

Moving From Enterprise Search to Cognitive Exploration



Executive Summary

The good news is that enterprise search has been around for decades and usually does an adequate job. The bad news is that the adequate job only provides lists of search results and a way to navigate and review the results. Finding documents that contain certain keywords? *Check*. Searching relational databases? *Check*. Coordinating search results from multiple internal data stores? *Check*. Searching content outside the company, such as premium information services? *Check*. Providing deep content analytics with mature supporting tools? Providing unified information applications, and collaboration? Providing advanced cognitive capabilities? Those capabilities remain out of reach for the majority of players in the enterprise search arena.

With Watson Explorer, you can keep enterprise search as the foundation and transform search into Cognitive Exploration. Leveraging technological advances such as deep search and exploration, advanced content analytics, and cognitive capabilities, IBM Watson Explorer provides a unified view of the information you need, combining data from multiple internal silos and a variety of outside datasets including social media. Stop limiting your search to traditional data sources in the new, non-traditional data world.

You Have Search. What You Need Are Solutions.

Enterprise search, a mature product category, is a standard part of the IT infrastructure in most large organizations. Market maturity means slowed innovation, few new vendors, and a definite lack of excitement among enterprise users. Part of the problem stems from the traditional enterprise search tools inadequately addressing the realities of the modern data world and the “consumerization” of search.

Data Growth Outpaces Intelligent Data Management

Data under IT’s direction used to be well-behaved. Contents of various systems came primarily from company operations only, not the outside world. Each company created its own data so stored information grew about as fast as the company grew.

Today, worldwide corporate data now doubles every 14 months. Analysts expect the total data created and copied to reach 44 ZB by the year 2020 (Analyst firm IDC)¹. After all, there are more than 204,000,000 emails launched every minute every day (Mashable.com)². How do you manage, search, and process that data and turn it into usable information?

Mountains of Unstructured Data Barely Managed

People rarely keep their data neatly organized. Spreadsheet information may be logical inside their rows and columns, but the data is not organized, nor are the files indexed. Text documents have no regular internal structure, and the same can be said for presentations. While emails are organized and indexed and sorted by users, their contents are unstructured and they often have text, presentation, or spreadsheet files attached. Of the estimated 80 Exabyte of data storage needed in 2014, nearly 90 percent, will be unstructured (RobertPrimmer.com)³.

Data Science Central calls this the “Rise of Data Anarchy.”⁴ Using the definition that any data not in a database or other formalized data structure is unstructured, count the types of data in your business. If your company follows the trend, four of five data sources inside the company will be unstructured.

Search developed into the workhorse it is today because it is the only way to work effectively with unstructured data sources. Unfortunately, searching may categorize unstructured data and provide organized results from a chaotic data set, but few systems provide intelligent analysis.

Data Silos Inside and Outside The Company

Multiple information silos hold data, often for exclusive use of various Line of Business groups, sometimes with the help of IT and sometimes despite IT. Mergers and acquisitions place incompatible systems together in a mashed-together IT fabric with many storage devices but few common denominators. Important data remains hard to find and nearly impossible to consolidate into usable formats.

Outside data will be grouped into virtual piles of unstructured data sources. Social media data is external, unstructured, and siloed by application. You can search and maybe find, but that provides no real organization.

Improved Search Is Not Enough

Every search vendor slaps an adjective of some type on their search product or optional module to entice upgrades and lure new customers. While indexing content for faster retrieval helps, the limitations of inadequate search tools remain even if results appear on screen ten times faster. Quicker search results may just give you incomplete answers in less time than the previous version.

Employees Demand Consumer Level Option

Employees today have high expectations as a result of experiences on the Web and by using consumer devices. A consumer may type a query in a Web search engine to find out what time the local pharmacy closes, where the closest fast food restaurant is, where a hit movie is playing, etc. They can ask their phone where the closest pizza restaurant is. Combining tagging for site relevance and a simple question and answer interface helps users feel empowered when looking for information. They get these tools free with their consumer devices. When IT offers no corresponding tools for enterprise search, users feel the company lacks modern knowledge management tools. Users may not understand the difference between consumer Web searching leveraging popularity ranking, geographic location and search engine optimization and industrial-strength enterprise search. Yet, they rely on these consumer tools and demand the same level of performance at the office.

Needed: A Unified View of Content Inside and Outside the Company

Simple consumer question and answer programs cannot address your employees' need to access information in your internal systems and combine

it in context with relevant data from outside your organization. To meet today's expectations, a search system must be able to access all of your important data sources and filter results based on a user's access permissions within the organization. Search must also include all data useful to the company that is outside its four walls, including password-protected premium information services. The more context the search program provides, the more useful the search results.

Unfortunately, many companies find searching external content easier than efficiently accessing internal content. Who owns the Customer Relationship Management (CRM) system and controls those data files? Sales? Does a different group besides IT own the ERP system, like finance? How does Sales get CRM data to feed the ERP sales pipeline reports? Does the ERP group know where research and development project files are stored?

Storage device sales top \$25 billion every year. Vendors actively target Line of Business executives on down to the department level. Does your IT department manage those scattered storage pools? Do they even know they exist?

Beyond Searching: Cognitive Exploration

Once companies realize indexing, tagging, and keywords fail to provide the information they need to make informed decisions, they search for the next step. The answer they seek? Cognitive exploration.

Far more than just a fancy word for searching, cognitive exploration encompasses a range of technologies and abilities far beyond the smartest search application. Cognitive exploration is the combination of search, content analytics, and cognitive computing. Not only can cognitive exploration accelerate the rate at which users can find and navigate information; by leveraging advanced technologies such as content analytics, machine learning, and reasoning it has the potential to augment human expertise.

Provide Information in Context

Search engines retrieve documents, Web pages and other objects based on keywords entered by a user, but they do not assist the user in actually understanding the presented information. Cognitive exploration systems have the ability to leverage a much deeper understanding of unstructured content to improve the

relevance of information returned to the user and to apply analytics that help users to grasp and act on the information.

Deeper Analysis of Unstructured Content for Trends and Insights

Taking the next step, a the new system that goes beyond searching must provide deep natural language processing. Capabilities such as entity and concept extraction can be used to improve exploration, as well as to implement content mining to expose trends and patterns in unstructured content to gain new insights.

Three Pillars of Cognitive Exploration

To improve your business results from enterprise search and step up to cognitive exploration, three features are critical. You must improve search beyond indices, tags and “smart” labels, analyze the content found, and apply cognitive computing to help users to more effectively leverage content. More importantly, you must be confident that users receive a complete view of all the information they need, presented in the proper context to make better decisions.

Let us be clear that cognitive systems include far more than simply better search. The term “cognitive” explicitly draws a parallel to how humans think. Cognitive systems mimic human cognitive processes by understanding human language, learning and forming hypotheses. As a result, they can be leveraged to extract and evaluate useful information from unstructured content such as dates, locations, emotional states and values and, in some cases, use that information to answer questions.

Search / Exploration

Searching can mean looking for a specific item, such as a record from a database, or for available information about a topic. The documents, Web pages or passages on a topic may be found internally or outside your organization. Even Web searches cover the majority of all available information on the surface Web, although penetration into the so-called “Dark Web” (private and sometimes illegal content) remains limited.

On the inside of your company, an enterprise search system should leverage all the available data sources and include them in a search when appropriate. A good cognitive system will then guide users with an intuitive interface. On the outside, a wealth of information about

your company resides on various social media sites, particularly sites with apps for smartphones. Most social media sites remain out of reach of the major Web search engines, leaving that information invisible to normal search. Enterprise search systems ignore social media because they are outside the company. Therefore, thousands or millions of data points about your business and products remain hidden.

Advanced Content Analytics

Enterprise search, business intelligence software, and enterprise resource planning systems actively work to improve analysis for the managers using them. These systems work with data warehouses and transactional systems, and many companies have successfully put these tools to good use. But the resulting information provided focuses on “what” happened, not why.

Very often, however, the answer to why something happened is hidden in human language--customer comments, operator’s notes, e-mail messages and a host of other unstructured content types. Advanced context analytics must be employed to analyze this content and push beyond the what, where, and when results to begin to answer why things happened and even predict future trends and patterns.

Most text analysis systems require technical experts to adapt the system to a particular organization's needs. Can your business problems wait months for answers? Probably not. Then look for a solution that enables business users and subject matter experts to tune the system without writing a single line of code. Advanced content analytics need to work with unstructured data from documents, emails, social media postings and even call center transcripts. All of these sources resist analysis by traditional analytics, reporting tools, and content analytics.

Cognitive Capabilities

Systems that read, understand, and analyze human language and augment human cognition now become critical for the modern enterprise. Questions like, "What is cognition," lead to many different discussions. Most can agree, however, that it includes the ability to understand language to learn and to form responses or hypotheses. Some critical pieces of this type of solution include:

The ability to understand natural language so the system can read and understand documents and communicate with humans. The capability to delve deep into specialized knowledge fields, like healthcare

or financial services, provides great value for vertical market applications.

The ability to adapt and learn. Data, particularly in dynamic forms from social media or Internet news feeds, changes constantly. Delivering insight to data captured last week provides far less value than insight on data analyzed today.

The capability to form hypotheses, which is the first step in offering answers that come with a confidence level, like you would get from a coworker.

For instance, what does the word "run" mean? Is it a baseball score? An exercise? Manufacturing a batch of product? Language imprecision confounds at every turn.

Documents full of baseball runs only annoy users looking for production information. No matter how smart search gets, the inability to distinguish different uses for the same words will result in useless returns.

Cognitive capabilities in your systems can have a dramatic effect on results. Intelligent solutions go beyond looking for search words in close proximity. They analyze the intention of the statements much like

a person does. Reading “that mess was incredible,” a cognitive system will not interpret the mess as wonderful or amazing. Effective natural language processing tools will read other statements concerning that mess and classify that statement as negative rather than praising that mess. Context matters.

All Presented in a Unified View

Search results organize data into an easily digestible format. But today you need more. Today your system should be a unified information application that gathers and displays data, analytics and cognitive insights on a particular topic in one place, in context and dynamically, in a way that best serves the needs of the user.

Finding the information most useful to these pursuits today means utilizing social media, the messiest of unstructured data. Discovering meaning for customer service may mean cross-referencing product mentions between different social media applications, your inventory, and your supply chain. Perhaps first call resolution rates would jump if agents were able to notice product mentions on social media. Then they could cross-reference call center complaints about failure symptoms, and tie those comments to

a particular vendor. Solving the problem before it becomes a problem is possible when you leverage cognitive capabilities across all the data sources relevant to your company.

Synergy between Search and Cognitive Computing

Professionals across a number of job functions can benefit from the combined effect of contextual search and cognitive computing.” For example, marketing professionals can search an external system such as public messaging apps for mentions of a potential customer and receive links to contacts working for that company.

Once the search results are returned, the system uses natural language processing to identify emotional words, such as broken, perfect, difficult, and so on. If your cognitive computing systems include insights into the personality of the speaker, your next action can take into account this deeper information about the people who wrote the messages.

Additionally, an M&A firm may constantly search news feeds for opportunities to partner compatible companies. Analyzing complex financial data with a

cognitive system versed in the target industry, such as healthcare equipment manufacturers, will make connections beyond the information on the Web news outlets.

Example 2: Relationship Extraction

Traditional entity extraction systems analyze unstructured content to identify people, places, organizations, and other nouns. This is extremely useful, and the results of entity extraction can be used to enhance search and exploration as well as to feed information to other analytic systems. The Watson Developer Cloud Relationship Extraction service takes this a step further by not only identifying entities, but also by determining the relationships among those entities, and the verbs and other nouns in the content.

For example, a relationship extraction tool could take unstructured news files and find connections between nouns, verbs, subject, and objects. Company names, number of items, dates, and people (names as well as nouns like accountant or police) would be studied for non-obvious relationships. Perhaps a description of a police detective's day would mention visiting a local bank where an assistant vice president was mentioned as being indicted for embezzling funds. The interesting relationship might be that the detective arrested the bank officer.

Example 3: Personality Insights

Sales and marketing departments want to know as much about prospects as possible. Reading customer text-based social media and viewing images provide plenty of insight, but doing so manually is inefficient and impossible to scale.

If your search and cognitive exploration software understands language and provides the necessary analytics modules in an automated fashion, your sales and marketing departments will become quite excited. The right software can read the unstructured textual data from a mass of prospects and learn their personality characteristics using multiple methods: the Big Five (Costa and Norman), needs such as excitement, harmony, and curiosity, and values such as helping others, conservation, and openness to change. The resulting output might identify good candidates for a follow-up marketing email.

Example 4: AI for the Visual World

The term “unstructured data” appears many times in this document. What may not be obvious is that photos and other images make up a substantial amount of that unstructured data.

In 2013, one popular social media site reported that 350 million photos were uploaded to that service every single day. By now, that number may well be 400 million photos per day. Another popular social media site claims 70 million photos are uploaded daily. Thanks primarily to smartphones, people take and share more photos each day than ever before. The Internet contains a vast number of images representing products and other useful information. Clearly there is a wealth of information available that is expressed in images rather than text. And yet most systems are unable to use this information effectively. If you sold sunglasses, would you like a list of social media users who appeared in photos wearing a style of sunglasses like you sell? Repeat business could boom, especially if those wearers were enticed with an email about a newer style of sunglasses that are now available.

Imagine the wealth of information that could be generated with a cognitive system that can accurately tag images with information such as identifying specific objects represented in pictures, recognizing individuals, or providing age range and gender from photos in unstructured data streams.

Consider IBM Watson Explorer

The First Cognitive Exploration and Analytics Platform

IBM Watson Explorer is a cognitive exploration solution that combines content analytics and cognitive computing with search to help users learn from hidden connections within their data. It presents all of these capabilities in unified information applications tailored to the user's role and current activities. You have search. What you need are solutions, and IBM Watson Explorer provides those.

Content analytics tools in Watson Explorer provide meaning and insight from natural language content, both structured and unstructured. From online customer comments to research reports, Watson Explorer can analyze information in ways traditional enterprise search and BI never could. Your analysis results may help you better understand customers, identify previously invisible relationships, and make better decisions.

The cognitive capabilities that can be leveraged by Watson Explorer are provided by the IBM Watson platform. IBM Watson came to public attention in February 2011 when it trounced the two most successful human contestants on the TV quiz show

Jeopardy! Described as, “A cognitive system enabling a new partnership between people and computers,” Watson is now applied in fields such as oncology diagnosis and treatment options, financial services, and a variety of other business applications.

A powerful combination of data exploration and content analytics capabilities, IBM Watson Explorer equips users with the information gathering and evaluation techniques that will help deliver real-time results.

Cognitive and Information Analysis Services

IBM Watson Explorer includes advanced content analytics, including shallow and deep natural language processing. Information in all forms is aggregated and analyzed to reveal hidden insights and patterns.

An Application Builder brings information, analytics, and cognitive insights together into a single view of data, including both structured and unstructured, internal and external. Users receive actionable insights from a cognitive exploration-powered presentation. Watson provides a set of tools that understand and present a “unified information application.”

IBM Watson Developer Cloud Capabilities

Watson Explorer delivers information and analytic capabilities to users which help them deliver better performance and real-time results. Teams arrive at decisions more accurately and efficiently than before when Watson Explorer combines content and data, then adds analytics.

Working in partnership with IBM Watson Explorer, the Watson Developer Cloud is a collection of APIs and SDKs that use cognitive computing capabilities to solve complex problems. Most modules are already in general release, while some are still in public beta. You can try them for free through Bluemix, the IBM Cloud Platform.

Firms working in areas with multiple languages may benefit from pulling in the Language Translation modules to identify the language of written text or translate text into another language. Special knowledge domains include language terms unique to specialties such as medical terms. Also available are modules to provide streaming Speech to Text transcription or Text to Speech services.

Other modules include Tradeoff Analytics, Personality Insights, Natural Language Classifier, Concept Insights, Concept Expansion, Message Resonance, Question and Answer, Tone Analyzer, Relationship Extraction, Visual Recognition, and Visualization Rendering. Users may research the documentation for each module and run trials to test the IBM Watson Developer Cloud for themselves.

Now part of the IBM family, AlchemyAPI offers 13 semantic text analysis APIs using natural language processing to better understand conversations, reports, and documents. The cloud platform enables companies using minimal programming to turn vast amounts of unstructured data such as websites, tweets, emails, and images into facts and knowledge about the context surrounding products and services. Documentation, tools, and SDKs are available today, along with case studies and sample solutions such as Social Media Monitoring, Ad Targeting, Business Intelligence, Public Relations, and Content Discovery.

More than Words – Meaning

The combination of natural language processing and personality insights models give IBM Watson Explorer users more insight into customers' minds than ever

before. Combining the psychology of language with data analytics algorithms means users can learn more about prospects and customers with IBM Watson Explorer than any other system.

Watson's Personality Insights service analyzes unstructured data and describes Needs, such as the aspects of a product that will resonate with that person. Is your prospect more about curiosity and closeness, or love and liberty? The same research describes Values, or the motivating factors that influence a person's decision making. Knowing which prospects score high on hedonism versus helping others will help sales fine-tune a winning message.

More than Search – Solutions

When sales drop, who is to blame? The easy answer is the sales department. Have failures in prospecting and pipeline building come to the forefront? Perhaps cognitive exploration will reveal another option: customer service needs to improve.

For example, by looking at data from all areas of the company, structured and unstructured, along with outside data, management might find heightened aggravation among the customer base over a product

failure. Checking the support and help center calls confirms customer unhappiness. Natural language processing discovers that a fair majority, maybe 75 percent, of users of a particular product express negative statements on social media. Yet when customers call for help, the call center has no prepared answer, so customers become more frustrated and vent on more social media sites.

Solution? Replace the part that fails too often and too early. Reach out to your vendor for an updated part with lower failure rates. Update call center scripts to acknowledge the problem and soothe customers with suggestions. Monitor social media as users get their replacement parts or warranty service to make sure the mood of your customers perks up. Watch increased customer satisfaction scores boost sales.

Conclusion

Today, more data is under your control than ever before. Yet the percentage of internal data that provides actual value keeps dropping as the data volume keeps building. A recent Forrester study shared that a mere 12 percent of total company data is being analyzed today. This means

that the vast majority of company data, and all the potential inside that information, remains useless. While aimed at Big Data customers, this report demonstrates that every company leveraging only enterprise search to analyze their data will be leaving valuable information unmined.

IBM Watson Explorer unlocks the value within your data, utilizing that information to help employees make well-informed decisions, provide better support, and identify more customers and business opportunities. By reaching across multiple silos of information within your enterprise, search results will include information never previously integrated into single solutions. Users will benefit from search results from all the data in your company, structured and unstructured, and include data from outside as well. Rather than trying to make good decisions with limited insight, cognitive exploration users can now extract and understand all of the valuable information at their fingertips.

Turning data into information requires effort. Turning information into solutions that benefit multiple areas of your company such as sales, marketing, and customer service requires IBM Watson Explorer.

Citations:

Amount of data stored and copied in 2020: <http://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm>

204 million emails per minute: <http://mashable.com/2014/04/23/data-online-every-minute/>

Unstructured data: <http://www.robertprimmer.com/blog/structured-vs-unstructured.html>

Rise of Data Anarchy: <http://www.datasciencecentral.com/profiles/blogs/structured-vs-unstructured-data-the-rise-of-data-anarchy>

1. <http://www.zdnet.com/article/the-internet-of-things-and-big-data-unlocking-the-power/>

2. <http://mashable.com/2014/04/23/data-online-every-minute/>

3. <http://www.robertprimmer.com/blog/structured-vs-unstructured.html>

4. <http://www.datasciencecentral.com/profiles/blogs/structured-vs-unstructured-data-the-rise-of-data-anarchy>