



Enterprise search – trends and developments

Intranet Focus Ltd Research Note 06/12

June 2012

Intranet Focus Ltd
12 Allcard Close
Horsham RH12 5AJ, UK
+44 1403 267030
www.intranetfocus.com

Enterprise search - trends and developments

Summary

After a long period of benign neglect it would seem that there is a renaissance in enterprise search, with major acquisitions and a number of new application areas emerging. One result of this is to question whether 'enterprise search' is the correct description for either the technologies or the outcomes of the process of information discovery and analysis. At the Enterprise Search Europe 2012 conference there was considerable unanimity that the factors influencing this renaissance were Big Data, mobile search and cloud services. This Research Notes sets out some of the opportunities and issues arising from these and some other factors, and looks at the emergence of Big Data, search-based applications, unified information access, federated search, mobile search and federated search. Recent major acquisitions in the search sector are reviewed.

Contents

1. Introduction	3
2. Industry consolidation	3
3. Microsoft and search	4
4. Open source search software	5
5. Big data	5
6. Unified information access	6
7. Search-based applications	7
8. Federated search	7
9. People and expertise search	8
10. Mobile search	8
11. Cloud-based search	8
12. In conclusion	9
Resources	11

Research Notes

This is the sixth in a series of Research Notes that Intranet Focus Ltd are publishing in 2012. For further information see <http://www.intranetfocus.com/resources/downloads>. Previous Research Notes covered enterprise mobile strategy development, enterprise search team management, digital workplaces, virtual teams and legal issues for intranet managers.

Enterprise search - trends and developments

1. Introduction

After a long period of benign neglect it would seem that there is a renaissance in enterprise search, with major acquisitions and a number of new application areas emerging. One result of this is to question whether 'enterprise search' is the correct description for either the technologies or the outcomes of the process of information discovery and analysis. At the Enterprise Search Europe 2012 conference there was considerable unanimity that the factors influencing this renaissance were Big Data, mobile search and cloud services.

At last in-depth market and customer information is available on the use of enterprise search. Initial results of a study of over 170 organisation in North America and the EU sponsored by Findwise, the Swedish search integration company, have recently been presented at the Enterprise Search Summit in New York and the Enterprise Search Europe conference in London. The full analysis will be available in the near future.

This Research Notes sets out some of the opportunities and issues arising from these and some other factors.

2. Industry consolidation

Compared to most sectors of the enterprise applications business the enterprise search business is quite small. The total annual sales of search software may only amount to \$3billion at most, which in IT terms is a niche market. In total there are probably no more than 80 companies in the business at present. Most of the vendors have revenues of less than \$50million and many may have revenues of less than \$10million.

There have been a number of high profile acquisitions over the last ten months, starting with the acquisition of Autonomy by HP in 2011 for around \$8billion. The fit between an entrepreneurial company of 2000 employees and HP with over 350,000 employees was always going to be a difficult one to engineer. Since the acquisition many senior Autonomy managers had left the company and the contract with Dr. Mike Lynch, the founder of Autonomy, was terminated. Next in line came Oracle, acquiring Endeca for an un-disclosed amount. Oracle has been in the search business for many years, launching Oracle Secure Search in 2006. Unlike IBM, Oracle has been more than happy to sell the product to non-Oracle customers. Oracle moved quickly to integrate Endeca into its product catalogue.

Probably the most surprising entrant into the search business has been Lexmark, which was founded in 1991 when IBM divested its printer business. In 2012 Lexmark diversified into the search business acquiring Brainware and then Isys-Search, a leading vendor of midrange enterprise search applications. Lexmark had previously acquired Perceptive Software in 2011, a supplier of enterprise content management and business process management software. Finally IBM stepped into the market. The company has a long history of research and development in information retrieval and over the years it has acquired many small companies with search expertise, rapidly integrating them into its core product range. Interestingly IBM's Omnifind search product is based on the Apache-Lucene open-source search application. In 2012 IBM acquired Vivisimo, a search application developed at the Carnegie-Mellon University, on the basis that the Vivisimo clustering technology would be of benefit in developing content analytics and data mining applications.

With all these companies there are some trade-offs between benefits and risks when choosing their search solutions. The benefits are that these companies have the financial resources, sales teams and commercial incentives to ensure that the costs of acquisition are defrayed as soon as possible. They will almost certainly have existing contracts with major companies and will quickly be in a position to offer

Enterprise search - trends and developments

enhanced search applications. Procurement departments will feel comfortable that the companies will remain in existence for some years to come, and IT managers know that they will be able to call on service and support teams in most countries of the world.

The risks are that as search is a small element of total sales enterprise application sales teams will probably not have the detailed product knowledge that is essential in ensuring that user requirements can be met by the vendor's search solutions. In the case of IBM and Oracle, and to a lesser extent Lexmark, the companies have multiple search applications to offer. Organizations who have already implemented Autonomy, Isys-Search, Endeca and Vivisimo will find that the people they have been dealing with for some time may not be able to make commitments to solving problems with the same degree of rapidity as they have done in the past.

3. Microsoft and search

Microsoft came late to enterprise search. The search functionality of the 2007 release of SharePoint was poor and in an effort to catch up Microsoft acquired the Norwegian software company FAST Search and Development in 2008. FAST ESP was a very powerful enterprise search application running on Linux servers. Microsoft moved the application to Windows servers and used the expertise of the company's developers to enhance the search functionality of SharePoint 2010.

The search functionality of SharePoint 2007 (often referred to as MOSS07) was poor and many organisations used one of the many third-party solutions (such as those from BA-Insight and Coveo) that were designed specifically for SharePoint. As a result few Microsoft Partners developed any expertise in search implementation, and had to build this capability with the arrival of SharePoint 2010.

The range of search applications available from Microsoft is quite complicated to understand and includes

- Search Server Express is a free product that can be downloaded from Microsoft and installed on a single server. It can be used to index up to 300,000 items.
- Search Server 2010 has the same functionality as Search Server Express but can index up to 10 million items per server, and up to 100 million items on multiple servers. There is a per-server license fee.
- SharePoint Server 2010 is the entry-level search for SharePoint 2010. It is bundled in to the Standard CAL (Client Access License).
- FAST Search Server 2010 for SharePoint brings much (but certainly not all) of the functionality of FAST ESP to SharePoint search. The power and complexity of the product are both substantially greater than SharePoint Server 2010.
- FAST ESP is one of the most powerful search applications on the market, but has not been developed in any way since it was purchased by Microsoft in 2008. It goes out of main-stream support in July 2013.

Many organizations are confused by the FAST prefix to FAST Search Server for SharePoint 2010 and think that they have purchased the FAST ESP product. That is not the case. Certainly FAST Search Server for SharePoint 2010 provides considerable search functionality but it is configured to run inside SharePoint. As a result the processing power and ability to customize the application are somewhat reduced and some of its features depend on the use of SharePoint 2010 in creating the original content.

SharePoint 2010 was released in May 2010, and it is likely that the next release of SharePoint will be available in 2013. The indications are that there will be further enhancements to the search products within SharePoint, and these may well provide a migration path for customers who are currently using

Enterprise search - trends and developments

FAST ESP but wish to move away from this application because of the cessation of main-stream support from Microsoft next year.

Despite the improvements to search in Microsoft SharePoint 2010 a number of vendors see significant opportunities in providing enhanced search services for SharePoint implementations. BA-InSight has been in this business for some time, starting with SharePoint 2007, but in quite a surprising move Recommend (which has traditionally focused on e-discovery) has recently launched a SharePoint-specific solution.

4. Open source search software

The open source enterprise search business is going to develop substantially over the next few years. The spate of acquisitions in 2011 and 2012 and the question marks over the future of FAST ESP have both caused organizations (other than maybe those implementing SharePoint) to look more carefully at the potential benefits of a customized open-source development option.

Companies providing open-source solutions emphasise that “open-source = freely available” and not “open-source = free”. The pros and cons of commercial vs open-source vs appliance are debated endlessly but without any benefit. Selecting a search solution requires a careful analysis of current and anticipated requirements, and as we discussed in Research Note 01/12 the requirements for technical and business support post-implementation have to be taken into account at the outset to gain a sensible view of the total cost of operation over (say) a three year period.

The development of open source software dates back to the early 1980s and the launch of the GNU Project by Richard Stallman at MIT and some small-scale open source search applications were developed in the 1990s, mainly for web site search. In 1999 Doug Cutting released Lucene as a SourceForge project and donated the code to the Apache Foundation in 2001. Around the same time Yonik Seely developed the Solr application for CNET and then donated the code to the Apache Foundation in 2006.

Lucene itself is an indexing and search library and does not contain crawling and HTML parsing functionality. However, several projects extend Lucene's capability, such as Solr and ElasticSearch . In December 2004, Google Labs published a paper on the MapReduce algorithm, which allows very large scale computations to be run in parallel across large clusters of servers. Cutting, at that time working for Yahoo!, took the concepts in the MapReduce algorithm and created the open-source Hadoop framework that allows applications based on the MapReduce algorithm to be implemented on large server clusters.

Although there is now a high level of awareness of Apache Lucene/Solr as an open source enterprise search application there are a number of others, including Xapian and Sphinx.

There are three business/implementation models

- An organization can download the software code and use internal developers
- Use a company with expertise in the development of the selected open source application
- Implement a 'productized' version of the open source application, for example from Lucid Imagination.

Open-source applications often require the integration of modules from other sources to meet particular user requirements, and are certainly not out-of-the box applications which can be implemented without expertise not only in (usually) Java programming but also in the basic principles

of information retrieval and enterprise search optimization.

5. Big data

“Big Data” has appeared from no-where to become one of buzz-words of 2012. The Exalead definition is that a data collection is considered “Big Data” when it is so large an organization cannot effectively or affordably manage or exploit it using conventional data management tools. The size is relative rather than absolute. It is not just a ‘Big Company’ issue. Another approach at defining Big Data approaches it from the characteristics of Volume, Velocity, Variety and Variability. ‘Velocity’ takes into consideration both the rate of change of data sets and the impact that even a small data item may have on a much larger data set. ‘Variety’ is a reflection of the number of different database formats and master data management schemas that may be involved.

In the context of enterprise search there are a number of issues and opportunities arising from the publicity around Big Data.

- It is putting enterprise search much higher up the list of ‘must have’ enterprise applications as senior managers start to focus on the ability of the company to find information for probably the first time ever,
- The major IT companies see the solution of Big Data problems as a very important market opportunity, hence the acquisitions by Oracle and IBM in particular. Google has launched its Big Query web service and Amazon and Microsoft offer similar services. Autonomy has had a private cloud service for some time.
- Companies are starting to discover just how much information they have in databases, and are finding that not only are the existing tools inadequate to meet the potential demand for Big Data analysis but that they have no employees with the skills needed to develop these solutions. In the USA in particular the concept of the ‘data scientist’ is gaining ground very quickly.
- When a query is made to a SQL database it is grounded in the structure of the database, but the power of Big Data analysis is in finding patterns across multiple databases, some of which may be in the public domain and released under Open Data initiatives. As a result there are some very challenging semantic analysis issues around the extent to which data sets are homogeneous and so can be mapped against each other.

6. Unified information access

In 2008 the Forrester Group published a report on Unified Information Access, making the following observation in the introduction to the report

“Search and business intelligence (BI) really are two sides of the same coin. Enterprise search enables people to access unstructured content like documents, blog and wiki entries, and emails stored in repositories across their organizations. BI surfaces structured data in reports and dashboards. As both technologies mature, the boundary between them is beginning to blur. Search platforms are beginning to perform BI functions like data visualization and reporting, and BI vendors have begun to incorporate simple to use search experiences into their products. Information and knowledge management professionals should take advantage of this convergence, which will have the same effect from both sides: to give businesspeople better context and information for the decisions they make every day.”

Other major consulting companies, notably Sue Feldman at International Data Corporation (IDC) take a similar position. Probably the company doing more than anyone else to get UIA on the agenda of senior management groups is Attivio. The Attivio solution is based on the Apache Lucene open-source

Enterprise search - trends and developments

software but with a lot of proprietary code on top. Both CEO Ali Riaz and CTO Sid Probstein were at FAST Search and Transfer prior to its acquisition by Microsoft.

As with Big Data (and meaning-based computing from Autonomy) Unified Information Access has no concise definition but it is indicative of an increasing level of integration between text-based enterprise search, business intelligence, content analytics, text and data mining and big data applications. All the major IT vendors (and systems integrators) can sense significant market opportunities, and many smaller search vendors are looking to take advantage of the visibility of the value of information discovery and sell solutions to companies that do not have enterprise-level agreements with companies such as IBM and Oracle.

7. Search-based applications

The concept of search-based applications is just as difficult to define as Big Data and Unified Information Access. Sue Feldman at International Data Corporation offers probably the best description.

- They are built on a search backbone to enable fast access to information in multiple formats,
- Are designed as a unified work environment to support a specific task or workflow, for example, e-discovery, fraud detection, voice of the customer, sales prospecting, research, or customer support,
- Integrate all the tools that are commonly needed for that task including information access, authoring, reporting and analysis and information visualization
- Unify access to multiple repositories of information in multiple formats,
- Integrate domain knowledge to support the particular task, including industry taxonomies and vocabularies, internal processes, workflow for the task, connectors to collections of information.

Another way of looking at these applications is that it is search without a search query box. The queries are created from the working processes of the user. One simple example is the Rightmove house agency in the UK which offers users the ability to draw a complex polygon to define the area in which they are looking for a house. <http://www.rightmove.co.uk>.

What these applications take advantage of, and the same is true of unified information access, is that the technology that has been developed over many years to handle unstructured content is equally adept at not only searching structured databases but then integrating the information from both source categories at very high levels of performance.

8. Federated search

Federated search, the ability to search for information across multiple repositories and applications and then provide an integrated set of results, is a fundamental requirement of enterprise search. The usual model is for a module of the master search application to send the query to the search applications in each of the target repositories. The results from each are then integrated in some way and presented to the user. In theory it is easy, and in practice it is extremely difficult. Each of the individual search applications will have calculated a relevance ranking on the basis of the content in the repository, so normalising the results to provide a rational overall ranking is not a reliable solution. There are almost certainly going to be performance delays, especially if the repositories are located around the world, and these performance delays will be exacerbated if the repositories have different security models. Single sign-on for all applications is still rarely achieved.

Enterprise search - trends and developments

Then there is the challenge of de-duplicating content from the various repositories. There are solutions for this when a single language is being used, but the situation is much more complex with multiple languages.

Another option is to create a master index of all repositories, search the master index and then download the relevant items from each repository. This option runs into some serious index performance management challenges.

A substantial amount of research and development is being undertaken into achieving good performance from federated searching as this is a core requirement of unified information access and search-based applications. Despite the best endeavours of search vendors high-performance federated search applications providing access to a list of relevant, de-duplicated information is still some way in the future. If this is an important requirement for your organisation then always ask to see your selected vendor's implementation of federated search for another customer.

9. People and expertise search

Tracking down who knows what in an organisation remains a priority search application and yet little attention is paid to whether it is working as well as it needs to. At the Enterprise Search Europe 2012 conference there was an excellent paper from Sinequa and Atos about the development of a very powerful expertise application that makes extensive use of entity extraction across 78,000 employees world-wide that searches across cvs, projects, emails, reports and many other 'expertise' resources to try to provide a ranked list of experts in a specified field.

Several companies now provide details of leading subject experts at the top of a search results page as a form of 'personal best bets'. For certain what is not going to be sufficient is an approach based only on SharePoint MySite information.

However often the need is just to track down a person by name, and here all sorts of problems arise with the transliteration of names, the use of shortened names and name formats which are not always in given name – family name sequence. Many companies have multiple employee databases, often managed by personnel departments for the purposes of employee management and compliance and not as the basis for creating networks in increasingly social businesses.

10. Mobile search

Mobile search is going to be a very important technology in the future. As with cloud-search we are still in the early days of mobile search but already a considerable amount of innovation is evident, based on a careful analysis of what users want from search on a mobile smartphone and a tablet. In both cases users want actionable information and not long lists of search results to browse through and perhaps print out. Voice query and voice output are clearly going to be important, as the level of use of Siri on the Apple iPhone4S illustrates.

The quality of mobile search is going to be a benchmark for enterprise search applications, much as Google public web search is used as a benchmark for enterprise search no matter how inappropriate the comparison is. If employees are going to use their own devices to access enterprise content then they will use Google, Apple and Microsoft as benchmarks. Apple is now taking search very seriously with the Siri interface on the iPhone4S but and Siri is likely to be upgraded with the release of iOS6 later this year. Already Siri supports over a dozen languages. Google will also be upgrading their search capability, and given the speed of development of Google Translate the company may be looking to launch cross-language search services. The support of mobile devices through cloud-based applications

Enterprise search - trends and developments

is one of the reasons why the cloud service suppliers are majoring on search.

11. Cloud-based search

Another important development is the availability of cloud-based search-as-a-service applications. Hosted search services have been used for web sites for well over a decade but have failed to make any inroads into the enterprise sector because of concerns over security management, data protection (from companies in the EU) and customisation to meet specific requirements. Over the last year a number of companies have started to highlight the benefits of using cloud-based search services, mainly in terms of getting something started and then being able to accommodate growth without the need to switch vendor to do so. Exalead and Autonomy have already made major commitments to enterprise cloud search services.

In the last couple of months there have been important announcements from Amazon (CloudSearch) and Microsoft Azure. Azure provides access to the Lucid Imagination Lucene/Solr stack and offers four service packages

	Micro	Small	Medium	Large
Document limit	10,000	100,000	1,000,000	On application
Search queries/month	10,000	500,000	2,000,000	
Index frequency	Once/month	Once/day	Unlimited	
Index targets	Website	Website, file, DB, XML content	Website, file, DB, XML content, ODBC, connectors	
Search collection	Single	Multiple	Multiple	
Disk space	5GB	50GB	150GB	
Cost	Free	\$150 per month	\$500 per month	

Search Technologies, a major search systems integrator, has set up a demonstration of Amazon CloudSearch using Wikipedia as the content source. Doing back-to-back comparisons of searches on the Wikipedia and Search Technologies/Wikipedia demonstration is an instructive exercise.

It would seem likely that Google will also be offering cloud-based enterprise search services. Google already has a document-based pricing structure for its ESA product which would migrate easily to a cloud framework, whereas other vendors with server-based licenses are going to have to be careful about the impact of cloud pricing on their current license income.

There have been some negative comments about cloud-based search, many about the lack of security and the need to upload documents. These comments fail to take into account that cloud-based search, and indeed other cloud-based applications, are still in their infancy.

12. In conclusion

In just a year the commercial enterprise search business has almost totally restructured and open-source solutions are now offering functionality as good as, if not better, than most commercial applications. The commercial vendors, and the larger open-source search suppliers, are putting significant resources into educating the business community about the value of search, and finding IT and business managers are increasingly aware of the need to provide employees with good access to

Enterprise search - trends and developments

corporate information assets. The problem is that this will need additional financial and skills resources at a time when budgets are stretched and few companies have invested in a search support team with the skills needed to make sure that a good return is made on the search investment by both the company and its employees.

For organisations with a major commitment to SharePoint 2010 and considering the benefits of upgrading to the next release a very close look at user requirements is essential.

One of the dangers of cloud services is that IT departments will see it as a convenient way of buying search services and still fail to invest in the skills needed for a successful implementation.

Never has there been a more important time to have an information discovery strategy that takes a user-centric and task-based approach to finding quality information and data. Finding people and expertise has to be a business-critical component of this strategy.

Enterprise search - trends and developments

Resources

Findwise global search survey

<http://www.slideshare.net/mobile/findwise/results-from-the-enterprise-search-and-findability-survey>

McKinsey Global Institute report on big data

http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation

The future of FAST ESP

<http://www.enterprisearchblog.com/2011/07/the-future-of-fast-draft.html>

http://www.tnrglobal.com/index.php?option=com_content&view=article&id=166&catid=43&Itemid=215

Mobile search

<http://www.designcaffeine.com/designing-search-ux-strategies-for-ecommerce-success/>

<http://isquared.wordpress.com/2012/04/03/the-information-needs-of-mobile-searchers/>

There is a section on enterprise search on the Intranet Focus Ltd web site, listing reports, books, blogs and a list of search vendors

<http://www.intranetfocus.com/enterprise-search>

The key source of news and analysis of enterprise search trends and developments is the Beyond Search blog from Stephen Arnold

<http://arnoldit.com/wordpress/>

There is a LinkedIn Group for Enterprise Search Engine Professionals. This is a closed group and membership is by application

http://www.linkedin.com/groups?gid=161594&trk=hb_side_g

The Enterprise Search Summit Fall takes place in Washington DC on 17-19 October 2012

<http://www.enterprisearchsummit.com/Fall2012/>