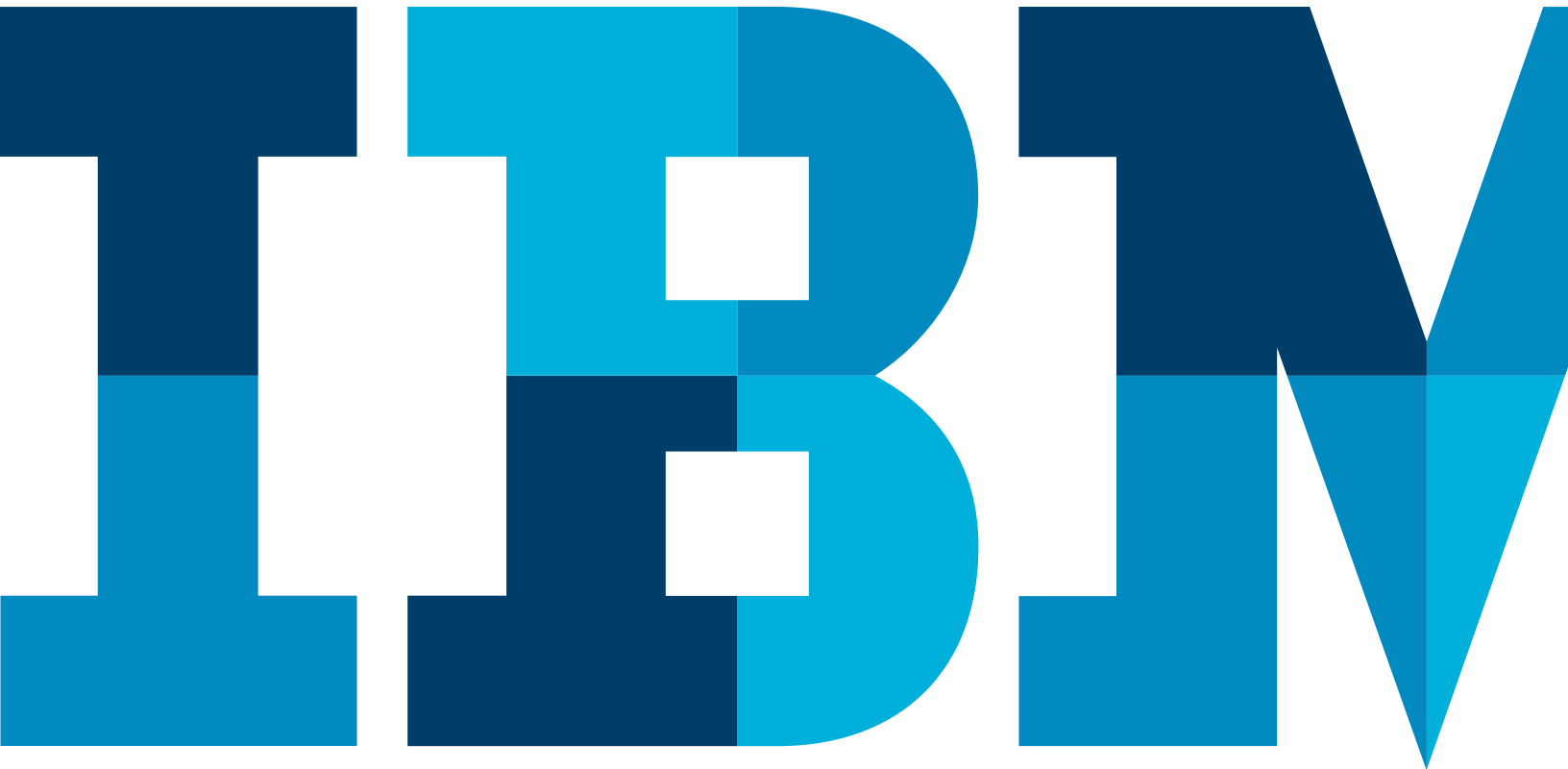


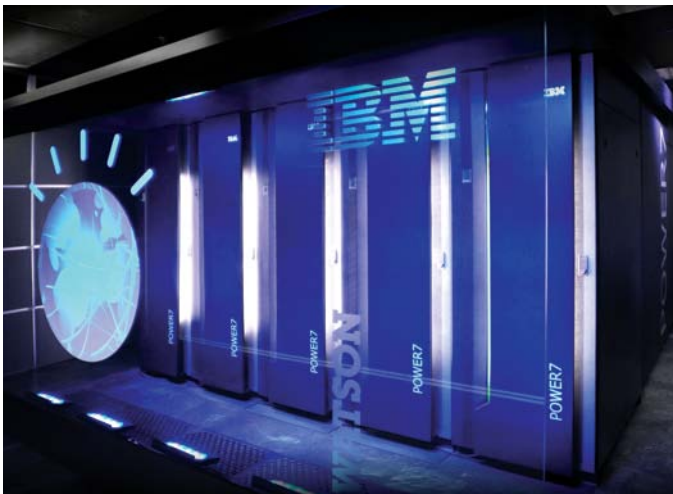
# IBM Content Analytics software powers Watson to a Jeopardy! win

*What can it do for your business?*



*“The technology behind Watson represents the future of data management and analytics. In the real world, this technology will help us uncover insights in everything from traffic to healthcare.”*

— John Cohn, IBM Fellow, IBM Systems and Technology Group



## What is Watson?

IBM Watson DeepQA is the world's most advanced question-answering machine, which uncovers answers by understanding the meaning buried in the context of a natural language question. By combining advanced natural language processing (NLP) and DeepQA automatic question-answering technology, Watson represents the future of analytics, systems design, and content and data management. Watson leverages core content analysis, along with a number of other advanced technologies, to arrive at a single, precise answer within a very short period of time. The business applications for this technology are limitless, from clinical health-care and customer care to government intelligence and beyond. For more information about Watson, visit [ibmwatson.com](http://ibmwatson.com).

## How does Watson work?

Watson works like the human brain to analyze the content of a Jeopardy! question. First, it tries to understand the question to determine what is being asked. In doing so, it first needs to analyze the natural language text. Next, it tries to find reasoned answers by analyzing a wide variety of disparate content, mostly in the form of natural language documents. Finally, Watson assesses and determines the relative likelihood that the answers found are correct, based on a confidence rating.

A great example of the challenge is described by Stephen Baker in his book *Final Jeopardy: Man vs. Machine and the Quest to Know Everything*. “When 60 Minutes premiered, this man was U.S. President.’ Traditionally it’s been difficult for a computer to understand what ‘premiered’ means and that it’s associated with a date. To a computer, ‘premiere’ could also mean ‘premier.’ Is the question about a person’s title or a production opening? Then it has to figure out the date when an entity called ‘60 Minutes’ premiered and then find out who was the ‘U.S. President’ at that time.” In short, it requires a ton of contextual understanding.

## Why does this matter to your business?

Jeopardy! poses a set of contextual information challenges similar to those found in the business world today:

- Over 80 percent of information being stored is unstructured (text based).
- Understanding that *over 80 percent* isn’t simple. Like Jeopardy!, subtle meaning, irony, riddles, acronyms, abbreviations and other complexities present unique computing challenges not found with structured data, in order to derive meaning and insight. This is where NLP comes in.

The same core NLP technology used in Watson is available now to deliver business value. Watson uses the same capabilities found in IBM Content Analytics software to unlock the value embedded in the massive amounts of unstructured information in the many systems and formats you have today. Understanding the content, context and value of this unstructured information presents an enormous opportunity for your business. By exploiting the technology within Watson, IBM software solutions bring the power of Watson to your business.

IBM Content Analytics software will help your business rapidly gain valuable insight, which is currently trapped in unstructured information:

- Aggregating and extracting content from multiple internal and external sources and types, including enterprise content management (ECM) repositories, structured data, social media, call center logs, research reports, transcripts, email, safety reports and legal contracts
- Organizing, analyzing and visualizing enterprise content (and data) using NLP and other analytics to understand meaning and identify trends, patterns, correlations, anomalies and business context
- Interactively searching and exploring to derive rapid insight by confirming what is suspected or by uncovering something new—all without building models or deploying complex systems

## Case studies

The following organizations are gaining valuable insight from their content today.

**Healthcare research**—Like most healthcare providers, BJC Healthcare had a treasure trove of historical information trapped in unstructured clinical notes and diagnostic reports that contained information essential for the study of disease progression, treatment effectiveness and long-term outcomes. Its existing biomedical informatics resources were disjointed, not interoperable, available to only a small fraction of researchers and frequently redundant, with no capability to tap into the wealth of research information trapped in unstructured clinical notes, diagnostic reports and the like.

With IBM Content Analytics software, BJC and university researchers are now able to analyze unstructured information to answer key questions that were previously unavailable. Questions such as the following: *Does the patient smoke? If so, then how often and for how long? If smoke free, then for how long? What home medications is the patient taking? What was the patient's diagnosis? What procedures were performed on the patient?* BJC now has deeper insight into medical information and can uncover trends and patterns within the content, providing better healthcare to its patients

**Customer satisfaction**—Identifying customer satisfaction trends about products, services and personnel is critical to most businesses. The Hertz Corporation and Mindshare Technologies, a leading provider of enterprise feedback solutions, are using IBM Content Analytics software to examine customer survey data, including text messages, to better identify car and equipment rental performance levels for pinpointing customer satisfaction levels and making the necessary adjustments to improve them. By using IBM Content Analytics software, companies like Hertz can drive new marketing campaigns or modify their products and services to meet the demands of their customers.

---

*“Hertz gathers an amazing amount of customer insight daily, including thousands of comments from web surveys, emails and text messages. We wanted to leverage this insight at both the strategic level and the local level to drive operational improvements.”*

— Joe Eckroth, chief information officer, The Hertz Corporation  
([ibm.com/press/us/en/pressrelease/32859.wss](http://ibm.com/press/us/en/pressrelease/32859.wss))

---

**Research analytics**—To North Carolina State University, the essence of a university is more than education—it is the advancement and dissemination of knowledge in all its forms. One of the main issues faced by North Carolina State University was dealing with the vast number of data sources available to it. The university sought a solution to efficiently mine and analyze vast quantities of data to better identify companies that could bring its research to the public. The objective was a solution designed to parse the content of thousands of unstructured information sources, perform data and text analytics, and produce a focused set of useful results.

Using IBM Content Analytics software, the university was able to reduce the time needed to find target companies from months to days. The result is the identification of new commercialization opportunities, with tests yielding an approximately 300 percent increase in the number of candidates. By obtaining insight into

North Carolina State University's extensive content sources, its Office of Technology Transfer was able to find more-effective ways to license technologies created through research conducted at the university.



---

*“What makes the solution so powerful is its ability to go beyond conventional online search methods by factoring context into its results.”*

— Billy Houghteling, executive director, Office of Technology Transfer,  
North Carolina State University ([ibm.com/software/success/cssdb.nsf/CS/SSAO-8DFL BX?OpenDocument&Site=software&cty=en\\_us](http://ibm.com/software/success/cssdb.nsf/CS/SSAO-8DFL BX?OpenDocument&Site=software&cty=en_us))

---

## About IBM enterprise content management

IBM enterprise content management software enables the world's top companies to make better decisions, faster. By gaining control of unstructured information, companies can access, collaborate and influence business decisions in new ways, making content a first-class source of insight. With industry-specific IBM ECM solutions, companies can capture, manage and share content throughout its lifecycle, helping ensure compliance, reduce costs and maximize productivity. The IBM ECM portfolio includes a wide array of capabilities that integrate with existing systems to help organizations maximize the value of information, including document capture and imaging; social content management; advanced case management; information lifecycle governance; and content analytics. More than 13,000 global companies, organizations and government organizations rely on IBM ECM solutions to improve performance and remain competitive through innovation.

Put the technology of tomorrow to work for you today by leveraging the same IBM Content Analytics capabilities that help power Watson. To learn more about all IBM ECM products utilizing Watson technology, visit these sites:

IBM Content Analytics software:  
[ibm.com/software/data/content-management/analytics](http://ibm.com/software/data/content-management/analytics)

IBM Classification Module software:  
[ibm.com/software/data/content-management/classification](http://ibm.com/software/data/content-management/classification)

IBM eDiscovery Analyzer software:  
[ibm.com/software/data/content-management/products/ediscovery-analyzer](http://ibm.com/software/data/content-management/products/ediscovery-analyzer)

IBM OmniFind® Enterprise Edition software:  
[ibm.com/software/data/enterprise-search/omnifind-enterprise](http://ibm.com/software/data/enterprise-search/omnifind-enterprise)

---

© Copyright IBM Corporation 2011

IBM Corporation  
Software Group  
3565 Harbor Boulevard  
Costa Mesa, CA 92626-1420  
U.S.A.

Produced in the United States of America  
April 2011  
All Rights Reserved

IBM, the IBM logo, and [ibm.com](http://ibm.com) are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at [ibm.com/legal/copytrade.shtml](http://ibm.com/legal/copytrade.shtml)

JEOPARDY! (c) 2011 Jeopardy Productions, Inc. JEOPARDY! is a registered trademark of Jeopardy Productions, Inc. All rights reserved.

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. Client success stories are available at [ibm.com/software/success/cssdb.nsf](http://ibm.com/software/success/cssdb.nsf)

The information contained in this documentation is provided for informational purposes only. While efforts were made to verify the completeness and accuracy of the information contained in this documentation, it is provided “as is” without warranty of any kind, express or implied. In addition, this information is based on IBM's current product plans and strategy, which are subject to change by IBM without notice. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, this documentation or any other documentation. Nothing contained in this documentation is intended to, nor shall have the effect of, creating any warranties or representations from IBM (or its suppliers or licensors), or altering the terms and conditions of the applicable license agreement governing the use of IBM software.

IBM customers are responsible for ensuring their own compliance with legal requirements. It is the customer's sole responsibility to obtain advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulatory requirements that may affect the customer's business and any actions the customer may need to take to comply with such laws.



Please Recycle