



Content analytics

Tapping unstructured content for
business advantage

Contents	
3	Business problems Unstructured content is exploding Companies must harness unstructured content
4	Business drivers The changing nature of analytics Proven technologies to handle burgeoning content An essential tool for any business The move to business-driven decision-making
8	The solution IBM Cognos Content Analytics Trustworthy insights, all-inclusive technology
11	Conclusion

Abstract

Content analytics is an emerging field of analytics that enables companies to gain new business insights from “unstructured” and “semi-structured” content. This content can include forms, documents, comment fields in databases, Web pages, customer correspondence and other information that is not stored in structured data fields.

Content analytics makes it possible to access, analyze and report on unstructured and semi-structured content both inside and outside an organization. It is a natural extension of business intelligence, which uses structured data for measurement, analysis and reporting.

IBM Cognos® Content Analytics gives organizations the ability to harness unstructured content to gain greater business insights for better decision-making. With IBM Cognos Content Analytics, companies can move from data-driven decision-making to business-driven decision-making for enhanced business insights that lead to competitive advantage in an ever-changing marketplace.

Overview

Business leaders are always seeking new business insights to drive their organizations forward. They want to identify the root causes of problems, spot opportunities more quickly and anticipate and exploit future market trends. To gain this insight, they turn to business intelligence solutions to track key metrics in sales, marketing and operations. This insight relies on structured, stored data that is internal to the organization.

But, that’s not the whole picture. Across the enterprise and on the Web, the market speaks volumes. Customers, prospects and influencers write comments, fill out forms and talk to sales, marketing and customer service personnel. This information should be factored into decision-making. Instinct and anecdotal information also contribute to good decision-making. However, this sort of information is hard to track and analyze objectively. It often resides in unstructured and semi-structured content such as documents, written notes, Web pages and more.

Smart companies recognize the untapped business value in unstructured content. For example, companies can use this content to answer questions such as:

- Do call center inquiries signal a risk of customer attrition?
- Are there early warning signs of fraud in our insurance claims?
- Are late deliveries costing us sales?

Companies need a systematic way to find, distill and analyze the massive amount of unstructured content that exists today. Content analytics solutions provide the answer.

This paper addresses the need for content analytics in an organization. It discusses:

- The need to tap unstructured content for business insights
- The reasons why content analytics solutions should be implemented now
- The IBM Cognos Content Analytics solution and its ability to discover, refine, visualize and deliver new insights through the analysis of unstructured content.

Business problems

Unstructured content is exploding

Unstructured content can be found everywhere. It is in the text of e-mails, documents, blogs, Web sites and handwritten notes. It is contained in enterprise systems and in cloud-based networks, inside or outside an organization. It is estimated that as much as 80 percent of information generated daily is unstructured. Some of this information contains potentially relevant business information.

This semi-structured or completely free-form information often escapes careful scrutiny. It is difficult to corral and utilize because it is not organized. It cannot be queried like a database. Yet this information contains critical business information that can be used to identify customer preferences, support concerns, sales issues, fraudulent claims, warranty activity, supplier feedback and other crucial information that can help an organization be more successful.

Companies must harness unstructured content

Companies put themselves at a disadvantage if they ignore the growing repository of unstructured content and the business insight it contains. Yet given the time and cost of reviewing, filtering and interpreting this information, it is cost prohibitive for most companies to analyze this data on their own. Even if they do, the resulting information is often subjective, anecdotal and difficult to track.

For companies to harness unstructured content in a systematic, cost-effective way, a content analytics solution is required.

Business drivers

The changing nature of analytics

Content analytics is an emerging field of analytics that enables companies to unlock the insights contained in unstructured content. Content analytics offers the ability to access, sort and analyze unstructured content and then combine it with structured data and other existing information resources and applications for reporting and analysis.

Content analytics is a natural extension of business intelligence. Many companies already use business intelligence for “data driven decision-making.” This is a decision-making process based on insights gleaned from records of past transactions and other structured information typically housed in data warehouses.

Companies can supplement these business intelligence methods with content analytics techniques, which can be used to expose trends within unstructured content. Content analytics has particular strengths for tracking shifting market information in near real-time, such as a sudden rise in warranty activity or customer complaints. It gives companies the ability to gather and analyze critical information that moves organically with the marketplace and offers a powerful extension of traditional business intelligence techniques to improve decision-making.

Proven technologies to handle burgeoning content

Content analytics capabilities are based on proven technologies that have been deployed for years. These technologies can handle the rapidly expanding amount of unstructured content that has resulted from the widespread use of the Internet, collaboration platforms, enterprise networks, cloud computing and other information technologies.

Several key technology components are used to power a fully featured content analytics solution. These components include:

- **Crawler, search and indexing technologies** to identify and tag relevant content in a wide variety of unstructured or semi-structured information sources
- **Text analytics technologies** to sort, filter and categorize content to distill the most relevant information and to automatically highlight trends and anomalies
- **Exploration and visualization technologies** for graphic display of content correlations, frequencies, deviations and trends in graphs and charts for fast consumption
- **Open integration technologies** to supply these insights to other applications and systems for more encompassing analysis and reporting.

When combined, these technologies offer the ability to identify, sort, filter, organize, visualize and deliver the critical insights found only in unstructured content.

An essential tool for any business

With the amount of unstructured content exploding, it has become a business necessity to invest in a content analytics system and use unstructured content for competitive advantage. For example, imagine if an insurance company could use content analytics to identify fraudulent claims and reduce them by 10 percent annually? A substantial return on investment can be achieved by companies who use content analytics to monitor and act on targeted areas such as customer complaints, warranty issues and fraud claims. A content analytics solution can manage the pervasive use of unstructured content and positively impact a company's bottom line.

Content analytics can be used in a wide range of industries and in numerous business situations. For example, it can be used to:

- Measure and monitor customer service metrics based on analysis of text in call center records
- Plan product release priorities based on warranty claims and product performance
- Develop a winning competitive selling strategy based on analysis of text in competitor filings and win/loss data
- Predict customer attrition or fraud based on contents of claims forms.

Virtually any kind of organization, from automotive to insurance to law enforcement, can benefit from the insights that content analytics provides.

The move to business-driven decision-making

Many companies rely on business intelligence software to derive key data-driven insights from operational systems and data warehouses. Good managers supplement these insights with anecdotal information about external market variables by reading blogs, perusing product reviews and talking to customers and field employees. However, this approach is time-consuming and subjective.

With content analytics, companies have a powerful, objective means of culling business insights from massive amounts of unstructured content. This enables an organization to move from data-driven decision-making to “business-driven decision-making,” which considers both internal and external variables for optimal decision-making.

Consider the example of a hotel chain that is making a decision about investing in better room furnishings. Using data-driven evidence, the company has solid information about the costs and sales trends associated with various grades of furnishings. By adding external market insight, the company can gain a more comprehensive view of the hotel marketplace, including customer preferences and perceptions. By combining data-driven evidence with external market insight, the hotel chain can make a more fully informed business-driven decision about whether to upgrade its furnishings.

Data-driven evidence	+	External market insight	=	Business-driven decision-making
Found in operational systems and data warehouses.		Found in unstructured content, such as blogs, notes in databases and Web sites.		Enables decision-making fully informed by cost, value and market perception.
Includes: <ul style="list-style-type: none"> • Cost to upgrade furnishings • Comparative analysis of sales trends by property • Percent of repeat business from customers • Customer survey results. 		Includes: <ul style="list-style-type: none"> • Customer perception of hotel properties • Customer perception of competitive lodging • Relative importance of customer lodging based on customer recommendation. 		Drives better decision-making: <ul style="list-style-type: none"> • The cost to upgrade the facilities is determined and compared to the value associated with edging out the competition, increasing brand equity and driving new/repeat business.

The solution

IBM Cognos Content Analytics

IBM Cognos Content Analytics is a complete solution that provides organizations with all the necessary tools to unlock the business value contained within unstructured content. Working through a dynamic, highly visual interface, this solution discovers important information by identifying and surfacing correlations, enabling organizations to discover hidden trends, augment business intelligence reports or enhance business processes with unstructured content.

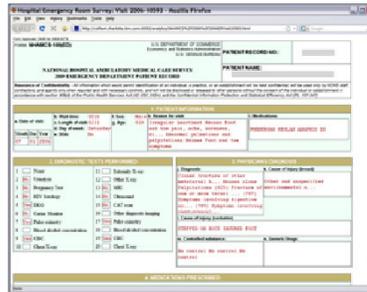
IBM Cognos Content Analytics can help companies discover, refine, visualize and deliver new business insights. It provides capabilities to:

Discover: Automatically identify and tag key attributes and entities within content by crawling almost any content source and identifying key words and phrases. Content is identified and tagged across a wide variety of content stores using sophisticated crawler technology for comprehensive information access.

Refine: Empower navigation and drill-down based on identified key attributes, entities and extracted dimensions. Crucial information is extracted by applying a series of “annotators” that identify the key concepts in a piece of text as well as the relationships among those concepts to present content relevant to data analysts and business users.

Visualize: Highlight deviations and anomalies through advanced visualization and enable exploratory mining for more informed business decisions and action. A variety of visualizations are provided, including time series views, facet views, deviation views and trends views through an easy-to-use interface.

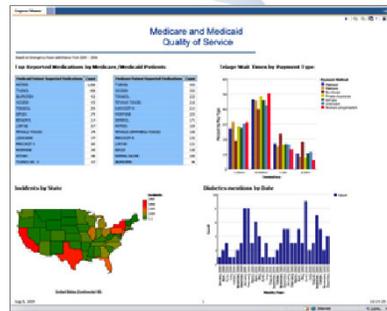
Deliver: Support for broad delivery of information to other processes and applications using open standards. These include ECM (Enterprise Content Management) repositories and CRM (Customer Relationship Management) applications, as well as IBM Cognos 8 BI for additional reporting and analysis.



Source Content
(ex: supplier invoice)



Interactive Discovery
(ex: trend & outlier analysis)



Operational Reporting
(ex: BI dashboard including content metrics)



An analysis of production records provides insights for business optimization.

IBM Cognos Content Analytics takes the subjectivity out of the analysis of unstructured content to deliver consistent, complete and cost-effective insights. It is scalable to hundreds or even thousands of Web sites, documents, database notes and other information sources. It can analyze unstructured content automatically, so a company stays up to date on current developments that affect its business strategies.

Trustworthy insights, all-inclusive technology

IBM Cognos Content Analytics is based on advanced technologies that IBM has been researching, developing and deploying for decades. This single all-encompassing content analytics solution uses sophisticated algorithms to analyze unstructured content in almost any format, including both text and data. It also works with semi-structured content such as XML documents, HTML pages, document metadata and databases with long text fields.

IBM Cognos Content Analytics can be integrated with existing processes and applications to extend their reach for more wide-ranging analysis and reporting. IBM Cognos Content Analytics can be immediately integrated with IBM Cognos 8 BI, InfoSphere Warehouse and IBM ECM systems. It can also be easily integrated with custom-built solutions to act as an ETL (extract, transform, load) solution for converting unstructured content into structured data. It uses open industry standards, such as Lucene and UIMA (Unstructured Information Management Architecture) to ensure the widest possible adoption.

Conclusion

By extending existing business intelligence functions to include content analytics, companies can move from data-driven decision-making to business-driven decision-making. They look beyond internal metrics and data to inform their decision-making with external information that enables qualitative analysis. These companies support both the art and science of decision-making.

IBM Cognos Content Analytics is a complete solution that provides organizations with the necessary tools to unlock the business value contained within unstructured content. It enables organizations to:

- Access virtually any type of structured, semi-structured or unstructured content found within the enterprise
- Explore and analyze content with a dynamic, highly visual interface that integrates easily with IBM Cognos 8 BI for additional reporting, analysis and information delivery
- Spot issues and important trends that may fall outside of the normal business reporting channels
- Deliver new insights to business users across the organization and help them turn insight into action with confidence.



© Copyright IBM Corporation 2009

IBM Canada
3755 Riverside Drive
Ottawa, ON, Canada K1G 4K9

Produced in Canada
December 2009
All Rights Reserved.

About IBM Cognos BI and Performance Management

IBM Cognos business intelligence (BI) and performance management solutions deliver world-leading enterprise planning, consolidation and BI software, support and services to help companies plan, understand and manage financial and operational performance. IBM Cognos solutions bring together technology, analytical applications, best practices, and a broad network of partners to give customers an open, adaptive and complete performance solution. Over 23,000 customers in more than 135 countries around the world choose IBM Cognos solutions.

For further information or to reach a representative: www.ibm.com/cognos

Request a call

To request a call or to ask a question, go to www.ibm.com/cognos/contactus. An IBM Cognos representative will respond to your enquiry within two business days.

IBM, the IBM logo and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at www.ibm.com/legal/copytrade.shtml.

References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates.

Any reference in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.