Global Headquarters: 5 Speen Street Framingham, MA 01701 USA P.508.872.8200 F.508.935.4015 www.idc.com

IDC EXECUTIVE BRIEF

The Information Advantage: Information Access in Tomorrow's Enterprise

October 2009

Adapted from *Hidden Costs of Information Work: A Progress Report* by Susan Feldman, IDC #217936, and from *Worldwide Search and Discovery Software 2009–2013 Forecast Update and 2008 Vendor Shares* by Susan Feldman, IDC #219883

Introduction

As information work has become increasingly electronic, the tools that workers need have not kept pace with changes in the workplace. In some ways, information workers have a less convenient work environment than they did 30 years ago. They spend significant time making up for the deficiencies of their workplace, manually bridging the gaps between one piece of software and another in their drive to gather, analyze, and manage a growing sea of information. That may be changing.

Organizations have begun to understand that in their drive to economize, they have cut as many jobs as they can; the next step is to streamline information work. Professionals spend significant time on tasks that could be automated. In many (perhaps most) work environments, applications are not integrated, requiring human effort to glue together information silos. Time and effort are wasted. Most of the last generation of IT innovation revolved around capturing interactions in databases. But today, the interesting interactions are at one or more levels up — between people and information or between individuals and groups. These Web 2.0 interactions happen outside the transactional framework and add new layers of complexity to the business environment.

The next generation of information work will be search based because search architectures are better suited to analyzing human interactions. These architectures are built to handle language with all its vagaries. While most people are familiar with basic search engines, they don't realize that the technologies that have developed to find, organize, and present text and rich media are fundamentally different from database architectures, which are aimed at managing and finding predictable and precise data.

Some of the earmarks of search technologies include the ability to find both exact and indirect or approximate matches, to handle an overload of text quickly, to understand key entities (such as people, companies, and topics) and meaning, or to answer questions ad hoc without needing to first restructure a database. The next generation has begun to resemble conversational systems that can carry on a dialogue with the user in order to determine what is being searched for and how best to find it.

As the demand for better work tools for professionals grows, so does the trend toward search-based applications. These new applications are built on a search technology backbone. They are designed to solve a particular business process that has a strong need for information, usually from multiple sources. Search-based application platforms combine search (usually as the underlying architecture, instead of a standard database) with workflow, domain knowledge (vocabularies, taxonomies, rule bases), task understanding, and collaboration tools.

This Executive Brief examines the high costs of information work today and the need for companies to invest in search and discovery software that can accelerate processes such as eDiscovery, e-advertising, customer interactions, categorization, publishing, and collaboration while reducing manual labor. The paper also discusses the particular advantages and challenges of search-based application platforms as potential solutions.

The High Cost of Information Work

When worker productivity increases, so do business profits. In the past century, much of manufacturing work has been automated, wringing more products out of every worker hour. But in an economy that is now more information based than industrial, increasing the productivity of the information worker has become imperative.

Moreover, as organizations seek to lower and control costs during this recession, they will need to streamline and automate information tasks and processes if they are to survive with fewer workers. Companies will need to ferret out instances of duplicated or repeated manual effort, for example. They will also need to invest in software that can speed up business processes such as eDiscovery, categorization, call center support, publishing, and collaboration — all while reducing manual labor.

Automating repetitive steps and eliminating those that waste time will increase information worker productivity and customer service and potentially save an organization millions of dollars. According to IDC research, the time spent searching for information averages 8.8 hours per week, for a cost of $14,209 per worker per year.

Analyzing information soaks up an additional 8.1 hours, costing an organization $13,078 annually. These tasks are all productive, but if they can be streamlined to reduce repetitive processes and to

improve the productivity of the aspects that require human judgment, then hours can be saved. This will free information workers to make the irreplaceable human judgments — to analyze, understand, innovate, or make decisions.

Two tasks — searching for and analyzing information — are obvious candidates for better automation. It makes sense that if workers are spending over a third of their time searching for information and another quarter analyzing it, this time must be as productive as possible. Moreover, as even more communications channels — e.g., Facebook, Twitter, or LinkedIn — need to be merged with traditional information sources, some sort of behind-the-scenes integration of all these sources becomes imperative.

The product of knowledge work is ideas, documents, data, decisions, and actions. Ultimately, high-quality information work should result in better decisions, innovative products, and more competitive organizations. A key byproduct of eliminating waste and streamlining processes should be more productive and more satisfied employees. Moreover, an integrated search-based platform approach to information work could solve many of the following common problems:

- Combine data from multiple communication and database sources
- Provide real-time access to more information — in the cloud, external or internal sources and applications
- Normalize data and content to a unified set of concepts
- Add fuzzy matching to data
- Help handle information overload by removing duplicates and pulling together related information automatically
- Mitigate risks of old, conflicting, incomplete information
- Streamline the task of gathering information to cut costs
- Provide better tools for information finding and analysis
- Broaden access to enterprise information with easier interfaces
- Leave information in place while providing access
- Integrate data, text, and rich media to pull in related materials from multiple sources

Because labor is one of, if not the biggest, expense in information-driven businesses, it makes sense to rethink the environment in which information workers operate. IDC expects that the next wave of investment will be in precisely this area.

Search and Discovery Applications on the Rise

The world may be in a recession, but the search and discovery software market grew 19% in 2008 to $2.1 billion. In addition to helping companies lower the high costs of information work, search and discovery software is seeing remarkable growth. The following market drivers are contributing to the growth of this technology:

- The world is migrating to the Web in all matters of culture, entertainment, commerce, and government.
- Business and its lackey advertising are following this wholesale migration. Further, advertising channels are mutating to find new angles of differentiation, many of which require better search technology for ad targeting.
- The continuing demand for better compliance and eDiscovery tools, especially given the turmoil in the financial markets.
- The recognition that enterprises are at risk if they do not have a single point of access to all their collections of data and content, including internal call center, marketing, or financial information as well as Web and media sources of opinions and trends.
- The need to improve online retail operations to increase revenue, coupled with the drive to reduce call center costs.
- The high percentage of government investment in the search and discovery market, which will continue in intelligence and in areas that will benefit from the stimulus package, such as healthcare, government automation, and construction projects.
- The increased use of embedded search technologies and language technologies in other applications (such as CRM, business intelligence, and social media) for globalization, translation, or reputation monitoring. In the digital marketplace, search and discovery applications are used widely for better ad matching and in recommendation engines, geolocation applications, or personalizing and customizing applications based on search histories.
- Better product differentiation among search and discovery applications for site search, intranet search, search-based applications, or integrated information access platforms.
- More applications with hybrid architectures that support unified access to both content and data in multiple formats and repositories.
- Widespread embedding of search technologies in software for information-heavy uses such as CRM, sales, research, publishing, business intelligence, and content management.
- Better and more pervasive analysis and reporting tools; increased and more effective use of visualization techniques.

- Increased use of open source applications such as Lucene, Solr, GATE, and other text analytics/NLP tools.

- Increased integration of NLP tools such as entity extraction and categorization and clustering in larger search applications.

- An emerging market for "ETL for text" — tools and information infrastructure components such as smart connectors, metadata extraction, linguistic analyzers, taxonomy builders, and categorizers.

Search and discovery applications create access to unstructured information and provide alternative access to structured data. This group of software applications analyzes, tags, and searches text, often in multiple languages and from evolving Web 2.0 sources and rich media such as audio files, video, and image files. This market, as defined by IDC, includes the following applications:

- Extended search platforms

- Search engines

- Question-answering applications

- Categorization/metadata tagging tools

- Categorizers and clustering engines

- Visualization tools for information navigation and analysis

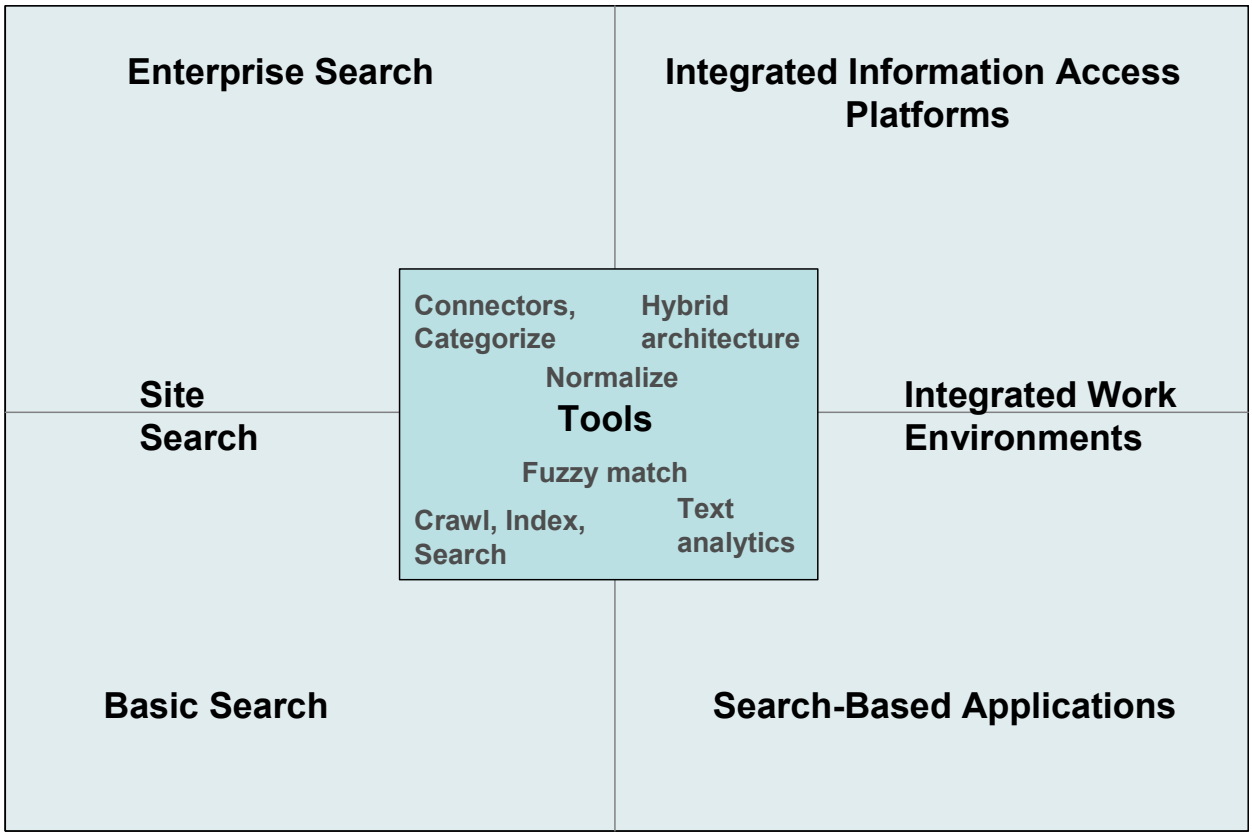
- Filtering and alerting tools and text analytics

- Translation and globalization software

Search vendors have begun to differentiate their products, and sophisticated buyers are realizing that when it comes to search, one product does not fit all needs. Potential buyers of search and discovery software can better sort out the myriad offerings by dividing current products into five segments that rarely compete against each other (see Figure 1):

Figure 1

2009 Search and Discovery Software Market Map



Source: IDC, 2009

- **Basic search engines** are single-purpose, standalone search products that provide out-of-the-box crawling, indexing, and searching capability. They are only minimally customizable.
- **Enterprise search applications** still focus on search as their main purpose. These search applications are highly customizable and often serve as a quick and noninvasive way to integrate multiple content applications and repositories across the enterprise.
- **Search-based applications** are designed to solve a particular business problem or streamline a business process that has a strong need for information, usually from multiple sources.
- **Integrated information access and management platforms** will be the high-end enterprise, intelligence, and ecommerce information infrastructure of the future. These platforms handle large volumes of queries and documents from multiple information sources with all kinds of formats — data and content

— including rich media. They provide a common platform for access to all enterprise information as well as connectors to external sources. These high-end platforms enable large organizations to build custom search-based applications for specific purposes.

- **Information access tools** gather, tag, organize, normalize, analyze, transform, or translate text. They extract and add meaning to text and rich media records in order to improve retrieval across multiple repositories and return more accurate results at retrieval time.

Of all these segments, search-based applications are the ones that foreshadow how information work will be performed in the future. They create a polished, integrated work environment for information workers for eDiscovery, sales, research, reputation monitoring, voice of the customer, or customer support. The work environment hides the complexity of the underlying multiple information sources and applications.

This is a rapidly emerging category that combines search (usually as the underlying architecture, instead of a database) with workflow, domain knowledge (vocabularies, taxonomies), task understanding, and collaboration tools. These applications can be installed quickly and have rich, easy-to-use interfaces and end-user tools to simplify a particular task.

Search-based applications are an excellent example of a larger market trend toward convergence of technologies into ready-to-use applications. They are also easier to market because they address information problems that are readily apparent to business users rather than the underlying complex technology. This trend toward using search as one of a suite of integrated tools to support a particular business problem shows a growing maturity in the market as well as the perception that information tasks need better tools if they are to be productive and mitigate the risks of poor decisions, slow reactions to customer complaints, or noncompliance with regulations.

There's a place for each of these application types in the market, but buyers need to know the purpose for which they are buying an information access application or platform because not all search engines are the same.

The Future of Search and Discovery

IDC believes that search technologies are already becoming a necessary part of most information-rich applications because of their flexible architecture and ability to handle and mine fast-changing information environments without restructuring. Combined with information management tools, collaborative technologies, and more advanced administrative and end-user tools, these software applications have become the de facto gateway to information for businesses and individuals on computers and mobile devices.

For the enterprise, we foresee a continuation of the current mixed environment, with departments and smaller businesses investing in quick-to-deploy search-based applications, at the same time that enterprise computing moves to integrated information access, management, delivery, and analysis platforms that enable easy exchange of information from multiple underlying sources.

This category of software is at the nexus of change and consolidation in enterprise computing. Therefore, it is becoming part of a larger information-centric software stack that may also include the following:

- Tools to manage the structure of information — metadata, schemas, and a semantic layer (schemas and metadata govern how information is integrated, retrieved, displayed, and analyzed)
- Content management systems
- Databases and data integration tools
- Master data management tools
- Collaborative tools, including wikis, blogs, and social networks
- Workflow engines

We are already seeing a shift in emphasis from transaction-based to information-centric or meaning-based computing. We are moving from managing and validating separate data points to synthesis and analysis. New applications will provide more real-time, ad hoc query, analysis, and reporting functionality. The interactions will be more conversational, as text joins data in creating a more complete view of the business or topic.

Ambiguity will need to be resolved, large amounts of text will have to be visualized effectively, and visions of what an information access and analysis application can and will do will broaden to include finding experts, locating and merging contract information with analyses of customer financial soundness or history, analyzing customer churn by mining call center calls, or predicting ecommerce purchases based on clicks or purchases by customers with similar profiles and interests.

That's a lot of data to crunch, from multiple sources, and new, powerful information tools will combine these technologies to help businesses understand their customers and market trends in real time. In fact, that is already happening.

Unified access and management platforms integrate structured and unstructured content in a unique way to address mission-critical applications for extended business intelligence, customer support, compliance, and other areas.

The best-in-class platforms offer customers scalability, reaching across multiple content repositories — including desktop, legacy applications, third-party outsourcing providers, and the Internet — to bring information access to new levels and decision-making intelligence to business professionals throughout the enterprise.

Challenges to Adoption

Needless to say, there are challenges to be met as the market migrates to a new computing platform. Changes of this magnitude don't happen overnight. For enterprises to move to a new computing platform, they must perceive themselves to be at a severe competitive disadvantage.

In 2009, we see early adopters, far-seeing enterprises that understand that they are information dependent, adopting the early forms of these platforms in order to integrate their streams of customer information into product planning cycles or their compliance infrastructure with their financial systems, their customer service, and their strategic planning process.

Information-centered organizations integrate their data streams and then mine them jointly to strategize as well as to avert disasters. These data integrations help them monitor their businesses continuously for early warning of problems or ideas for new products or services. The point is that a dynamic information system can point out surprises that won't turn up in a standard data warehouse–based report because anomalies are easier to find.

Other challenges to adoption of search-based applications and integrated information platforms include the following:

- Unwillingness to invest, especially in a slow economy
- Lack of perceived value of implementing highly usable search-based applications among IT departments that are used to database applications

IDC believes that the gradual shift to cloud computing will mitigate some of these risks by removing the need for large up-front investments and by allowing vendors to provide the needed expertise. As with all new generations of technology, we will reach a tipping point in the next three to five years when the market perceives these new, integrated information management and access platforms as the norm.

As the tools improve and become more pervasive, the accepted vision of what an information-driven organization should be will permeate business culture, and adoption will become the norm. But first movers are already gaining an advantage that will make them hard to catch.

Conclusion

The recession will continue to influence the search market in several ways, including new businesses substituting technology for staff. These companies will invest in affordable IT — cloud computing, inexpensive search, open source applications, and pay-as-you-go subscription models that can grow over time.

New businesses will also build impressive-looking online storefronts to serve unmet needs. These storefronts will be built on tools that help them merchandise, provide support to customers, and provide back-office infrastructure. To succeed, they will depend on affordable technology to sell services and goods, partner or develop channels, offer good customer support, fulfill orders, and keep in touch with a far-flung, loosely knit workforce of contractors.

Technology- and information-savvy entrepreneurs will start information-heavy consulting businesses that depend on their personal but global networks of colleagues. These global networks will support each other, providing complementary knowledge and skills to each other on an ad hoc basis.

They will need affordable, secure, and easy-to-use business network services to provide adequate bandwidth, collaboration tools, and business support services. They will also need affordable access to the high-value information that they are accustomed to using; the Web won't be enough.

IDC sees enterprise search and discovery applications evolving into the following two major areas:

- **Integrated information platforms that unify access to and management of information, with a healthy dose of collaboration thrown in.** The largest search vendors as well as a growing number of big software vendors have been moving in this direction already. These enterprisewide platforms address the need to consolidate related information work tasks into a single, streamlined workflow that hides the complexity of learning separate applications. Better interaction design and specific uses of the platform will be supplied, at least initially, through custom services. Although they may be relatively quick to deploy (months) in their basic form, full deployment with user interfaces, workflows, well-tuned relevance, knowledge bases, and taxonomies can take a couple of years. Much of that time is spent cleaning up duplicate and dirty data, harmonizing schemas, and normalizing terminology. The result should be better retrieval across multiple repositories and formats, giving enterprises a better understanding of their businesses rather than the partial and fragmented view they have today.

- **Search-driven applications that either are built on integrated information platforms or are packaged as separate applications.** Search-based applications, built on a search backbone but designed to facilitate a particular task and to

create an integrated work environment for users, will proliferate and flourish. Because they make sense to business users, they are already popular. Search-based applications streamline knowledge work, making information workers more productive. These applications embed search and discovery technologies as a component, but their selling point is that a worker can sit down and accomplish a job without having to move from one information source to another or from one application to the next. Successful vendors will build intuitive applications to facilitate sales, research, loan processing, marketing, financial analysis, eDiscovery, or call centers. These full-blown applications will develop integrated work environments in which the UI design hides the complexity of multiple information sources and applications.

The following trends and growth areas will help the search and discovery market continue to outperform the IT industry as a whole:

- **The need for automating information work.** As the labor force is cut, businesses must continue to operate. Automation becomes the only substitute for people.

- **Healthcare automation.** One of the first priorities in the United States appears to be to improve and streamline the healthcare industry. Providing online access to healthcare records across offices and repositories will be a significant part of this initiative.

- **Engineering and construction.** Investments in improving infrastructure for transportation of people as well as for data will take longer, but they should begin to have an effect, possibly by 4Q09.

- **Government intelligence.** The intelligence community has always been one of the primary investors in search and discovery technology. That will continue, particularly in the areas of homeland security, fraud, and terrorist detection and monitoring.

- **Government computing.** The federal government has often been accused, and rightly so, of lagging behind the commercial sector in its use of current, off-the-shelf technology. It seems clear that the new administration will increase its use of electronic communication and networking tools and software dramatically, changing how those inside and outside government interact with each other. Management of and access to the rich lodes of government information will be improved. Management of projects will rely more heavily on electronic infrastructure to not only manage but also monitor these projects. We can expect the federal government to lead these efforts, with cash-strapped state and local governments lagging behind.

- **Compliance, eDiscovery, risk mitigation, and reputation monitoring.** These four areas continue to be additional bright spots for the search and discovery market.

The seeds for new integrated information platforms have already been sown. With scaled-down workforces and an exponentially growing collection of information, organizations are already seeking new approaches to automating knowledge work. Large businesses have already invested in integrated information platforms for ecommerce, logistics, and business intelligence. Governments have been integrating intelligence information streams themselves for several years and are investing in commercially available large-scale systems to speed development time.

As the market realizes the competitive advantage and flexibility that these businesses have, adoption will accelerate. As the search and discovery market moves from emerging to mature, the unbounded potential of the technology will make it pervasive in software that requires a modicum of language understanding to improve user interactions, to perform business intelligence ad hoc and in real time, and to find similar as well as exact matches in both content and data.