IDC
Analyze the Future

WHITEPAPER

Better Search Experiences with Unified Information Access

Carrie Solinger and David Schubmehl

Sponsored by Lucidworks

White Paper

Better Search Experiences with Unified Information Access

Sponsored by: Lucidworks

Carrie Solinger David Schubmehl

June 2016

IDC OPINION

Information has a central role in most organizations today. Yet because there is significant complexity involved in managing and understanding the scattered information environment, organizations continue to struggle with implementing systems to help employees find the right information when they need it. Search and discovery technologies are often not well understood by internal IT organizations or their client business units. There is considerable confusion among buyers, who need to choose between generic enterprise search and discovery platforms and more specialized unstructured information access and analysis applications. The first generation of enterprise search succeeded by assisting organizations in handling file shares and intranet sites, but these systems did not fully address the vast amount of unstructured data integration needed within various applications and content repositories. Early implementation of these systems proved fruitful in providing value and solving problems within limited domains. However, these older enterprise search systems revealed limitations that demanded a new generation of solutions using semantic and linguistic capabilities, which we are calling "unified information access platforms."

A unified information access platform provides a single point of access to information from multiple heterogeneous sources and integrates and finds relationships across sources. An excellent example of a unified information access platform is Lucidworks Fusion. Lucidworks is a United States-based technology company that is the primary commercial sponsor of the popular open source search software Apache Solr. Lucidworks builds enterprise search solutions for some of the world's largest brands. Its application development platform, Fusion, provides all the capabilities needed to design, develop, and deploy intelligent search applications for unified information access.

The vast majority of money in organizations' IT budgets is spent on structured data analysis because unstructured content analytics and search are held to a higher standard than other, more widely deployed and broadly used techniques. In addition, IT and senior managers are much more familiar with structured data analysis techniques and management systems such as traditional CRM and ERP systems. Organizations traditionally have focused less energy on unstructured information analysis, but they should direct more resources to it because of the volume of unstructured information in their organizations and the potential value that can be unlocked. In our research, we are beginning to see organizations focus more resources on unified information access to realize the benefits that can be derived from the use of such platforms. Platforms like Lucidworks Fusion offer these benefits, and organizations are beginning to adopt these types of systems instead of more traditional enterprise search solutions.

IN THIS WHITE PAPER

This white paper discusses the imperative demand faced by organizations to create satisfying and relevant search experiences for their users. Executives who read this white paper will learn what it means to the overall organization to provide unified information access; the impact and benefits of and challenges to pursuing a strategy to unite information across an organization; and the strategies and examples of organizations that have effectively utilized tools to support unified information access in the pursuit of delivering a superior and more relevant search experience for their users.

SITUATION OVERVIEW

The State of Enterprise Search

Access to real-time unstructured content from external and internal data sources continues to grow. Approximately 90% of an organization's information is unstructured. Organizations need to contend with the increasing number of information sources and repositories that contain these massive amounts of unstructured data, in addition to structured and semistructured data. As a result, enterprise search platforms became the first generation of search and discovery solutions.

Enterprise search is a vertical search, software-based solution that takes content from multiple enterprise-type sources (e.g., file systems, databases, intranets, document management systems, and emails) and indexes, searches, and displays results to specified authorized users. These systems integrate structured, semistructured, and unstructured data; however, they handle unstructured data less successfully. Enterprise search platforms are also used to control access in enforcing security policies over their users.

Enterprise search systems require content to go through multiple phases as data is converted from source repository to a list of ranked search results. Components of traditional enterprise search systems include:

- Content awareness in a push and pull model (pushes new content to its APIs for real-time indexing and pulls new, updated, or deleted content from sources via connectors, respectively)
- Content processing analysis, which normalizes different document types for better search recall
- Indexing of the normalized text for faster searches without the need to store the full text of the document
- Query processing, which consists of any terms users may enter in their search process
- Matching, a processed query compared against the stored index to provide results, often the indexed document

Further, enterprise search systems include mechanisms such as queries, federated search, entity extraction, enterprise bookmarking, faceted search, text clustering, and access control. However, there are limitations to many of these functionalities: Enterprise bookmarking, a collaborative tagging system for capturing structured and semistructured enterprise data knowledge, does not include unstructured data; text clustering renders faceted search techniques less effective; and access control in an enterprise search environment is a complex task to cover comprehensively. As such, a large number of organizations are looking for better solutions to the shortfalls of enterprise search.

Some enterprise search systems use a variety of components to work effectively. Enterprise search systems that use a database for storing and accessing metadata and full-text index for word search can be expensive and complicated. Unified information access platforms remove these barriers by streamlining the interactions within the application layer. A single query produces all the information across all the sources versus multiple queries against multiple repositories, which is sometimes called federated search. The ability to unify information at the repository level rather than federate searches across multiple repositories and indexes provides better results and relevance and simplifies search. This allows information of all types — structured, semistructured, and unstructured — to be analyzed very quickly. This velocity leads to more rapid data-based decision making, which often results in increased return on investment (ROI).

As the volume of information grows exponentially, licensing schemes that require an organization to spend more money as it uses more information are not a viable solution given that enterprises are always accumulating more information. In other words, tethering the cost of information solutions to the amount of data an organization supplies or incorporates is cost prohibitive. Using free open source software such as Apache Solr is an option that removes that licensing expense concern. Solr is an open source enterprise search platform with roots in the Apache Lucene project. It is often referred to as Solr/Lucene, is written in Java, and delivers many of the major features that businesses require to move from enterprise search to unified information access. However, these open source software solutions can present a different set of complexity, implementation, and expertise concerns.

Enterprise search systems require accessing multiple information silos. An information silo is a collection of content that is available to specific users. These separate information silos are difficult to manage for a variety of reasons. Separate repositories often have duplicate data across two or more of the silos, which renders the process of finding, updating, and deleting duplicate data clumsy and difficult. In addition, permissions tied to the different data need to be configured across all the information repositories. This can prove unwieldy as changing permissions on even one folder containing thousands of documents can be costly and time consuming and opens up greater possibilities of security breaches. Unified information access systems offer security and access control capabilities that don't have to change the physical permissions of any files indexed by the system.

Many legacy enterprise search vendors used separate relational databases to store the user permissions. When a user initiated a query, the query was run against the search index and then the permissions database was used to filter the results as they went back to the user. This action was processed without security, resulting in poor performance and possible security leaks. In this less-than-ideal environment, organizations opted to not provide access to secure unstructured information or several silos at one time, let alone put unstructured information into a database. Unified information access platforms, unlike their enterprise search predecessors, provide a seamless solution by modeling the secured data as a database would: keeping documents separate. When users input a query, their credentials are used to identify the information they are authorized to see and then matched to the authorized data.

Many organizations remain reluctant to move from legacy enterprise search systems to better-equipped unified information access platforms that address the need for enhanced aggregation; filter personalization; analytics on content, users, actions, trends, events, navigation, browsing, and discovery; question-answering, trending, and location-based services; and advertising and marketing.

Why Unified Information Access?

End users do not need data — they need information and the tools to understand that information. Organizations have realized that "big data" is a comprehensive problem: Too much data equals information overload. As such, it is difficult to make the right information accessible. IDC estimates that more than a third of an average knowledge worker's day is spent searching for and consolidating information across a variety of data sources. In addition, these knowledge workers find the information required to do their jobs only 56% of the time.

This requires organizations to find solutions for more information and improved research to deliver data-based decision making. Increased understanding of how to use this deluge of information demands a need to unite all this disparate data to increase revenue, reduce risk, and reduce wasted time searching for information. Unified information access focuses on helping businesses increase productivity, cut costs, and address security and compliance concerns more efficiently.

Unified information access platforms, in comparison with legacy enterprise search platforms, provide a single point of access to information from multiple heterogeneous sources. They integrate and find relationships across sources in the following ways:

- Integration of access to unstructured, semistructured, and structured information
- Use of natural language processing to identify and extract key attributes and entities for improved recall
- Incorporation of machine learning to provide improved search and information relevance as the application gets "smarter" over time
- Combination of features of database, business intelligence, and search technologies in a single architecture
- Providing a modular, well-integrated set of tools and services to normalize, index, enrich, search, query, present, visualize, analyze, and report information
- Creation of a single platform for information gathering and analysis and decision support
- Accommodation to quickly update information through real-time or near-real-time updating and analytics
- High scalability
- Providing a platform for building search-based applications that support specific industries, tasks, and workflows

In addition to these unified information access platforms, IDC is seeing specialized applications that provide a targeted, integrated information work environment that is specific to a process or business unit. These applications:

- Are based on the task, not the technology
- Combine multiple technologies and tools
- Integrate information from multiple sources
- Incorporate knowledge bases for specialized workflows and terminology
- Hide complexity behind an easy-to-use, compelling user interface (UI)

These applications are often used in many different ways and support mobile and tablet options as well as desktop uses. An organization can use these types of applications to implement a customer portal to improve the overall value of customer service by reducing call times/volumes and answering

customer questions the first time while giving self-service access to customers so that they can solve their issues or check the status of issues.

Each time a sales professional interacts with a customer or prospect, he/she must be prepared to answer a host of questions and have knowledge at his/her fingertips. With a diverse set of complex products and services to sell, sales professionals often rely on many forms of dispersed knowledge assets, as well as distributed employees, for the support and information needed to move the sales process forward. Providing this information and support through the use of a targeted search application can help reduce the sales cycle significantly, improve prospects' satisfaction with the potential solution, and differentiate the account executive from his/her competitors.

Lucidworks, a big proponent of and contributor to the Apache Solr developer community, provides the gateway for organizations to adopt open source Solr as a unified information access solution. Solr's major features include full-text search, faceted search, real-time indexing, database integration, NoSQL features, rich document handling, distributed search and index replication, and an architecture designed for scalability. Solr runs as a standalone full-text server, using the Lucene Java search library for full-text indexing and search, with APIs that make it usable for the most popular programming languages. Solr's external configuration provides tailoring for many types of applications without Java coding. Solr has a plug-in architecture to support more advanced customization.

Solr is an excellent open source search solution, but it is not truly enterprise grade out of the box. Solr lacks advanced administration and management tools, data connectors, enterprise-level security, and a host of other capabilities needed to create an enterprise-grade unified information access platform. While Lucidworks provides enterprise search support to its customers that use Solr on their platform, organizations may still require a broader-based industrial-strength solution. The Lucidworks Fusion product meets these demands.

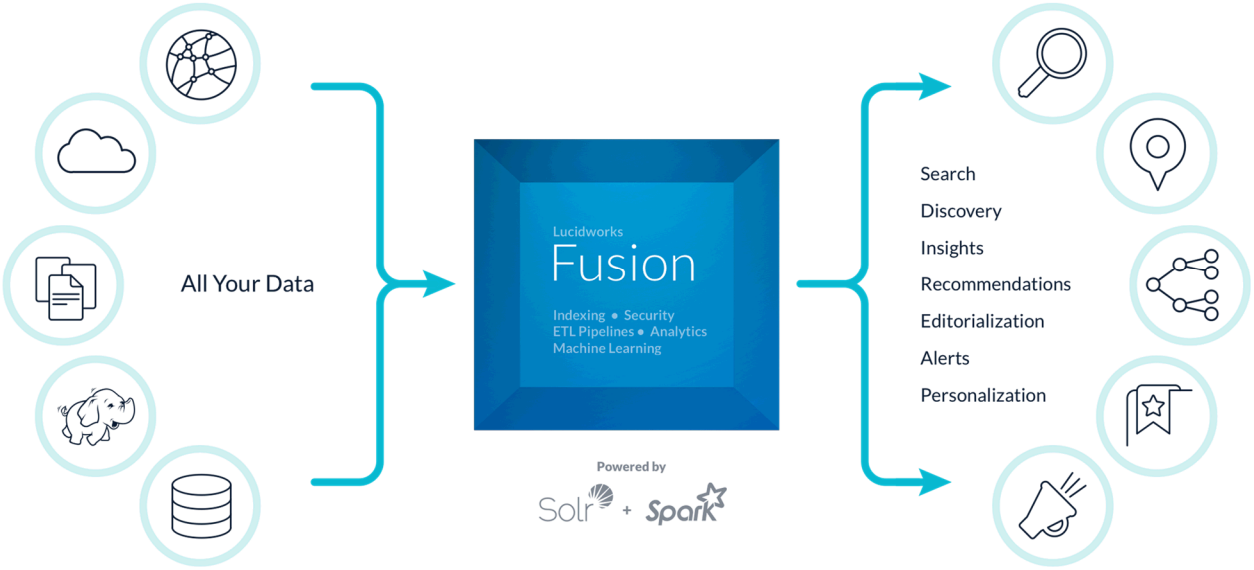
Lucidworks Fusion

While Apache Solr is a great search platform, it is not necessarily suited to the complete needs of an enterprise. Lucidworks worked to solve that challenge by offering Lucidworks Fusion for organizations that have realized the need for more than just search — that is, the need for unified information access as outlined previously. Lucidworks Fusion provides the enterprise-grade capabilities needed to design, develop, and deploy powerful search applications built on open source Apache Solr. Fusion works as a standalone search platform or as an extension of existing Solr implementations. As outlined in Figure 1, Fusion includes Apache Spark, advanced analytics, full enterprise security, machine learning, a robust set of data connectors, index pipelines, UI APIs, and single-point administration tools to provide distributed information search and access in a 24 x 7 environment.

Lucidworks Fusion's Spark integration within Fusion's data processing layer enables real-time analytics within the enterprise search platform, filling in a gap in Solr's capabilities with data analytics. Lucidworks has added Spark integration to the Solr architecture specifically to accelerate data retrieval and analysis. Developers using Fusion also have access to Spark's store of machine learning libraries for data-driven analytics.

FIGURE 1

Lucidworks Fusion Architecture



Source: Lucidworks, 2016

Connector technology is a key capability for successful unified information access applications. Many organizations want to extract content from SharePoint sites or Hadoop data lakes. Lucidworks Fusion's Connector Framework provides these kinds of capabilities and makes data ingestion simple and painless. Fusion's data connectors also extend and simplify the Apache Solr experience. With connectors to over 60 data sources and repository types, Lucidworks makes it easy to ingest gigabytes and terabytes of content seamlessly and make it indexable and available for analysis and retrieval. The indexing pipelines also streamline this ingestion capability, providing flexible and sophisticated tools for transforming and adjusting content as it makes its way into the indexing process.

Lucidworks Fusion connects to Solr through sophisticated APIs, extending its capabilities and making a powerful platform for a new generation of data-driven applications. Fusion enables teams to quickly build and deploy powerful search applications across the enterprise. In many ways, the enablement of a typical unified information access platform can help with the transition to more targeted, relevant search-driven applications.

For Gabe Arnett, senior director, Search & Data Strategy, Moody's Analytics, Lucidworks Fusion is its "easy button." According to Arnett, "[With Fusion], we started seeing the capacity to reduce our development efforts even further and to really shift search and the creation of the search experience itself to the business analysts with their strategy perspective so they're more in control. That's one of the biggest strengths that Fusion gives us, the reduced time to market with the easy user interface. I can get up and running indexing a collection of information from many different sources, testing it, and validating it several times in a matter of hours versus development efforts of months or quarters."

Lucidworks Fusion has enabled the organization to move farther and faster than it would have with a traditional enterprise search platform or with just Solr alone. Moody's has seen significant return on investment from the deployment of Lucidworks Fusion compared with its prior proprietary enterprise search solution, and it has also seen a significant reduction in development times while providing better, faster, and more accurate search results.

Other firms interviewed by IDC highlighted different aspects of Lucidworks Fusion, such as its connector technology as well as its security capabilities. For example, a search manager for a global financial institution mentioned how the company sees Lucidworks Fusion and Solr: "Our speed of deployment is of utmost priority. Timelines are everything. You have to meet the timelines. Having an environment that's already built out in Lucidworks Solr and is known and can pass security audits and is up 24 x 7 is a huge factor for us. Being able to develop something on it faster than anybody else can bring up something else is also a key attribute in our organization."

After government probes resulted in prosecution and fines for insider trading, a Fortune 100 financial services firm knew that it had to put practices and technology in place to make sure it never happened again. The Lucidworks Fusion team built a system that ingests, indexes, and monitors all incoming and outgoing communications — emails, chats, and telephone transcripts — searching for anomalies or patterns that might point to an employee trying to conceal possible collusion or other illegal activity. With powerful algorithms that process these data streams in real time, analysts can quickly identify potential bad actors and investigate them and seek swift resolution to keep the organization firmly within legal compliance and out of the purview of regulators. As the program manager for the firm said: "We did some analysis on the industry space to see who are the experts, and we narrowed in on Lucidworks. Their product, Lucidworks Fusion, which sits on top of Solr, has been a good fit for what we wanted to do."

CHALLENGES/OPPORTUNITIES

Unified information access is the next generation of enterprise search. The benefits are bountiful in comparison with enterprise search, with unified information access actually delivering the promise of enterprise search. As more information is created, there is a greater demand for newer technologies and automation techniques that will help organize and make this information accessible and actionable. Therefore, unified information access must continue to evolve. Unified information access still needs to tune its methods of:

- Information collection
- Data normalization and relationship mining
- Adaptive software to respond to changes in the data environment
- Visualization
- Reporting and alerting
- Collaboration and social business tools
- Better interaction design, tools, and intelligent workspaces

Supporting these methods is both a challenge for the enterprise organization and an opportunity for a company such as Lucidworks. Organizations need to innovate beyond traditional enterprise search, and Lucidworks Fusion provides the capabilities and functionality.

CONCLUSION

Organizations need to recognize that their two most valuable assets are their data and their employees, and helping employees make that data actionable and productive will yield long-term benefits to both the employees and the organization as a whole. Our research shows that successful businesses not only recognize that their employees and staff are absolutely essential to the organization's long-term success but also provide necessary funding to support improved employee productivity, collaboration, and knowledge sharing in order to grow the value of their businesses. Organizations implementing unified information access achieve significant ROI, saving money and also increasing revenue as their staff become more productive.

Consider the following questions:

- What is the dollar value of improved morale, satisfied customers, or more systematic capture and transfer of employee knowledge?
- What does a company lose by delaying implementing unified information access or implementing it in such a way that users still cannot find and use what they need to do their daily jobs? How will that impact employee productivity and satisfaction, and what will it mean for IT's reputation to provide the tools necessary for knowledge workers to compete effectively in today's world?

As organizations become more distributed, efficiently sharing information is increasingly important. Legacy search and discovery technologies are simply not agile enough to manage this data in a timely, distributed manner. However, unified information access can — it was designed with these challenges in mind. Unified information access allows information to be of any type and format. Unified information access systems combine access to structured, semistructured, and unstructured data in new ways, which allows end users to obtain a holistic view of a particular topic, situation, or subject.

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

5 Speen Street
Framingham, MA 01701
USA
508.872.8200
Twitter: @IDC
idc-community.com
www.idc.com