IDC ANALYST CONNECTION

Dave Schubmehl
Research Director, Cognitive Systems and Content Analytics

How Information Access and Analysis Technology Can Address the Enterprise Knowledge Crisis

June 2016

Advances in information technologies are pushing many domains of human activity into an era of increased software and machine intelligence. These advanced systems are beginning to understand the context of content and organize information across functional areas to make new connections, discover valuable insights, and aid in decision making. IDC believes that within a decade, a majority of enterprise research applications will be enhanced by the use of technologies such as machine learning, automated content aggregation, natural language processing, and advanced semantic capabilities. This next generation of research tools has the potential to offer significant benefits for enterprises.

On behalf of its line-of-business and executive customers, IHS Engineering and Product Design addressed the following questions to Dave Schubmehl, research director for IDC's Cognitive Systems and Content Analytics service.

Q. **With respect to optimizing performance, in what areas are manufacturers and other engineering-intensive companies struggling today?**

A.   These organizations are facing unprecedented challenges from the overwhelming amount of information that is available to them but often unusable in its current form. IDC's research shows that for most organizations in general, and manufacturers in particular, unstructured information remains fragmented and spread across the enterprise. 61% of researchers and knowledge workers access 4 or more systems on a regular basis and 13% need to access 11 or more systems to get the information they need to do their job. This information is locked in a variety of formats in network file shares, content management systems, intranets, bespoke applications, and other repositories that knowledge workers have to access directly. In addition, in global organizations, this content is literally scattered across enterprises in many different locations and is usually written in several different languages.

Enterprises need to be able to find, access, and use unstructured information as easily as they are able to find, access, and use structured information. However, organizations have typically underinvested in technologies that utilize unstructured information. Instead, unstructured data often sits apart from structured data. Enterprises should be able to relate their unstructured information to their structured data in order to present a holistic view of the topic being analyzed. In addition, they should be able to easily link and relate internal information to external data sources such as patent and research paper archives as well as scholarly content and government research. However, for most enterprises, this type of

access is merely a dream. Instead, researchers and knowledge workers must go to separate applications to do research. Once knowledge workers open applications, they spend valuable time searching for and often not finding what they need. IDC research shows that knowledge workers spend 26% of their time finding and analyzing information. All of this adds up to slower innovation that inhibits profitability while increasing competitive risk, duplication of effort, and compliance issues.

Q. **How have companies invested to try to address these challenges, and how successful have they been?**

A. Companies have invested billions in structured data applications like PLM and ERP, but these systems aren't delivering the benefits that organizations require. They are usually driven by a manufacturer's IT organization with applications built around structured and semistructured data. Unstructured information, such as reports and emails, isn't handled well by these applications.

In addition, for most IT organizations, unstructured information is something to be put in a repository rather than used to help with innovation. Although unstructured content accounts for 90% of all information, organizations have generally either significantly underinvested in processes for addressing unstructured content or invested in substandard processes. Most organizations feel they have addressed general-purpose search, but for the most part, these search applications don't solve the problems of relating information across repositories and helping understand specialized content such as research documents.

Q. **What new solutions exist to address these challenges, and how are they different from the solutions that have been used in the past?**

A. Information access and analysis technologies provide a solution to the issue of siloed access. Using connectors that crawl the content, this technology indexes the information in various repositories, enabling a type of "virtual information integration" rather than moving the information. The unified indexing approach enables data enrichment through text analytics, uncovers information relationships among data in vastly different systems, and enables automated metadata extraction and tagging. This serves to increase the contextualization of content, thereby helping engineers find information in the context of problems they are trying to solve. Once the data sources are connected and there is a unified index across all needed sources, organizations can use the index as the basis for unified search and relevance. With the addition of text analytics, other capabilities may include taxonomy generation, metadata extraction, auto-tagging, and the ability to recommend contextually relevant content or experts to make it easier for users to discover and locate the information they need, regardless of where that information "lives."

Many of these systems have the ability to use text analytics and natural language processing to extract key information found in unstructured information. This key information can be used as metadata to link records, documents, and other information. For example, company research reports within an organization can be linked to external research and patent applications as well as potential competitor information. The ability to identify and link all this information can be very valuable for anyone working on research-related projects. Once information is located and discovered, many systems make it easy to share and socialize data with employees and other stakeholders such as research partners and suppliers.

Further, we are beginning to see the emergence of cognitive systems. These systems consist of advanced computing capabilities that can be brought together to ingest high volumes of structured and unstructured data from a variety of internal and external sources, understand what the data means in a particular business context, and support human decision making

with more accuracy, confidence, speed, and agility based on broader and deeper bodies of evidence applied to a more comprehensive view of pertinent conditions.

Q. **How are companies benefiting from implementing these new solutions?**

A. IDC has found that organizations that are able to combine, reuse, and share information more effectively than their counterparts, thereby eliminating rework and increasing productivity, are five times more likely to experience business benefits that exceed their expectations than companies that don't. This ability is what sets many organizations apart from their competitors. We're also seeing that organizations that use these technologies effectively are able to innovate more quickly than their counterparts, often generating more revenue per employee. These organizations are also becoming more efficient with knowledge-intensive processes such as research and development, product maintenance, and finance, as well as a host of other activities.

In addition, we're seeing that organizations using these systems have better compliance risk mitigation based on having all available information at hand for decision making. This is especially important in highly regulated industries, such as the agriculture, pharmaceutical, and energy sectors. IDC strongly recommends that organizations develop measures and methodologies for determining success before any projects start. These measurements can take many forms, such as determining anticipated cost savings, based on improved workflows, and then measuring them after the system has been in place. Other methods of determining ROI include estimating the impact of improved time to market for new projects, quicker decision making in eliminating projects that won't be successful, and even measuring less tangible benefits like worker satisfaction and productivity. Organizations that explicitly measure their efforts are much more likely to gain significant benefits than those that don't.

Q. **What are the typical challenges that a company encounters in deploying these technologies, and what can it do to avoid common mistakes?**

A. Today, many solutions combine information access and analysis technologies with trusted data, including high-quality third-party information such as research reports as well as company information and even technical news. These solutions provide a single approach for applications ranging from pharmaceutical research to manufacturing plant maintenance. Many of the solutions are available now but are not well known by IT organizations or, in some cases, the affected lines of business. Only a very few of these types of solutions are designed and developed specifically for the engineering and R&D–based user population.

Another key factor is that many of these solutions are being deployed either as cloud-based applications or as cloud/on-premises hybrid combinations. This simplifies the capital outlays for many organizations and also reduces the need for massive IT involvement to get systems up and running. IDC suggests that companies create an organizational information access and analysis strategy to tie together structured and unstructured data sources — inside and outside the enterprise — virtually and implement content aggregation strategies that can effectively access siloed and legacy data sources because they may be great sources of potentially high-quality data. It's also important to develop and promote an organizational culture that understands and embraces the collection, use, sharing, dissemination, and collaboration of information as a key asset. In addition, companies should utilize technology and services partners to most effectively prototype, design, and develop an approach that works for their organization and the type of problem being solved.

ABOUT THIS PUBLICATION

This publication was produced by IDC Custom Solutions. The opinion, analysis, and research results presented herein are drawn from more detailed research and analysis independently conducted and published by IDC, unless specific vendor sponsorship is noted. IDC Custom Solutions makes IDC content available in a wide range of formats for distribution by various companies. A license to distribute IDC content does not imply endorsement of or opinion about the licensee.