics. The suite includes integration technology, knowledge engineering tools, a scalable server platform, patented extraction engines, business user applications and turnkey business solutions. Attensity can be installed at the customer site and is also offered as a hosted solution. Customers can use all or parts of the solution. Attensity offers [professional services](#) and training to enable its customers to leverage the power of Attensity's Text Analytics suite.

Below is a diagram that displays the Attensity Text Analytics suite. Attensity offers a wide-range of technologies and solutions for Text Analytics and is the only vendor in the market that offers [Exhaustive Extraction™](#), an approach that automatically finds important information hidden in text without extensive rules creation or pre-work.

clarabridge™

A Unique Approach | Solutions | Products | Services | Customers | Partners | About Us

Unlock the potential of your unstructured data.

Reliable analytics enable smart solutions to complex business problems.

Clarabridge™ transforms unstructured content into rich, structured data, unlocking the full power of information assets. As a result, notes, comments, and documents can be mined alongside databases using standard business intelligence (BI) tools and analysis techniques. Now you can ask any question of any information source using any analysis technique to address a host of real-world business needs.

That's true intelligence. That's [Clarabridge™](#).

Look at some of the examples of what Clarabridge™ can unlock for you:

▶ [Customer Relationship Management](#)

Use every bit of your information to improve customer retention.

▶ [Investigation](#)

Find descriptions, comments, and scenarios that indicate anomalous or fraudulent insurance claims.

Upcoming Webinars

▶ [Transform customer feedback into actionable analysis.](#)

[Join us for our most requested webinar](#)
April 26th, 2007 at 1 PM (EDT)

Recent News

▶ [Software Helps TurboTax Address Tax Filers' Complaints](#)

▶ [Clarabridge's Content Mining Platform 2.2](#)



Unlocking the Potential of
Unstructured Data.



Worldwide | Contact | Site Map | Store | Free Trials

SUBMIT

› PRODUCTS › SERVICES › SUPPORT › SOLUTIONS › CUSTOMERS › PARTNERS › COMPANY

[Home](#) / [Products](#) / [Query, Reporting, and Analysis](#) / [Advanced Analysis](#) / [BusinessObjects Text Analysis](#)

BUSINESSOBJECTS TEXT ANALYSIS

Turn Unstructured Data Into Insight

BusinessObjects Text Analysis "reads" text in 30+ languages, extracting key information so unstructured text data can be used as a data source for data integration or business intelligence, uncovering hidden information in CRM systems, Web, emails, and other text sources. Applications include Voice of the Customer, Root Cause Analysis, Link Analysis, Shareholder Value Analysis, Employee Satisfaction Analysis, and more.

[Product Benefits](#)

[Technical Information](#)

[Purchasing Information](#)

Use the Power of Text for Greater Insight

- Convert text into structured data to create a complete view of customers, partners, and market.
- Extend BusinessObjects Enterprise and Data Integrator deployments to access and analyze text using Crystal

"We originally developed Starlight to aid researchers in visually spotting patterns and anomalies in data. By integrating [BusinessObjects] entity extraction, we have added a powerful and precise way to identify exact entities. One example of the success we've had is a recent study of a food poisoning outbreak in a particular U.S. city

Contact sales now:

[LOCATE A SALES PERSON](#)

Learn more:

[VIEW DEMO](#)

Get More Information:

[Global contact list](#)
[Request more information](#)
[Find a reseller](#)



TEMIS

NEWS

SOLUTIONS

CLIENTS

PARTNERS

CONTACT

LUXID SOLUTION FOR PUBLISHERS CORE PRODUCTS SKILL CARTRIDGE™ LIBRARY



10:52:07 15-04-2007

CORE PRODUCTS

XeLDA®



Insight Discoverer™
Extractor



Insight Discoverer™
Clusterer



Insight Discoverer™
Categorizer



eXtraction
Terminology Suite



ONLINE MINER™

APPLICATIVE SOLUTIONS

» SOLUTIONS / SKILL CARTRIDGE™ LIBRARY

Overview

Generic

Analytics

Text Mining 360°

Competitive Intelligence

Human Resources Management

Life Sciences

Biological Entity Relationships

Medical Entity Relationships

Chemical Entity Relationships

OVERVIEW

What is a Skill Cartridge™?

A Skill Cartridge™ is a set of customizable knowledge components that defines the information to be extracted. The two major knowledge components of a Skill Cartridge are multi-lingual dictionaries that assign concepts to words and phrases, and multi-lingual extraction rules that establish relationships between defined concepts. An inference component provides the functionality to further interpret and normalize the extraction output. A Skill Cartridge™ is charged into the Insight Discoverer Extractor™ that performs the extraction.

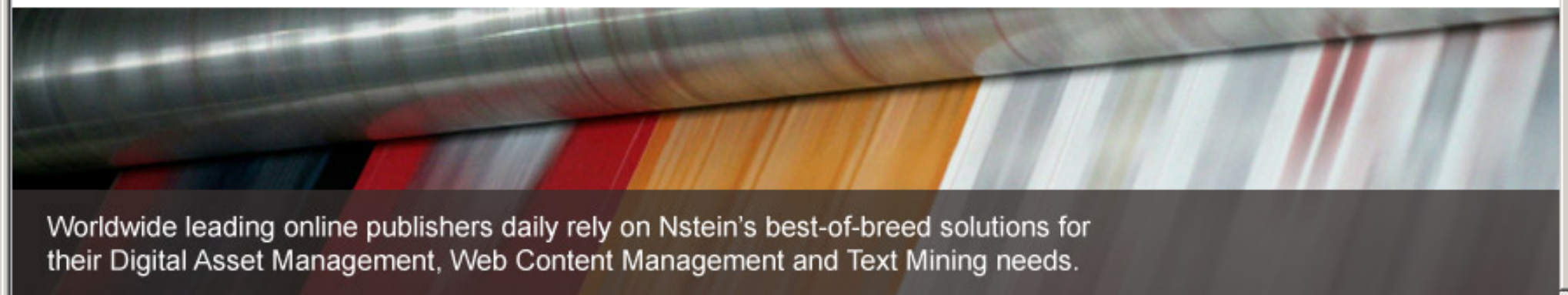
The Skill Cartridge™ Library

TEMIS Skill Cartridges™ is grouped together into a single collection called the Skill Cartridge™ Library. The collection includes:

- The general Skill Cartridges™



Products / Text Mining Engine



Worldwide leading online publishers daily rely on Nstein's best-of-breed solutions for their Digital Asset Management, Web Content Management and Text Mining needs.

PRODUCTS OVERVIEW
WEB CONTENT MANAGEMENT
WCM Features
DIGITAL ASSET MANAGEMENT
DAM Features
TEXT MINING ENGINE
TME Features
PICTURE MANAGEMENT DESK
PMD Features

Engaging Online Experiences that Generate New Revenue Opportunities

Nstein's Text Mining Engine (TME) is a powerful, multilingual solution that enables publishers to offer a more engaging experience to online readers by leveraging the "aboutness" of their content. With deep, wide websites that both attract more unique visitors and enhance site stickiness, publishers are then able to increase their advertising inventories rapidly and

Real ROI Nstein's TME helped:

- A news portal increase its website stickiness and ad inventory 10x by introducing contextual topics that draw readers to read more related stories
- A broadcaster unlock decades of rich historical information within its digital asset management environment by

TME Architecture Diagram

Click to enlarge



TME Data Sheet

Download TME's features, benefits and specs.





ABOUT US

LATEST NEWS

OUR SERVICES

OUR SOLUTIONS

OUR PARTNERS

SOLUTIONS

PARTNERS

SERVICES

COMPANY

Discover content.
Understand its impact.
Act on the message.



STOP ANIMATION

Be sure to visit our Blog: [LexaBlog](#)

Lexalytics, Inc. is the leader in text analytics software, providing entity extraction, sentiment analysis, document summarization and thematic extraction for the PR/Marketing, eDiscovery, Business Intelligence and Financial Services markets.

Today is Sunday, 27th July 2008

Latest News

Jun 08

EContent explores how SmartBrief integrated Lexalytics' Classification Solution

[more](#)

May 08

Boston Globe's Dave Copeland reports how companies are "listening" to customers online and gets input from Jeff Catlin

[more](#)

April 08

Jeff Catlin joins DM Radio for discussion on text mining...Listen Here

[more](#)

Welcome ...



Lexalytics is offering a fully-functional version of its text analytics software for you to try - for free! Simply click the image to the left and download your trial version. The Text Analytics Trial processes up to 50 documents of various types, including HTML, Word, PowerPoint, PDF and text. You can also select from a list of full-text RSS feeds for analysis. Once the information is processed, it returns pertinent metadata such as:

- People and companies in the Fortune 500
- Sentiment and Tone for those extracted entities
- Summarization of the documents

- [-] City
 - Amsterdam (1)
 - Boston (1)
- [-] Company
 - Analytics Firm Datops (3)
 - Boeing (1)
 - Cisco (1)
 - IDC (2)
 - Information Access Softw...
 - Inight (4)
 - Merrill Lynch (1)
 - Pfizer (1)
 - Philips (1)
 - Procter & Gamble (1)
 - Reuters (1)
- + Country
- + IndustryTerm
- + Organization
- + Person
- + Product

Join the Blog

About us

Get Alerts

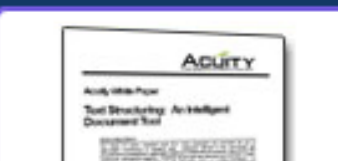
News & updates!

NEW!

Click here for exclusive interviews with...



Stefano Maria De' Rossi
CRM Manager at Telecom Italia



[Text Mining Boom Discussion to Take Place in Boston this June 2007](#)

Since 2005 senior level executives from data management, business intelligence and knowledge management have been coming together to discuss the amazing applications of Text Mining Technologies...

[The Content Intelligence \(CI\) market: to be or not to be?](#)

So does Ian Hersey, founder of Inight, think that the CI market is 'real' yet? "Definitely for the U.S. Federal government. They are ahead on text" he says. Inight has more than 125 government agency customers. The Financial Services industry is also showing interest, for example to extract and investigate data from regulatory archives and to spot fraud and insider trading. Publishers, Pharmaceutical and High Tech companies are also keen. Inight counts opinion leaders such as Procter & Gamble, Cisco, Boeing, Reuters, and Merrill Lynch among their 450 customers.

Ian believes that European adoption is lagging the U.S. currently. However, things are changing. The European Text Analytics Summit in Amsterdam in Apr includes speakers from Telecom Italia, Pfizer and Philips: "come and find out about their real life deployments" screams the promotion. The big vendor names at this Summit are SAS and SPSS, but the industry rising stars are Inight, Clarabridge, and Attensity.

[LexisNexis in Europe Acquires Analytics Firm Datops](#)

Datops Enables Analysis of Web and LexisNexis Premium Information Sources



Username: * Password: *

[Forgot Username](#)

- [Home](#)
- [About](#)
- [Gallery](#)
- [Tools](#)
- [Forums](#)
- [Blog](#)
- [News](#)
- [Register](#)



00:00 CALAIS: CONNECT. EVERYTHING

Calais: Connect. Everything.

We want to make all the world's content more accessible, interoperable and valuable. Some call it Web 2.0, Web 3.0, the Semantic Web or the Giant Global Graph - we call our piece of it Calais.

Calais is a rapidly growing toolkit of capabilities that allow you to readily incorporate state-of-the-art semantic functionality within your blog, content management system, website or application.

[▶ Get Started](#)

Latest News

Welcome to Calais 2.1!
This release includes many new features and tools to make Calais more usable, useful and accessible. Please take a

Featured Application

Meet Calais Tagaroo.
Tagaroo provides automated tag generation



Calais and You

Choose one of the user types below to see what you can do with Calais.



The Calais initiative is about enabling semantic applications by providing a metadata generation web service, sample applications using that service to jumpstart development efforts, and support for developers.

The Calais Web Service

The Calais web service automatically attaches rich semantic metadata to the content you submit. Using natural language processing, machine learning and other methods, Calais categorizes and links your document with entities (people, places, organizations, etc.), facts (person "x" works for company "y"), and events (person "z" was appointed chairman of company "y" on date "x").

Enter text here:




```
Sydney airport trial: check-in anger
Email Printer friendly version Normal font Large font Daniel Emerson
July 28, 2008 - 2:24PM
```

Page 1 of 2 | Single page

```
An ill-timed trial of a new baggage check-in system is being blamed by angry passengers for
chaotic scenes at Qantas's Sydney domestic terminal.
```

```
Flights were delayed, passengers said they were queueing for more than an hour, and scores
of passengers found themselves waiting in line at the time their planes were due to take
off today.
```

Submit

[Show RDF](#)[Entry Page](#)Entities:   City Company Facility PersonEvents &  

Title

1217231885556-3CB60A39-646981

Date

2008-07-28

Body

Sydney airport trial: check-in anger

Email Printer friendly version Normal font Large font [Daniel Emerson](#)

July 28, 2008 - 2:24PM

Page 1 of 2 | Single page

An ill-timed trial of a new baggage check-in system is being blamed by angry passengers for chaotic scenes at [Qantas's](#) [Sydney domestic terminal](#).

Flights were delayed, passengers said they were queueing for more than an hour, and scores of passengers found themselves waiting in line at the time their planes were due to take off today.

Issues

- Complex names:
 - John and Mary Smith
 - Proctor and Gamble
- Book and movie titles
- Not Every Initcapped String Is A Name

Text Analytics Components

- Search
- Text Categorisation and Clustering
- Entity Recognition
- Cross-Document Name Matching
- Text Summarisation
- Event Recognition
- Sentiment Analysis
- Question Answering



Feedback?

Click here.



Welcome, Guest

[Register](#) | [Sign in](#) | [Help](#)

Company Search

People Search

Job Search

Robert Dale

Search

[Advanced Search](#)
[Premium Services](#)

Refine Search

You Searched for:

Person Name

Robert Dale

Refine by:

- Title
- Company Name
- Industry Keywords
- Location

Refine

[Advanced Search](#)

Results 1-25 of 374 people

Sort By: Relevance (Default)

Introducing

Time saving tools help you
get better information faster.

Try it now

Name	Title	Company	
Dale, Bob <input type="checkbox"/>		Dale Studios Inc	✉ ☎
Dale, Robert <input type="checkbox"/>	Professor	Macquarie University	✉
Dale, Robin L. <input type="checkbox"/>	Associate Editor RLG DigiNews	The Research Libraries Group Inc	✉ ☎
Dale, Robert V. <input type="checkbox"/>	Board Member	Genesco Inc.	✉
Dale, Bob B. <input type="checkbox"/>	Vice President , Research and Development	Polymer Group , Inc.	
Dale, Bobby T. <input type="checkbox"/>	the Reviewer	Leeking Inc.	✉

3giphone.mobefun.com *mobile subscription

Professor⁴
[Macquarie University](#)
 Headquarters Address:
 Balaclava Rd
 North Ryde, NSW 2109
 Australia
 Website: www.mq.edu.au
 Phone: +61 2 9850 4185
 Fax: +61 2 9850 9231

Macquarie University in Sydney, Australia

Director of the Centre for Language Technology⁴
[Macquarie University](#)

Business Manager⁵
[Macquarie University](#)

Director
 Macquarie University's Centre for Language
 Technology

Teacher
[Centre for Language Technology](#)

Editor
[Journal of Computational Linguistics](#)

Director
[Microsoft Research Institute at Macquarie University](#)

Chief Technology Officer

1. www.ai08.org
www.ai08.org/call4paper.htm - [Cached]
 Published on: 5/30/2008 Last Visited: 6/3/2008

[Robert Dale, Macquarie University, Australia](#)

2. Research Active
www.researchactive.mq.edu.au/s - [Cached]
 Published on: 5/2/2008 Last Visited: 5/2/2008

Professor [Robert Dale](#)

More than number crunching

"Traditionally, computational work in the world of finance has been to do with number crunching stock prices and market movements, the clever analysis of numerical trading data, to try and find out what's going on in the marketplace," explains CLT Director, Professor [Robert Dale](#).

"All the software is doing is looking for anything at all that distinguishes these market sensitive documents, without trying to understand them in any sense," says [Dale](#). "It's a dumb approach in a way, but it's actually very effective."

So effective, in fact, that not only did the computer come close to matching the accuracy of human attempts to classify market sensitive documents, they even picked up some instances where humans had not classified appropriately.

"Other experiments that have been done on wider

Robert's Network

Join Robert's Network

[Click here to create your](#)

XING Network



Robert Dale AU Search [Clear](#)
First Name Last Name City State Country

Robert Dale, Australia

[Personal Web Spaces - MySpace](#)

Robert. "", Male, 21, Queensland, **AU** ...
[search.myspace.com/index.cfm](#) - Deep Web

[Professional Profile & Networking - LinkedIn](#)

Robert Dale, , **AU**, Professor ...
[www.linkedin.com](#)

[Business Contacts - Jigsaw \\$\\$\\$](#)

Robert Dale, North Ryde, **AU** - Professor, Macquarie University ... Dr **Robert Dale**,
Church Point, **AU** - Principal Consultant, Language Technology Group ...
[www.jigsaw.com](#) - Deep Web

[Patent Applications - US Patent & Trademark Office](#)

1, 20060041378, Method and system for adaptive navigation using a driver's route ...
[appft1.uspto.gov/netacgi](#) - Deep Web

[Case Records - Federal Court of Australia](#)

DALE, ROBERT, BRG204/2006, Nicol Robinson Halletts (A Firm) v **Robert Dale**,
BANKRUPTCY ...
[esearch.fedcourt.gov.au/Esearch](#) - Deep Web

[Nominal Rolls](#) (10)

DALE, ROBERT BARRY, NX6422, 21 Aug 1912, BOURKE, NSW, Army ... **DALE, ROBERT**
ELMO FALIERE, PA34, 9 Aug 1910, BIRKENHEAD, SA, RAN ...

Issues

- Name variation:
 - Alexander Smith, Alex Smith, A Smith, Mr Smith
 - IBM = Big Blue; Qantas = the Red Kangaroo
- Two sides to the coin:
 - One name, many people
 - Many names, one person

Text Analytics Components

- Search
- Text Categorisation and Clustering
- Named Entity Recognition
- Cross-Document Name Matching
- Text Summarisation
- Event Recognition
- Sentiment Analysis
- Question Answering

Types of Summarisation

- Informative vs Indicative
- Extractive vs Abstractive

Informative vs Indicative

- An informative summary represents or replaces the original document:
 - Needs to contain all the core information and omit ancillary information
- An indicative summary suggests the contents of the article without giving away detail on the actual content
 - Can serve to entice the user into retrieving the full form
 - Examples: book jackets, card catalog entries, movie trailers

Extractive vs Abstractive

- Extractive summaries (extracts) are made up of fragments (whole sentences or phrases) of the source text.
- Abstractive summaries (abstracts) may contain content not directly present in the source – may require generalization or abstraction

Techniques for Extractive Summarisation

- Basic idea:
 - Extract key sentences
- Approaches:
 - Rule-based vs machine-learned
- Relevant features
 - Position of sentence in text; frequency of words appearing in sentence; density of named entities

Related Techniques

- Sentence compression
 - Geoff Dixon, CEO of Qantas, today announced that the company will sell off ...
 - Qantas will sell off ...
- Microsoft's IntelliShrink
 - DrctDepositPymntsWllBAvbleInYrAcctWthn3BsnssDysFrmPymnt
DteBlw
- 'Snippet' or 'teaser' extraction

Web [Images](#) [Maps](#) [News](#) [Shopping](#) [Gmail](#) [more](#) [Sign in](#)



text analytics

Search

[Advanced Search](#)
[Preferences](#)

Web

Results 1 - 10 of about 547,000 for [text analytics](#). (0.16 seconds)

[Text analytics](#) - Wikipedia, the free encyclopedia

The term **text analytics** describes a set of linguistic, lexical, pattern recognition, extraction, tagging/structuring, visualization, and predictive ...

[en.wikipedia.org/wiki/Text_analytics](#) - 32k - [Cached](#) - [Similar pages](#)

[Text mining](#) - Wikipedia, the free encyclopedia

Inxight - provider of **text analytics**, search, and unstructured ... VantagePoint [3] - **Text mining software** which includes tools for data cleanup, **analysis**, ...

[en.wikipedia.org/wiki/Text_mining](#) - 55k - [Cached](#) - [Similar pages](#)

[More results from en.wikipedia.org »](#)

[4th annual text analytics summit](#)

The 4th Annual **Text Analytics** Summit 2008 is taking place 16 - 17 June, 2008 in Boston. If you're a user of **text analytics** technology in any application or ...

[www.textanalyticsnews.com/](#) - 27k - [Cached](#) - [Similar pages](#)

[4th annual text analytics summit](#)

Brand new to this years summit, **Text Analytics** News, is proud to present two pre ... Looking to implement **Text Analytics** solutions in your organization? ...

[www.textanalyticsnews.com/4thannual08/agenda.shtml](#) - 72k - [Cached](#) - [Similar pages](#)

[More results from www.textanalyticsnews.com »](#)

[Text Analytics Wiki: Welcome to the Text Analytics Wiki](#)

This wiki aims to be a one-stop site for everything related to **Text Analytics** (also known as

Sponsored Links

[Free SMS Text](#)

Great Deal On Texts, Unlimited SMS \$3 Per Week, 1 Week Free!
[www.YourMoco.com](#)

[Textkernel Web Mining](#)

Extract valuable data feeds from the web...
[www.textkernel.com](#)

[Text Mining Solutions](#)

For Unstructured Multilingual Data. Foreign Language **Text Analysis**.
[www.basistech.com](#)

[Lexalytics, Inc](#)

Metadata extraction engine
Entities, Sentiment, Quotes, Themes
[www.lexalytics.com](#)

[VisualText](#)

Build it. NLP, IE, parsers, smart

Issues Impacting on Performance

- Real summarisation requires real understanding
- Quality of 'knowledge-free' summarisation relies on aspects of the document other than content
- Problems with out-of-context or dangling references:
 - 'He said the price offered in the buyback scheme would be ...'



IMAGINE IT

[WHO WE ARE](#)[CHALLENGES WE SOLVE](#)[MARKETS WE SERVE](#)[TECHNOLOGIES WE PROVIDE](#)[OUR PARTNERS](#)[Integrated Search](#)[Categorization](#)[Contextual Explanation](#)[Conceptual Comparison](#)[Multilingual and Cross-Lingual Text Analysis](#)[Name Tracking and Disambiguation](#)[Relationship Discovery](#)[Summarization](#)

SUMMARIZATION

One of the biggest challenges of analyzing large volumes of material is in knowing where to begin. Content Analyst technology dramatically accelerates the analysis process by instantly identifying key sentences that most accurately represent the prevailing concepts within a single document or across a collection of documents.

The interface for this feature allows the user to quickly change the length of the summaries they would like to view. Content Analyst technology can summarize any document that can be represented in the Unicode® encoding system.

[Use Case Example 1:](#)

SPOTLIGHT

SOLUTION SPOTLIGHT

ANYDOC/CONTENT ANALYST

Infiniworkx gives you the freedom and flexibility to incorporate automated document and data capture, classification and processing into a workflow that matches your company's existing business processes.

[LEARN MORE](#)[SEE MORE SOLUTIONS](#)

MARKET SPOTLIGHT

Intl Legal Technology Association 2008

August 25-28, 2008
Gaylord Texan,
Grabevine, Texas



TERAGRAM
A sas COMPANY

TECHNOLOGY

[Technology](#) **[Solutions](#)** [OEM Products](#) [Services](#) [About](#)

[CONTACT](#)
[NEWS](#) • [JOBS](#)

TERAGRAM SUMMARIZER™

The Teragram Summarizer distills documents and creates concise summaries. The Teragram Summarizer allows your users to access your information faster and more efficiently and even enables the distribution of documents on devices with small viewing areas such as cell phones and SMS. Based on Teragram Linguistic and parsing technologies, The Teragram Summarizer composes short summaries of variable lengths based on your needs. The Teragram Summarizer condenses documents in real time and is deployed into your workflow. It drastically improves the efficiency at which information is conveyed by allowing the readers to focus on the important concepts. The Teragram Summarizer uses Teragram's core Linguistic technologies to produce editor quality abstracts.

This technology can be deployed in as a client server environment and is delivered either as stand-alone software or as a software development toolkit. It is available for European and Asian languages.

Copyright © Teragram Corporation

>> SEARCH

- Overview
- Federated & Desktop Search
 - Inight SmartDiscovery Awareness Server
 - Inight Search Extender for Google Desktop
- Text Analysis
 - Inight SmartDiscovery Extraction Server
 - Inight Categorizer
 - Inight Summarizer
- Data Cleansing
 - Inight SmartDiscovery Metadata Management System
- Visualization
 - Inight SmartDiscovery VizServer
- OEM Solutions
 - Inight LinguistX Platform
 - Inight ThingFinder
 - Inight StarTree
 - Inight TimeWall
 - Inight TableLens
 - > Inight Summarizer

Inight Summarizer™



Fast abstracts of any document

The Inight Summarizer™ SDK (software development kit) allowing applications developers to incorporate into their products an intelligent solution to many problems inherent to online searches. By focusing on the relevant key sentences contained within a document, Summarizer technology enables end-users to browse quickly through volumes of information and extract the documents most applicable to their search requirements. Summarizer utilizes consistent sentence-selection criteria that match the conceptual content of documents. End-users save precious time and effort since they do not have to download and read each retrieved document to determine its relevancy. They experience easier navigation through Web sites, faster access to pertinent information and increased productivity.

Summarizer can summarize a typical document in a fraction of a second and so enables users to more of their time utilizing data, not just trying to find it. Also, to expedite search functions, Summarizer can be "trained" to find key sentences based on the structure of specific document types. Information is accessible by the length of key sentences or the number of key phrases. The end-user can control the weight of phrases by query phrase or drop phrases.

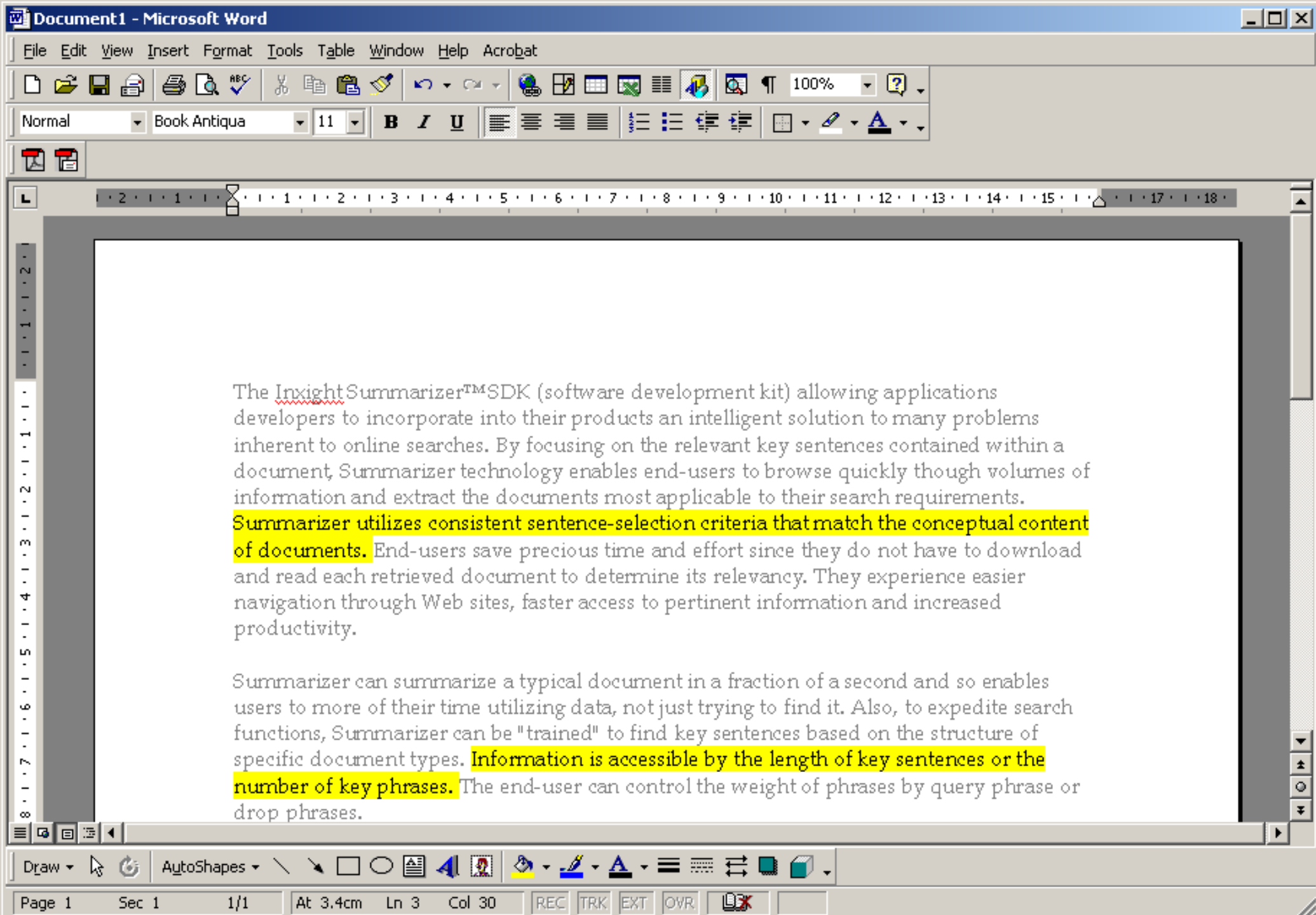
- > Overview
- Features
- Technical Specs

White Papers

-  ["Inight Summarizer: Managing the Information Deluge"](#)
-  ["Linguistics: Adding Value to e-Publishing/e-Content"](#)

Case Studies

-  ["NewsReel and Inight Give Business Value to the Web"](#)



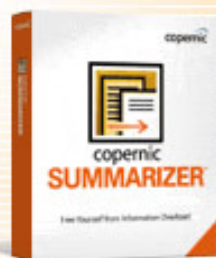


Search and Discover

Search the Web:

[Home](#)[Products](#)[Downloads](#)[Store](#)[Support](#)[Contact Us](#)[About Copernic](#)[English](#)[Français](#)[Products Home](#)[Copernic Agent](#)[Copernic Desktop Search](#)[Copernic Tracker](#)[Copernic Summarizer](#)[Copernic Meta](#)[home](#) / [products](#) / [copernic summarizer](#)

Copernic Summarizer



copernic
SUMMARIZER

**Free Yourself from
Information Overload**

- ✓ 30-Day Money-Back Guarantee
- 🔒 Secure Online Ordering
- 💰 Includes Free Software Updates

\$59.95 Buy Now

Download Trial

Free 30-Day Trial



→ Demo

Copernic Summarizer
Animated Flash Tour

Quick Links

- [Copernic Summarizer](#)
- [Download Trial](#)
- [Buy It](#)
- [Demo](#)
- [Features](#)
- [Installation Help](#)
- [What's New?](#)
- [Screen Shots](#)
- [FAQ](#)

Copernic's easy-to-use summarizing software dramatically increases your productivity and efficiency by creating concise document summaries of any file or Web page so you spend considerably less time reading without missing any important information.

Using sophisticated statistical and linguistic algorithms, it pinpoints the key concepts and extracts the most relevant sentences, resulting in a Web site or document summary that is a shorter, condensed version of the original text.

Feature Overview

Creates document summaries of any text file (including PDFs) or Web page

Awards

Text Analytics Components

- Search
- Text Categorisation and Clustering
- Named Entity Recognition
- Cross-Document Name Matching
- Text Summarisation
- Event Recognition
- Sentiment Analysis
- Question Answering

Event Recognition

- Working out who did what to whom and when
- Also known as 'Information Extraction'

Doing It Yourself

- GATE, from the University of Sheffield
- IBM's UIMA

An Example Document

San Salvador, 19 Apr 89 (ACAN-EFE) -- [TEXT] Salvadoran President-elect Alfredo Cristiani condemned the terrorist killing of Attorney General Roberto Garcia Alvarado and accused the Farabundo Marti National Liberation Front (FMLN) of the crime.

...
Garcia Alvarado, 56, was
guerrillas on his vehicle
intersection in downtown

...
Vice President-elect Fra
general's car stopped a
an individual placed a b

Incident: Date	19-Apr-89
Incident: Location	El Salvador: San Salvador (CITY)
Incident: Type	Bombing
Perpetrator: Individual ID	urban guerrillas
Perpetrator: Organization ID	FMLN
Perpetrator: Confidence	Suspected or Accused by Authorities: FMLN
Physical Target: Description	vehicle
Physical Target: Effect	Some Damage: vehicle
Human Target: Name	Roberto Garcia Alvarado
Human Target: Description	attorney general: Roberto Garcia Alvarado
Human Target: Effect	Death: Roberto Garcia Alvarado

Techniques

- Identify named entities (see earlier)
- Identify stated relationships between named entities:
 - Generally achieved via pattern matching
 - As pattern matching becomes more complex, it effectively becomes syntactic parsing

Example: GainSpring

- Problem:
 - 100k+ company announcements per year issued through the Australian Stock Exchange
- Goal:
 - Extracting key facts from documents automatically and deliver them via different mechanisms – web page, email alerts, SMS, synthesized voice

Getting started

Find out about...

- HOME
- MARKET STATISTICS
- COMPANY RESEARCH
- ASX MARKETS
- ASX SHAREHOLDER INFORMATION
- FLOATS
- INVESTOR EDUCATION
- ABOUT ASX

CBA company announcements

The following announcements have been found for [CBA](#) in your search period.

[Search again](#)



- SITE SEARCH
- SITE MAP
- GLOSSARY

	Summary	Published	Sensitive
<input type="checkbox"/>	Change in substantial shareholding for KYC	20/04/2000 12:08:39	
<input type="checkbox"/>	Change in substantial shareholding for KYC	20/04/2000 12:08:39	
<input type="checkbox"/>	Change in substantial shareholding for BQD	08/02/2000 17:23:00	
<input type="checkbox"/>	Change in substantial shareholding for BQD	08/02/2000 17:23:00	



Notice of change of interests of substantial holder

To: Company Name/Scheme SOUTHERN STAR GROUP LIMITED ("SSR")
 ACN/ARSN 003 321 266

1. Details of substantial holder (1)

Name SOUTHERN CROSS BROADCASTING (AUSTRALIA) LIMITED and each of the companies set out in Annexure A ("SCB Group")
 ACN/ARSN (if applicable) 006 186 974

There was a change in the interests of the substantial holder on 31/03/04, 1/04/04 – 2/04/04, 5/04/04 – 9/04/04, 13/04/04
 The previous notice was given to the company on 31/03/04
 The previous notice was dated 31/03/04

2. Previous and present voting power

The total number of votes attached to all the voting shares in the company or voting interests in the scheme that the substantial holder or an associate (2) had a relevant interest (3) in when last required, and when now required, to give a substantial holding notice to the company or scheme, are as follows:

Class of securities (4)	Previous notice		Present notice	
	Person's votes	Voting power (5)	Person's votes	Voting power (5)
Ordinary	77,017,392	67.26	96,583,668	84.31%

3. Changes in relevant interests

Particulars of each change in, or change in the nature of, a relevant interest of the substantial holder or an associate in voting securities of the company or scheme, since the substantial holder was last required to give a substantial holding notice to the company or scheme, are as follows:

Example Template Fill: Becoming a Substantial Holder

<u>Element</u>	<u>Example Contents</u>
DocumentCategory	02001
AcquiringPartyASX	TCN
AcquiringParty	TCNZ Australia Investments Pty Ltd
AcquiredPartyASX	AAP
AcquiredParty	AAPT Limited
DateOfTransaction	4/07/1999
NumberOfShares	243,756,813
ShareType	ordinary shares
PercentageOfShares	79.90%


[Home](#) > [Demo](#) > Recent Announcements

ASX Code	Summary	Published	Sensitive
CBA	Change in substantial holding for PWR	2000-12-29 16:52:35	
PWR	Change in substantial holding from CBA	2000-12-29 16:52:35	
PWR	Change in substantial holding from CBA	2000-12-29 16:52:35	
CBA	Change in substantial holding for HPX	2000-12-29 16:46:28	
HPX	Change in substantial	Commonwealth 16:46:28	

[more announcements...](#)

Shareholding Company: Bank of Australia (CBA)

Issuing Company: Hpal Limited (HPX)

Date Of Transaction: 2000-12-28

Change: ↑ from 6,300,000 (5.67) to 7,778,500 (7.00) ordinary shares


[Who we are](#)
[What we do](#)
[Technologies](#)
[Partners](#)
[Members area](#)
[Home](#) > [Demo](#) > [Company details](#)

Latest Director's transactions

Directors	Type	Shares	Date Notified	Previous Notification	Announcement
David Victor Murray	Change	↓ 44,372	2000-07-01	1998-04-18	Announcement
John Theodore Ralph	Change	↓ 2,605	2000-04-14	1999-09-30	Announcement
John Theodore Ralph	Change	↓ 2,605	2000-04-14	1999-09-30	Announcement
John Michael Schubert	Change	↓ 6,476	2000-04-14	1999-11-19	Announcement
Norman Ross Adler	Change	↓ 6,222	2000-04-14	1999-09-30	Announcement
Barbara Kay Ward	Change	↓ 1,837	2000-04-14	1999-09-30	Announcement
Frank Joseph Swan	Change	↓ 1,922	2000-12-04	1999-09-30	Announcement
Fergus Denis Ryan	Change	↓ 4,000	2000-04-14		Announcement
Anna Christina Booth	Change	↓ 1,131	2000-04-13	1999-09-30	Announcement
Warwick Gordon Kent	Other	↓ 6,237			Announcement
Warwick Gordon Kent	Other	↓ 6,237			Announcement

A JAPE Rule from GATE

Rule: FirstName

```
{Lookup.majorType == person_first}:person
```

```
-->
```

```
{  
    gate.AnnotationSet person = (gate.AnnotationSet)bindings.get("person");  
    gate.Annotation personAnn = (gate.Annotation)person.iterator().next();  
    gate.FeatureMap features = Factory.newFeatureMap();  
    features.put("gender", personAnn.getFeatures().get("minorType"));  
    features.put("rule", "FirstName");  
    outputAS.add(person.firstChild(), person.lastChild(),  
                "FirstPerson", features);  
}
```

Issues Impacting on Performance

- Difficult to achieve high recall and precision because of variations in language
- Success depends on the predictability of document content
- High performance really requires full natural language processing

Text Analytics Components

- Search
- Text Categorisation and Clustering
- Named Entity Recognition
- Cross-Document Name Matching
- Text Summarisation
- Event Recognition
- Sentiment Analysis
- Question Answering

Sentiment Analysis

- Goal:
 - To automatically determine the polarity of sentiment expressed towards a company, product or individual
- Increasingly seen as important because of the blogosphere
- This year's hot topic, aka 'Voice of the Customer'

Approaches to Sentiment Analysis

- Identify specific key words that are indicative: The General Inquirer
 - Positives (1914 words) eg: decent, ardent, wholesome, triumph, ...
 - Negatives (2293 words) eg: allege, despicable, harmful, worry, terrorize, ...



PRODUCTS

Please select a product

SOLUTIONS

- Document management →
- SharePoint replication →
- Maritime replication →
- Military SharePoint replication →

Sentiment For BI, PR, and marketing.

Sentiment provides communications professionals with an instant understanding of the news and its impact on their brand.

Sentiment puts the world's news at the fingertips of BI, PR and marketing professionals to help them make better informed decisions.

Product information

Technical Specifications
Please contact us for detail on Sentiment's specifications.

System requirements
Intel Pentium 4 or eq
512 Mb Ram & 200 M
space Windows 2000

[home](#)[learn more](#)[what you'll get](#)[why us?](#)[who is this for?](#)[typical uses](#)[free trial](#)[order now](#)[contact](#)**SENTIMENT METRICS**[login](#)

Blog and Media Monitoring - Listen . Learn . Profit

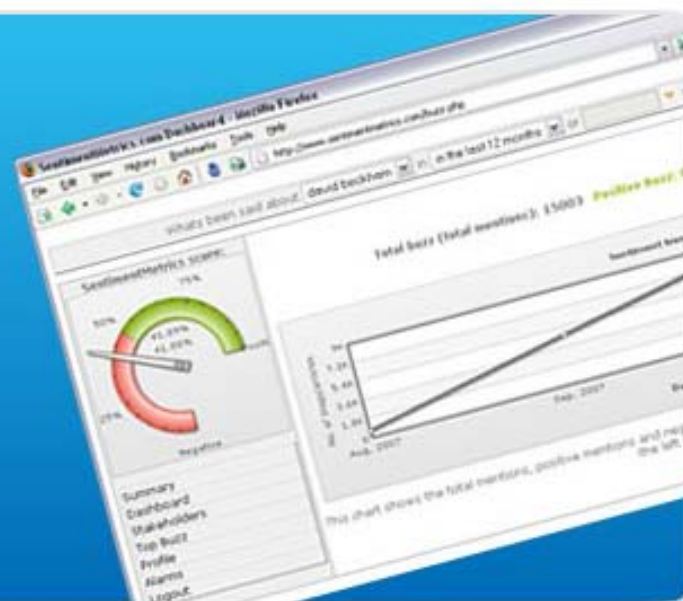
14 Days Trial



sentiment metrics

Try our blog and media monitoring service for 14 days absolutely FREE!
No charge, No contract, No worries!

Click here to find out what people think about your brand, products and services.



1. Listen

SentimentMetrics locates and stores all the content about your brand, from millions of blogs

2. Learn

SentimentMetrics then cleans and analyses the data, presenting it in clear easy-to-read tables

3. Profit

The insight you gain from SentimentMetrics allows you to grow customer satisfaction, increase

Are you Listening???

People are talking about you, your brand, your products, your business. On tens of millions of blogs, forums, message boards and on mainstream media. What they are saying



[Contact Us](#) | [Sup](#)

- Home
- Products & Services
- Resource Center
- Clients
- About Us
- Biz360 Client L

- Biz360 »**
- Opinion Insights »
- Media Insights »
- Social Media Insights »

Transforming information into media and market insights that build your business.



Introducing Opinion Insights this Fall, 2008

Fact: Online product opinions influence consumer purchase decisions.

Biz360 in the News

Biz360 Names Darian Shimy Vice President of Technology

[request a demo](#)

[contact us »](#)



Contact | Sitemap | Search this site:



Products & Services Solutions Approach Press & Events CGM Resources About

Welcome to Nielsen **BuzzMetrics**, the global measurement standard in Consumer-Generated Media.

[learn more >>](#)



Who We Are

Our Products

Resources

Nielsen BuzzMetrics is the global measurement standard in Consumer-Generated Media. With solid data-mining technology, superb research and Nielsen's unrivaled experience in media measurement and client services, we help today's companies, brands and business professionals better understand the influence and impact of CGM on products.

Our **BrandPulse** suite of products measure consumer-generated media (CGM) and online word of mouth to help companies understand and track consumer "buzz," including opinions, preferences, issues and coming trends.

From your desk, log into a **BrandPulse** project 24/7 to track data driven reports

Connecting the Dots Blog

Visit our Blog to read insights into CGM, analytics and much more!

[Click Here >](#)

Technology at Work

To see BuzzMetrics' text-

Upcoming Webcasts

Upcoming Events

[Click Here to See Schedule of Upcoming Events and Webcasts](#)

Contact Us

Want to know more about Nielsen BuzzMetrics products and services?

Contact us today for additional information, pricing, demonstrations and more:

Email:
sales@buzzmetrics.com

Amazon.com: Apple iPod nano 8 GB Green (3rd Generation): Electronics

http://www.amazon.com/Apple-iPod-nano-Green-Ge

amazon.com Hello. Sign in to get personalized recommendations. New customer? Start here. Need a Last-Minute Father's Day Gift?

Your Amazon.com Today's Deals Gifts & Wish Lists Gift Cards Your Account | Help

Shop All Departments Search Electronics GO Cart Your Lists

Electronics Browse Brands Bestsellers Camera & Photo Computers & PC Hardware Audio, TV & Home Theater Cell Phones & Service Office Electronics GPS Today's Deals

Prime To get this item by **Wednesday, Jun 18** order within 20hr 53min.

Get Free Shipping for a full month with a Free Trial of Amazon Prime [learn more](#) FREE Upgrade to Two-Day Shipping on this item with Amazon Prime

Apple iPod nano 8 GB Green (3rd Generation)

Other products by [Apple](#)

★★★★☆ (738 customer reviews) [More about this product](#)

Size Name: 8 GB
First Select Size Name

Color Name: Green
Then Select Color Name

List Price: ~~\$199.00~~
Price: **\$179.00** & this item ships for **FREE** with **Super Saver Shipping**. [Details](#)
You Save: **\$20.00 (10%)**

In Stock.
Ships from and sold by **Amazon.com**. Gift-wrap

Quantity: 1 Add to Shopping Cart
or Sign in to turn on 1-Click ordering.

More Buying Choices

J&R Music and Computer World	Add to Cart
\$179.00 + \$8.35 shipping In Stock	
TigerDirect	Add to Cart
\$188.50 + \$5.24 shipping In Stock	
Outlet	Add to Cart
\$179.00 + \$15.00 shipping	

See larger image and other views

Pluribo computed the following summary by analyzing hundreds of reviews (see numerical scores)

This is a good choice; despite concerns with the tendency to scratch, people are happy with the design, display, and great sound.

Done

Amazon.com: Apple iPod nano 8 GB Green (3rd Generation): Electronics

http://www.amazon.com/Apple-iPod-nano-Green-Ge

amazon.com Hello. Sign in to get [personalized recommendations](#). New customer? [Start here](#). Need a Last-Minute Father's Day Gift?

Your Amazon.com Today's Deals Gifts & Wish Lists Gift Cards Your Account | Help

Shop All Departments Search Electronics GO Cart Your Lists

Electronics Browse Brands Bestsellers Camera & Photo Computers & PC Hardware Audio, TV & Home Theater Cell Phones & Service Office Electronics GPS Today's Deals

Prime To get Get Free Shipping Free Trial of Am

Overall opinion about **scratching** (relative to similar products):



[bad] [good]

What people are saying:

- *...scratches...
- *...scratched...
- *...scratch...
- *...back scratches...
- *...few small scratches...
- *...chance for scratches...
- *...impossible to scratch...
- *...backing scratches...
- *...scratched cd...
- *...occasional scratch...
- *...probably scratch...
- *...nice scratch...
- *...tiny scratch...

Quantity: 1 Add to Shopping Cart or Sign in to turn on 1-Click ordering.

More Buying Choices

J&R Music and Computer World **\$179.00** + \$8.35 shipping In Stock Add to Cart

TigerDirect **\$188.50** + \$5.24 shipping In Stock Add to Cart

Outlet **\$179.00** + \$15.00 shipping Add to Cart

See larger image and other views

Pluribo computed the following summary by analyzing... This is a good choice; despite concerns with the tendency to scratch, people are happy with the design, display, and great sound.

Done

Amazon.com: Apple iPod nano 8 GB Green (3rd Generation): Electronics

http://www.amazon.com/Apple-iPod-nano-Green-Ge

amazon.com Hello. Sign in to get personalized recommendations. New customer? Start here. Need a Last-Minute Father's Day Gift?

Your Amazon.com Today's Deals Gifts & Wish Lists Gift Cards Your Account | Help

Shop All Departments Search Electronics GO Cart Your Lists

Electronics Browse Brands Bestsellers Camera & Photo Computers & PC Hardware Audio, TV & Home Theater Cell Phones & Service Office Electronics GPS Today's Deals

Prime To get this item by **Wednesday**, Jun 18 order within 20hr 52min.


Get Free Shipping for a full month with a Free Trial of Amazon Prime [learn more](#) FREE Upgrade to Two-Day Shipping on this item with Amazon Prime

Apple iPod nano 8 GB Green (3rd Generation)

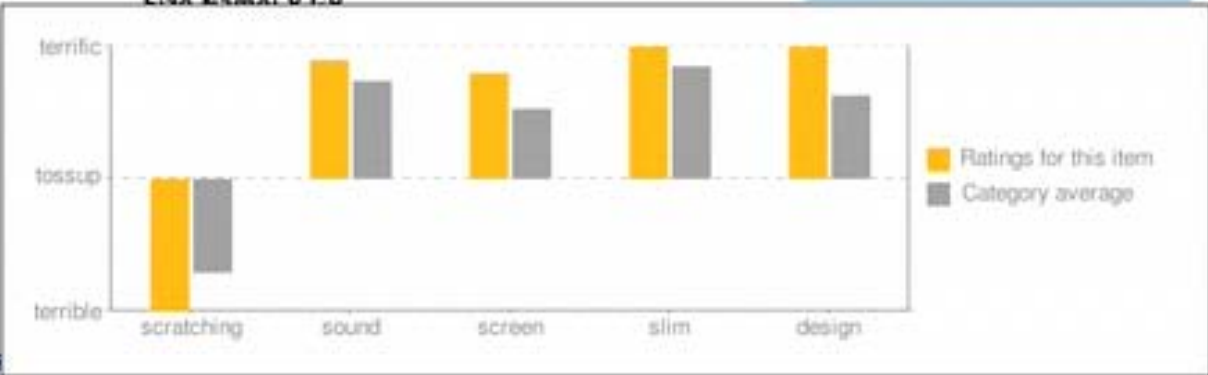
Other products by [Apple](#)

★★★★★ (738 customer reviews) [More about this product](#)

Quantity: 1 Add to Shopping Cart or Sign in to turn on 1-Click ordering.



See larger image and other views



Category	Ratings for this item	Category average
scratching	Low	Low
sound	High	Medium
screen	High	Medium
slim	High	Medium
design	High	Medium

Pluribo computed the following summary by analyzing hundreds of reviews (see numerical scores)

This is a good choice; despite concerns with the tendency to scratch, people are happy with the design, display, and great sound.

Why Sentiment Analysis is Hard

- This laptop is a great deal.
- A great deal of media attention surrounded the release of the new laptop.
- This laptop is a great deal ... and I've got a nice bridge you might be interested in.

[Example from Lilian Lee]

Text Analytics Components

- Search
- Text Categorisation and Clustering
- Named Entity Recognition
- Cross-Document Name Matching
- Text Summarisation
- Event Recognition
- Sentiment Analysis
- Question Answering

Question Answering

- Current paradigm:
 - Type in a query, get back a list of documents
- Future paradigm:
 - Type in a question, get back an answer

Powerset

Powerlabs : Sign

Ready. Powerset. Go.

search wikipedia articles

Explore

search tips & examples

watch demo video



Discover Factz

For most people, places and things, Powerset shows a summary of Factz from across Wikipedia.

Unlock Meaning

Powerset finds articles related to the meaning of your query. And sometimes even direct answers.

Scan Summaries

Powerset summarizes pages for easy b with tools that follow you as you read a explore.



what is text analytics?

search

Text analytics Text mining

source: freebase (view topic) ?

watch demo video

ADD THIS

add Powerset to your search bar

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... Read enhanced Wikipedia article

Wikipedia Articles: results 1 - 10 of 373 advanced ?

- Text analytics The term text analytics describes a set of linguistic, lexical, pattern recognition, extraction, tagging/structuring, visualization, and predictive techniques.
- Noisy text analytics While Text analytics is a growing and mature field that has great value because of the huge amounts of data being produced, processing of noisy text is gaining in importance because a lot of common applications produce noisy text data.
- Text mining Fair Isaac - leading provider of decision management solutions powered by advanced analytics (includes text analytics).
- Infonic Text Analytics
- ClearForest ClearForest is a software company that develops and markets text analytics and



BETA
hakia[®]

Enter a question, phrase, or keywords

A new
Semantic Search Engine
dedicated to quality

Compare hakia

longest strike in the history of railroads - [hakia vs Google](#)

Syndication web services

Offer your visitors semantic search experience and much more.. See the [web services](#) or [search box](#) option.

Mobile Applications

Semantic Search on your cell phone. Download now!

BETA
hakia

Enter a question, phrase, or keywords

what is text analytics?

search



Brilliant query indeed. How about: Text Analytics ... offering can be broadly categorized as Information Extraction and Text Analytics: ... [See this page.](#)

[Text Analytics](#)

Text-Analytics provides a guide to text analytics news and techniques and to companies in the field. ...

Unified, integrated Text Analytics suite that offers ...

<http://www.text-analytics.com/index.html>

[Defining Text Analytics | Intelligent Enterprise Blog](#)

The Intelligent Enterprise Weblog presents the Web's best lineup of expert ... I've been writing and speaking and consulting on text analytics for years. ...

http://www.intelligententerprise.com/blog/archives/2007/02/defining_...

[Defining Text Analytics](#)

In a nutshell, text analytics parses textual documents and transforms ... The low adoption rate shows that text analytics is just emerging as a new ...

<http://www.tdwi.org/Research/display.aspx?id=8355>

[Text analytics marketplace trends | Text Technologies](#)

It was tough to judge user demand at the recent Text Analytics Summit because, well, very few users showed up. And frankly, I wasn't as aggressive at pumping

<http://www.texttechnologies.com/2007/07/22/text-analytics-marketplac...>

[Text Analytics article](#)

semanti



Web [Images](#) [Maps](#) [News](#) [Shopping](#) [Gmail](#) [more](#)

[Sign in](#)



what is text analytics?

Search

[Advanced Search](#)
[Preferences](#)

Web

Results 1 - 10 of about 473,000,000 for what is [text analytics](#)?. (0.25 seconds)

[Web definitions for Text analytics](#)



Text analytics is a process of information extraction whose goal is to automatically extract structured or semistructured information from ...
en.wikipedia.org/wiki/Text_analytics - [Definition in context](#)

[Text analytics - Wikipedia, the free encyclopedia](#)

The term **text analytics** describes a set of linguistic, lexical, pattern recognition, extraction, tagging/structuring, visualization, and predictive ...

en.wikipedia.org/wiki/Text_analytics - 32k - [Cached](#) - [Similar pages](#)

[Text mining - Wikipedia, the free encyclopedia](#)

This Overall **Analysis** System for Intelligence Support (OASIS) integrates among the most advanced **text analytics** and **text mining** technologies available on ...

en.wikipedia.org/wiki/Text_mining - 55k - [Cached](#) - [Similar pages](#)

[More results from en.wikipedia.org »](#)

[TAPoR @ UAlberta - What is Text Analysis?](#)

A range of resources including information about **text analysis**, the TAPoR workshop, electronic **text** collections, and journals in the field of digital arts. ...

tapor.ualberta.ca/Resources/TAIntro/ - 55k - [Cached](#) - [Similar pages](#)

[Text Analytics Wiki: Welcome to the Text Analytics Wiki](#)

This wiki aims to be a one-stop site for everything related to **Text Analytics** (also known as **Text Mining** or **Information Extraction**). ...

textanalytics.wikidot.com/ - 20k - [Cached](#) - [Similar pages](#)

NEWSLETTER SIGN UP:

[WHO WE ARE](#) | [WHAT WE DO](#) | [CONTACT](#)
Searching: MEDLINE Abstracts. [Details](#)

For Cognition to better understand you, enter a phrase with **proper capitalization**.
brain encephalitis herpes lobe: 81 files —

[Human herpesvirus-6 encephalitis after unrelated cord blood transplantation.](#)

Tanaka M, Taguchi J, Hyo R, Kawano T, Hashimoto C, Motomura S, Kodama F, Kobayashi S, Okabe G, Maruta A, Nagao T and Ishigatsubo Y

(2005) Leuk Lymphoma, 46 (4) 561-6

Magnetic resonance imaging (MRI) of patient 1's **brain** showed low-intensity signals at the gyri of the bilateral lateral **lobes** on T1-weighted images and [...]

[Direct link \(no highlighting\)](#)

[Adult-onset MELAS presenting as herpes encephalitis.](#)

Sharfstein SR, Gordon MF, Libman RB and Malkin ES

(1999) Arch Neurol, 56 (2) 241-3

RESULTS: Magnetic resonance imaging scan of the **brain** showed hyperintensity on T2-weighted images involving temporal, parietal, and **occipital lobes** [...]

[Direct link \(no highlighting\)](#)

[Serial diffusion-weighted imaging of neonatal herpes encephalitis: a case report.](#)

Kubota T, Ito M, Maruyama K, Kato Y, Miyajima Y, Ogawa A, Kuno K, Okumura A and Watanabe K

(2007) Brain Dev, 29 (3) 171-3

The following word meanings were selected.
Use dropdown menus to change the meanings.

brain:

encephalitis:



[true knowledge]™

Home | Forum | Blog | Wiki | Recent Activity | League Table | Add Knowledge

Logged in as rdale | Log Out

Welcome to [true knowledge]™

[click here to say who you are](#)

What is [true knowledge]™?

[true knowledge]™ is a technology that can directly answer questions put to it on any topic.

What can I do with it?

You can **ask it questions**, such as: [What is the height of the Eiffel Tower in yards?](#) or [is canada a member of the g8?](#)

You can **view profiles**, such as: [Abe Lincoln](#) or [Queen Elizabeth](#)

You can also **add new knowledge**, such as: [Pigs are omnivores](#) or [Madonna is married](#).

You can look [here](#) to find out more about the beta test and what we're trying to do.

Simply type whatever you want in the box and press GO.

Why don't you **make your mark** by [adding something you know about](#)? As soon as you add



[true knowledge]™

How can I help?

Examples of what you can type:

[Mount Kilimanjaro is in Tanzania](#)

[What's the time?](#)

[\(More examples...\)](#)

[add knowledge](#)

Answering queries based on **101,595,254** facts about **3,676,655** things

[\(More details...\)](#)

Type in the box below to send us feedback including bug reports, feature requests or anything else about this page! (If you want to get an idea discussed, please use the [forum](#))

 (won't affect what you're doing)

[true knowledge]™

(More examples...)

[Home](#) | [Forum](#) | [Blog](#) | [Wiki](#) | [Recent Activity](#) | [League Table](#) | [Add Knowledge](#)

Here is the answer that I found:

- [the integer 4284379](#)

This conclusion is based on a single fact in the knowledge base:

[4284379 has been the population of Sydney since at least June 2007](#) ([endorse](#)) ([contradict](#))

Thanks to [londonhilton](#) who added knowledge necessary to answer this question.

(I understood your question to mean: **Which integer is the number of people normally resident in Sydney, the city in New South Wales, Australia at the current time?**)

There were other interpretations of your request based on rarely used names. If the interpretation above is not what you were expecting, you can [try again with the rare cases included](#).

I followed this chain of reasoning:

By calculation (generator: ["now@trueknowledge.com"]) I know that:

The current time is [03:29:34 UTC on the 28th of July 2008](#)

I know from locally stored knowledge that:

Fact 1: [4284379 is the population of Sydney](#) ([fact: ["124152385@trueknowledge.com"]])

Fact 1 is true for the period from [June 2007 onwards](#) ([fact: ["124152386@trueknowledge.com"]])

By calculation (generator: ["timeperiodtopoint@trueknowledge.com"]) I know that:

Fact 1 is true at [03:29:34 UTC on the 28th of July 2008](#)

How can I help?

Examples of what you can type:
[da vinci code author](#)
[How far is Sydney Australia from New York City](#)
[\(More examples...\)](#)

[true knowledge]™ [Home](#) | [Forum](#) | [Blog](#) | [Wiki](#) | [Recent Activity](#) | [League Table](#) | [Add Knowledge](#)

If there are any answers, I couldn't find any.

If you find this out from another source, you can help out by [adding this information](#) to the Knowledge Base.

(I understood your question to mean: **Which integer was the number of people normally resident in Sydney, the city in New South Wales, Australia in the year 1900?**)

There were other interpretations of your request based on rarely used names. If the interpretation above is not what you were expecting, you can [try again with the rare cases included](#).

External web pages (using standard web search):

[1900 - Wikipedia, the free encyclopedia](#)

1900 From Wikipedia, the free encyclopedia Jump to: navigation , ... 1930s Years : 1897 1898 1899 - 1900 - 1901 1902 1903 1900 by topic <http://en.wikipedia.org/wiki/1900>

[DEMOGRAPHIA: Demographics Development Impacts Market Research & Urban ...](#)

... Louis St. Petersburg Stockholm Sydney Tianjin Tokyo-Yokohama Toronto Vancouver ... USA Metro Areas: USA Metros 1900: USA Smart Growth: Brief USA Cities 2000 USA ... <http://www.demographia.com/>

What you can do

First, you can **tell the system about the existence of an entity** that it doesn't already know about (note that we don't currently support the addition of fictional entities) registering it with the system:

add a new person

add a new place

add a new business

add something else

(Don't worry too much about whether the system already knows about the thing you wish to add. If it does, you should discover this shortly into the process and no harm will be done.)

Secondly, you can **tell the system a new fact**. If the things mentioned in the fact are unknown to the system you can add them as part of this process:

add a new fact

Advanced users

People who are already familiar with the technology can add knowledge by selecting one of these choices:

add a new relation

add a new class

add a new object

Issues Affecting Performance

- Hard to break the 2.3-word-query habit:
 - In our four year 'Just Ask!' experiment, less than 10% of invited questions were actually questions
- 'Deictic' and other context-specific questions:
 - are my phys149 results available?
 - are practicals on this week?

Other Related Areas Worth Watching

- Machine Translation
- Audio Search and Audio Mining

Where We're At

- Search provides 80% of the value of text analytics – but this decreasing as document datasets get bigger and bigger
- Text categorisation provides value if implemented appropriately
- Text clustering can help if 100% accuracy is not required
- Text summarisation needs to be tailored to document sets
- Named entity recognition already improves on basic search
- Event recognition is hampered by low accuracy
- Sentiment analysis is this year's hot topic

What's Coming

- Fact extraction and aggregation: 'open' information extraction
- Cross-document entity tracking
- Multi-document summarisation

Finding Out More: Industry

- Check out the companies mentioned in this presentation
- Go to the Text Analytics Summit
- Type 'text analytics' into any search engine ...
- Read my 'Industry Watch' column
 - <http://www.ics.mq.edu.au/~rdale/publications/industrywatch/>

Finding Out More: Research

- Conferences:
 - The Association for Computational Linguistics
 - Empirical Methods in Natural Language Processing
 - COLING
- Journals:
 - Computational Linguistics
 - Natural Language Engineering
- The ACL Anthology: <http://aclweb.org/anthology-new/>

Finding Out More: Do a PhD!

- Current topics being pursued at the Centre for Language Technology include:
 - Extracting tabular information from documents
 - Identifying requests and commitments from emails
 - Tracking entities across documents
 - Automatic agenda construction from UN documents
- High-value PhD scholarships available via the Capital Markets Co-operative Research Centre

Questions?