



European  
Commission

# JRC SCIENTIFIC AND POLICY REPORTS

## Enterprise Search in the European Union: A Techno-economic Analysis

Authors: Martin White, Stavri G Nikolov

Editors: Shara Monteleone, Ramon Compañó,  
Ioannis Maghiros

2013



Report EUR 26000 EN

Joint  
Research  
Centre

European Commission  
Joint Research Centre  
Institute for Prospective Technological Studies

Contact information

Address: Edificio Expo. c/ Inca Garcilaso, 3. E-41092 Seville (Spain)

E-mail: [jrc-ipts-secretariat@ec.europa.eu](mailto:jrc-ipts-secretariat@ec.europa.eu)

Tel.: +34 954488318

Fax: +34 954488300

<http://ipts.jrc.ec.europa.eu>

<http://www.jrc.ec.europa.eu>

This publication is a Scientific and Policy Report by the Joint Research Centre of the European Commission.

Legal Notice

Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of this publication.

Europe Direct is a service to help you find answers to your questions about the European Union  
Freephone number (\*): 00 800 6 7 8 9 10 11

(\*) Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed.

A great deal of additional information on the European Union is available on the Internet.  
It can be accessed through the Europa server <http://europa.eu/>.

JRC78202

EUR 26000 EN

ISBN 978-92-79-30493-4 (pdf)

ISSN 1831-9424 (online)

doi:10.2791/17809

Luxembourg: Publications Office of the European Union, 2013

© European Union, 2013

Reproduction is authorised provided the source is acknowledged.

Printed in Spain

## **Preface**

This report contributes to the work being carried out by IPTS on the potential of Search, providing a techno-economic analysis of Enterprise Search in the EU and a discussion of the main challenges and opportunities related to the current state of the Enterprise Search market in Europe. This study is part of CHORUS+ - an initiative supported by the Directorate General Information Society and Media. Information about CHORUS+ and its related activities is available at <http://avmediasearch.eu>



## Table of Contents

<b>Preface</b> .....	<b>1</b>
<b>Executive Summary</b> .....	<b>5</b>
<b>Methodology</b> .....	<b>11</b>
<b>Part I: Managing Enterprise Information</b> .....	<b>13</b>
1.1 The enterprise repository.....	13
1.2 Reasons for complexity of ES repository.....	15
1.3 The technology of enterprise search.....	17
<b>Part 2: Market Considerations</b> .....	<b>25</b>
2.1 The value chain for enterprise search.....	25
2.2 The enterprise search business structure.....	27
2.3 The EU market for enterprise search.....	32
2.4 Making a business case for enterprise search .....	35
<b>Part 3: The Choice of Enterprise Search Solutions</b> .....	<b>43</b>
3.1 Selecting and implementing enterprise search applications.....	43
3.2 Search implementation and user satisfaction.....	48
3.3 Enterprise search skills availability.....	51
<b>Part 4: Analysis and Policy Considerations</b> .....	<b>53</b>
4.1 Technology assessment and forecast.....	53
4.2 SWOT analysis .....	61
4.3 Opportunities for EC support actions. Some policy briefs.....	62
<b>Appendix A: List of enterprise search vendors – alphabetical</b> .....	<b>71</b>
<b>Appendix B: Corporate profiles of selected enterprise search vendors</b> .....	<b>75</b>
<b>Appendix C: Delphi summary tables</b> .....	<b>79</b>
<b>Appendix D: Enterprise search industry analysis consultancies</b> .....	<b>83</b>
<b>Appendix E: Workshop "Exploring the future of enterprise search"</b> .....	<b>85</b>
<b>References</b> .....	<b>87</b>



## Executive Summary

### ***The value of enterprise search***

The term 'enterprise search' (ES) is used as a generic description for information retrieval applications that use a range of different core technologies to search enterprise repositories. For the purpose of this report, it includes the search of organisations' external web sites, intranets and other electronic text held by the organisations in the form of email, database records, and documents on file shares. This is often referred to as 'unstructured' information.

Enterprise search technologies date back to the late 1960s when they were developed to search large online databases of scientific, commercial and legal information and to support the legal teams working on a number of large anti-trust suits in the USA – the breakup of AT&T being one example.

There are three main technical approaches to ES: Boolean, vector space and probabilistic. Though there are some differences between the requirements of searching web sites and searching other enterprise applications, primarily around security management, it is possible to use the same enterprise search application for both purposes.

The development of enterprise search applications requires a wide range of specialised skills, in particular mathematical approaches to set theory, probability and computational linguistics.

Enterprise repositories of unstructured information are growing rapidly because of the widespread adoption of social media, increased compliance and regulatory requirements and a lack of resources to remove redundant information. According to research in the USA, large companies (i.e. with more than 1,000 employees) have accumulated over 100 terabytes of information, and many have more than 1 petabyte. Surveys indicate that senior managers are aware of the importance of unstructured information but few are taking action to provide employees with adequate tools to access this information.

### ***Motivators***

Motivators for the development of an enterprise search market, as emerged from the surveys mentioned in this report and also from the workshop organized by JRC-IPTS, "Exploring the future of Enterprise Search", in Seville in October 2011 are:

- There is increasing information everywhere: more than 200 billion emails per day; 80% of enterprise information is unstructured.
- Digital data growth is enormous: it is expected to be 35 zettabytes in 10 years' time. In particular, it seems that 94% of organizations are collecting and managing more business data than just a few years ago and business information collected/managed has increased by 86% in the last few years.<sup>1</sup>
- The cost of poor data management: organizations are seemingly losing revenue each year (on average, 14%) as a result of not being able to fully leverage the information they collect. That translates to circa \$130 million in lost opportunity each year for a \$1 billion organization.<sup>2</sup>
- Legal compliance of the enterprise: obligation to store and find all enterprise documents, business communications for legal reasons. Enterprise data is all over the place. ES has to federate all the information existing in both structured data (databases) and unstructured data (text, reports, mail).

---

<sup>1</sup> Source: Oracle Survey, *From Overload to Impact: an industry scorecard on big data business challenges*, 2012.

<sup>2</sup> *Ib.*

In other words, if one reason for adopting ES is the growth in data generation, a more worrying reason is the fact that this huge amount of information is largely unstructured. It is estimated that about 80% of the information stored is either unstructured or has no adequate metadata for the needs of employees.

As noted by Findwise in its recent Enterprise Search and Findability Survey (2012), quick access to information is of strategic importance in the Information Economy: "The fault does not lie with the technology; rather, most organisations have failed to take a strategic approach to enterprise findability". Accordingly, many companies implement search in some way, but few of them do it with the thorough thought needed for the solution to leverage the expected value or the expected "return on information".

### ***The enterprise search business***

The enterprise search business has a **complex value chain**. Some enterprise search vendors provide direct support to customers for implementation and development. Others work either through large systems integrators or through companies specialising in enterprise search implementation.

In the past, commercial **proprietary** products for ES dominated the market but there is now also an increasing interest in the use of **open-source software**. Although search vendors have significant skills in software development, supported by expertise in computational linguistics, applied mathematics and enterprise architecture management, they also make use of specialised modules from other companies, such as document filters, entity extraction applications and stemmers. There is also a substantial OEM business, in which search applications are included in applications such as business intelligence, document management and customer relationship management applications.

Although the issue is acknowledged to be of extreme importance, only a small number of companies benefit from provision of dedicated enterprise search technologies and products. The global enterprise search business probably has no more than 200 companies. There are around **60 enterprise search vendors** who together account for probably over 90% of enterprise search software sales, excluding open source products where there is no license fee. **Six** vendors, all of them multinational IT companies, have a major impact on the development of search technology but only a limited impact on the development of the enterprise search market through promotional activities. These can be termed Type 1 vendors, and are Google, HP, Lexmark, Microsoft, IBM and Oracle. **Type 1 companies** in general sell enterprise search as part of an overall enterprise application suite, and not as stand-alone products. The long-term implications of the acquisition of Autonomy by HP in 2011 will not be clear for some time.

The other companies form a **large Type 2 category**. The primary characteristic of Type 2 companies is that they have developed proprietary search software and are mostly funded by venture capital and private equity investments. Because of their small size and the fact that they are private companies, there are no requirements to publish detailed accounts of revenues and costs. Most of these companies have revenues of less than \$20 million. Examples of Type 2 companies with headquarters in the EU are Fabasoft (Austria) and Sinquea (France).

**Type 3 companies** build products around open-source software such as Lucene/Solr, with Intrafind (Germany) being an example.

There are **no comprehensive lists** of search vendors operating in the EU market. As a result, when procuring an enterprise search application, the primary source of information used by IT managers is the web sites of vendors. Although there are some industry analysts (mostly based in the USA), in the reports and services they provide most do not list more than a core group of Type 1 and larger Type 2 vendors. The most comprehensive list of ES vendors is on Wikipedia but even this has a number of omissions.



### ***Specifying and selecting search software***

One main barrier to making a business case is a **lack of awareness of the functionality** of enterprise search applications and the benefits that effective search can have for the enterprise. Enterprise search is of potential value to most, if not all, employees, but no single department wishes to take responsibility for making a business case. There is evidence that implementing enterprise search is not a high priority; the reason being that one single business unit is probably unable to make a business case for enterprise search. Organisations have not usually done any research into the most important tasks carried out by employees and the extent to which enterprise search would improve operational effectiveness.

IT departments have a role to play in the technical evaluation of enterprise search applications and in the initial installation, but it is important that **business requirements** are well defined in developing a business case for selecting or replacing an enterprise search application. IT managers, however, may have little formal teaching about information retrieval technologies and may not be aware of how to evaluate search applications or how to plan adequately for implementation and subsequent optimisation.

### ***The EU market for enterprise search***

It is very difficult to assess the size of the EU market for enterprise search in terms of revenues. There is very little published information even on the global market for enterprise search and none that provides an analysis of the EU market. However, we can assume that the market potential is still considerable, particularly for mid-sized companies that could benefit from using ES solutions.

Among the leading vendors of enterprise search software are IBM (Omnifind and Vivisimo), Oracle (Oracle Secure Search and Oracle Endeca), Google (Enterprise Search Appliance) and Microsoft. The largest EU search software company is Exalead, which was acquired by Dassault Systems in 2010 and Autonomy, which was acquired by HP in 2011.

In particular, from the limited amount of information it is possible to infer that:

- The total revenues of EU-headquartered search vendors are between €100 million and €200 million depending on the view taken of Autonomy search revenues.
- Sales by IBM, Oracle, Google and Microsoft in the EU could be of the order of €300 million. This would give a total market in the EU of at most €500 million, or 25% of the global market.
- There is no dominant supplier of search software in the EU

It is equally difficult to establish an installed base figure for enterprise search in the EU, for three main reasons:

- Many large multinational EU companies may be using enterprise search applications that have been purchased in the USA or another country outside the EU;
- Only a few companies disclose, even in broad terms, the installed base of their software;
- Many companies may be using enterprise search applications embedded in enterprise application suites such as Microsoft SharePoint and Oracle Web Logic/Web Center.

For the purposes of making an estimate of the market potential it may be reasonable to assume that a company with more than 500 employees would have a reasonable business case for an enterprise search application. In the EU, no information is published on the number of companies by size other than those with 250 or fewer employees. In the USA, there are 18,000 companies with 500 or more employees out of a total of 5 million. In Europe, there are 21 million companies of all sizes, but scaling up number of US businesses with 500 or more employees to 72,000 would not be a realistic comparison. For the purposes of this analysis, the assumption is made that there are 20,000 companies in the EU with 500 or more employees, to provide a very conservative market potential.

To this corporate market potential, central and local government and agencies, hospitals, public services and charities and the academic sector have to be added.

Despite the general consensus amongst experts that ES has strong potential, the unexpectedly poor market data seem to indicate that there are barriers to rendering enterprise search mainstream.

Nonetheless, companies would increase their ability to manage (i.e. to discover, reuse, modify, extract and combine) all kinds of information assets rendering them more efficient and thus more competitive.

### ***Achieving high enterprise search performance: challenges***

A number of **challenges** have been identified that need to be addressed. For instance, efforts to meet changing business requirements, the lack of support post-implementation, or the lack of a search support team. With most enterprise applications (such as financial applications) the development work is carried out before implementation. In general these applications are replacements or upgrades of existing applications and there will be a good body of knowledge about the applications in both the IT and relevant business departments in the organisation. This is not the case with enterprise search, where the installation can be carried out quite quickly, but afterwards there is an on-going requirement to adapt the application to meet changing business requirements, such as the acquisition of a company or investment into a new business area.

In a recent survey, 60% of organisations reported that search is business-critical to them, but in only 10% of organisations are very satisfied with the performance of the in-house search application. One possible reason for this is that only 20% of the organisations have more than one person supporting the search application. When users do complain about the quality of search there is anecdotal evidence that the decision is made to 'upgrade' the search application, on the basis that clearly the current search implementation is not adequate. Since the underlying issue is one of a lack of support post-implementation the results from the replacement search application are usually no better.

Within the organisation there will be a requirement not only for IT support to ensure that the hardware and software applications are working to agreed technical performance standards, but also for a search support team that has a combination of business and technical skills in order to monitor the use of the application and to ensure that it meets changes in business requirements.

For a large enterprise search implementation there are a number of roles that need to be filled. These are a Search Manager, a Search Technology Manager, an Information Specialist with a strong business background, a Search Analytics Manager and a Search Support Manager. For all these posts the team need to have a good background in the technology and implementation of enterprise search. However, the EU seems to have a significant lack of academic institutions that are offering taught courses in information retrieval. There are around 30 institutions in the EU undertaking research into enterprise search applications, but there are no full-time three year undergraduate courses. As a result there is a **shortage of skilled professionals** to join search vendors as development and implementation engineers, and to join enterprise search support teams.

### ***Information retrieval and enterprise search***

There are over thirty institutes in the EU undertaking research into information retrieval. Few of these teach courses in information retrieval and the coverage of enterprise search development and implementation is very low. As a result it seems that search vendors are finding it difficult to recruit development and support professionals with a good knowledge of the fundamental principles of search and companies are not able to recruit staff with good search implementation expertise for search support teams. Although the EU supports a wide range of information retrieval research projects the extent to which these can then be incorporated into commercial or open source search applications may be quite limited.

## **Technology forecast**

The outcomes of the Delphi study, carried out by IPTS in collaboration with Intranet Focus Ltd in 2011, indicates **six important areas** of technical development for enterprise search over the next few years, namely

- Integrated search of structured and unstructured content;
- Search as an integration platform (unified access platforms);
- Search incorporated into business intelligence applications;
- Search-based applications;
- Text mining, and
- Enterprise mobile search applications.

All the above are evolutions of current search technologies and products. Because most of the Type 2 independent search vendors are quite small businesses, they face a difficult challenge in balancing the investment in solution development with the difficulty of convincing prospective customers that these developments will make a significant difference to operational performance. All the Type 1 companies have substantial global research and development activities.

### ***Future trends in ES: cloud-based and user-demand approach, open data models, interoperability***

According to our Delphi-type study and to the results of the workshop on Enterprise Search (organized in Seville in October 2011), the technologies that are regarded as most important to the demand and adoption of ES are **Search-Based Applications (SBA)** and **integrated search platform (unified access platform)**. Integrated platform and search-based applications are key solutions because they promise to provide semantic linking (combining structured and unstructured data) and semantic search (allowing intelligent analysis of query).

SBA can rely on faceted search (facets are visualizations of semantics that users can understand) on semantic databases and on search engines. Given that search engines can handle the semantics of databases, and that facets allow for business intelligence-type reporting, SBA can use the power of search engines (intuitive, scaling, agility) to extract and merge information from databases and text.

Despite the need to improve technological building blocks, it is not the absence of effective technologies that prevents ES from developing as a market. Stakeholders and experts do not consider a lack of specific ES technology to be the limiting factor for the take-up of enterprise search.

This does not mean that there are no technological challenges. On the contrary, improvements in re-use of components or better interoperability are examples of missing elements that would be beneficial for the development in the short to medium term. In the long term, the growth of semantic web applications and of the 'Internet of things' will impact on ES, both in terms of new applications and in terms of innovative business models to be developed.

Also, **visual enterprise search**, though still quite limited and applied for specific scopes, is increasingly demanded, as enterprises have more and more multimedia assets.

A common future trend for ES we identified is represented by "domain-specific semantics", i.e., the development of intelligent query interpretation (intelligent analysis of query to extract relevant terms).

The **mobile search** landscape has changed as well over the last few years. Nowadays, tablet devices play an important role in search as well. This type of device has had an impact on mobile

search and search in general. Innovations that originate from mobile search are being adopted in traditional search, so this aspect should also be taken into account for future development in ES.

Finally, a possible future evolution of ES could be towards the application of cross-sessions search models in which there is increasing interest. The assumption underlying these models is that while some simple needs can be satisfied with a single query, others require a series of queries over the long term, depending on the task users are trying to accomplish. In other words, users may be interested in retrieving and grouping search results in the future, when they have to do with cross-session search tasks (e.g. a vacation planning task). Searchers behaviours extending over multiple search sessions are analyzed so that knowledge of previous queries on the same long-term task enables the search tool to support the user in his task.<sup>3</sup>

### **To sum up:**

- There is no dominant supplier of search applications in the EU.
- Although there are no reliable data nor are there any revenue analyses for the EU enterprise search market, we can assume that the market potential is still considerable, particularly for medium-sized companies that could benefit from using ES solutions.
- The total revenues of EU-headquartered search vendors are between €100 million and €200 million.
- Some of the main challenges that the ES market has to face are: how to define a stable and open architecture over which a European ecosystem could develop and mature, how component technology testing could be performed (see benchmark section), what would facilitate technology transfer to industry (see tech transfer section), or how to remedy the shortage of skilled professionals.
- A general issue to be addressed in the future is the fact that boundaries between enterprise search, text and data mining, business intelligence and content analytics are becoming very blurred.<sup>4</sup>
- To our knowledge, this report provides the most detailed and comprehensive techno-economic analysis of the enterprise search market in the EU and includes an up-to-date list of ES vendors.

---

<sup>3</sup> A. Kotov, P. N. Bennet, R.W. White, S.T.Dumais and J. Teevan, *Modelling and Analysis of Cross-Session Search Tasks*, In Proceedings of the 34th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval ([SIGIR 2011](#)), Beijing, China, July 2011

<sup>4</sup> See the Report by PWC, *Technology forecast 2012*, pp 45-53, available at <http://www.pwc.com/us/en/technology-forecast/2012/issue1/index.jhtml>

## Methodology

The research for this report is based upon a Delphi-like exercise, which was carried out in two phases.

- In the first phase, a questionnaire was e-mailed to 30 different enterprise search companies. Responses were collected over the period from July to September 2011. The questionnaire was designed by IPTS in collaboration with Martin White, Intranet Focus Ltd and sent to search vendors, integrators and industry analysts in Europe, US and Canada. A total of 18 responses were received. Most of the companies that replied were invited to discuss the results of the questionnaire at the expert workshop 'Exploring the future of Enterprise Search', held in Seville.
- In the second phase, the questionnaire results were presented at the 'Exploring the future of Enterprise Search' expert workshop, organized as part of the CHORUS+ project by IPTS, in Seville, on 13-14 of October 2011. At this workshop, the techno-economic and socio-economic trends that currently characterize the enterprise search solutions market were discussed. This discussion of the questionnaire results was summarized in a SWOT analysis, which has contributed to the results of this report.
- The Delphi-like survey was not the only element of this study: a substantial amount of desk research has been undertaken, search industry conferences in the EU and the USA have been attended, and search vendors, integrators, academics and organisations with experience of implementing search applications have been interviewed. Although the research was completed in late 2011, significant events related to ES companies occurred during the publication process that required an update of the report and therefore postponement of its publication.



## Part I: Managing enterprise information

### 1.1 The enterprise repository

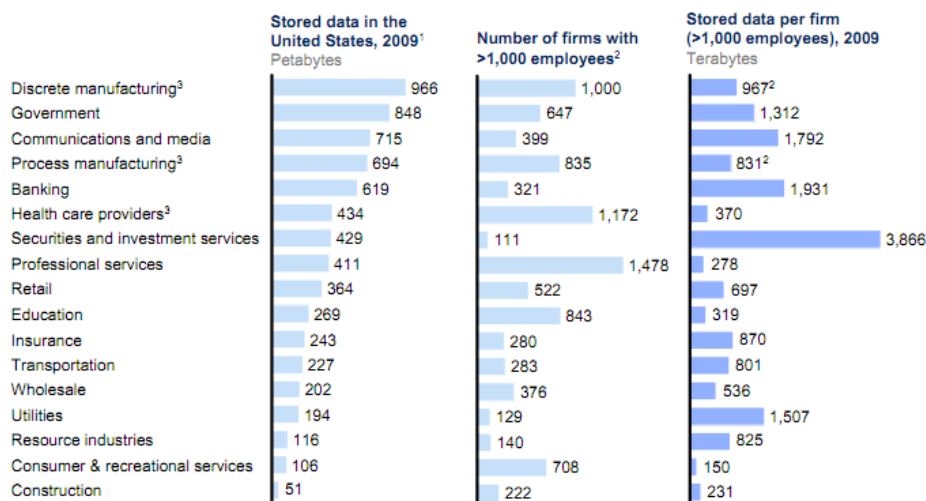
Marshall McLuhan has noted that the one thing that fish are totally unaware about is water, as they have no anti-environment that would enable them to perceive the element they swim in (1). They take it for granted, and business and society do just the same with information. The move towards a service economy in the EU over the last decade, with manufacturing increasingly being outsourced, has also been a move towards an information economy. Especially in the financial sector very important decisions are made on the basis that the information available represents all the information that an organisation holds, and that the quality of the information is flawless. However it could be argued that the financial crisis of 2008 was a failure to manage information, as banks were not aware of the poor quality of the loans they had made.

All **organisations track** the extent of their assets, including their financial situation, the number of employees, equipment used in offices and factories, intellectual assets (patents, trademarks and licenses) and orders and deliveries. This information is required for regulatory reasons and to enable the Board of Directors to manage the organisation effectively.

Very few organisations have any measure of the amount of information they hold or the quality of the information. This is especially the case with what is often referred to as '**unstructured information**', such as information stored in document management, customer relationship management and records management applications and in shared drives. In reality there is a considerable amount of structure to this information, such as a title, author, date, and headings to sections and sub-sections.

In a report published in 2011 the McKinsey Global Institute calculated the scale of enterprise repositories in a number of vertical industry sectors in the USA (2).

#### Companies in all sectors have at least 100 terabytes of stored data in the United States; many have more than 1 petabyte



1 Storage data by sector derived from IDC.

2 Firm data split into sectors, when needed, using employment

3 The particularly large number of firms in manufacturing and health care provider sectors make the available storage per company much smaller.

SOURCE: IDC; US Bureau of Labor Statistics; McKinsey Global Institute analysis

It is important to appreciate that 1 terabyte is equivalent to around 200,000 documents each 500k in size. This should only be taken as a very indicative estimate, but for an employee seeking to find a document that contains a specific piece of information that is business-critical they need to be able to find relevant documents amongst the millions that a large company may hold going back many years.

Although this data is for US companies it would be a reasonable assessment that companies in the EU were in a similar situation.

The **value of enterprise search** is in enabling the employee to find this information in ideally no more than two minutes (3) However finding a specific piece of information on which to base a business decision is becoming a very challenging. The US IT consulting company International Data Corporation estimates the time spent searching for information averages 8.8 hours per week and analyzing information consumed an additional 8.1 hours. (4) On that basis almost half the working week is involved in finding and managing information.

In 1999, Don Marchand, in his seminal book "Competing with Information" (5), showed information to create substantial business value in four key areas:

- *Minimising risks* arising from changes in markets, new financial and legal requirements, the actions of competitors and disruptions to supply and distribution channels;
- *Reducing costs* by simplifying business processes through the effective use of information;
- *Adding value* to customers by being able to analyse the wealth of information that the business accumulates from the transactions with customers, and by the feedback they give on product and/or service quality; and
- *Creating new reality* by developing new and enhanced products and services that meet emerging user requirements as quickly as possible, and certainly before competitors can take advantage of the market opportunity.

In 2008 CapGemini studied the impact of good information management on FTSE350 companies (6). Over 80% of survey respondents acknowledged information exploitation to be a critical driver or determinant of business performance and both a personal and organisational priority for the vast majority.

The senior executives surveyed saw information quality and information systems as the biggest barriers; other barriers including ineffective policies and procedures, a lack of staff skills and training, the user culture and business processes. Many respondents felt that badly informed decisions affected the bottom line. For example, the likelihood of increased operational costs were cited by 47% of respondents, followed by financial losses (43%), loss of customers (38%), and lack of confidence (35%) as likely outcomes.

**Software** applications that can search through large repositories of text information dates back to the 1960s, but the level of **adoption** of enterprise search is still quite **low**. Research by International Data Corporation indicates that global sales of enterprise search applications in 2010 were around \$2 billion.(7) For comparison the forecast by Forrester for the 2010 global market for enterprise resource planning (ERP) was \$43 billion.(8) Although most large multi-national businesses have implemented enterprise search many have not yet decided to do so. Indeed in the IBM Global CIO Survey 2009, based on 2,500 individual interviews with CIOs in high-growth and low-growth companies only 67% of CIOs in high-growth companies felt that their employees had access to the information they needed. This value dropped to 51% in the case of low-growth companies. (9)

What might be termed the '**information management paradox**' is visible in a survey undertaken in 2011 by MarkLogic, a provider of unstructured database management applications. (10) The survey was conducted among nearly 500 data managers in the USA. 57% of the survey respondents said unstructured information was either extremely important or very important to their businesses but only 17% regarded themselves as highly aware of the information they held.

As with the McKinsey analysis, the Mark Logic surveys relate only to the situation in the USA, but it would be reasonable to assume that if the surveys were repeated in the EU that a similar pattern would emerge.



In 2012 the first global survey of enterprise search adoption and implementation was undertaken by Findwise<sup>5</sup>. 60% of the respondents considered that it was either very hard or moderately hard to find the right information. Only 11% stated that it is fairly easy to search for information and as few as 3% considered it very easy to find information. The ease of finding information clearly has a connection with the size of the organization. For organizations with less than 1000 employees 31% of the respondents feel that it is moderately/very hard to find the right information, while the corresponding percentage for organizations with 1001 or more employees is 77%.

For the period to 2015, 52% said that the amount of information in the organization will significantly increase. Findwise suggests that there are or/and will be inefficiency in organizations, as a majority of the respondents (75%) said that the findability of relevant information is critical success factor for their business. The survey data clearly shows that satisfaction with enterprise search is low. The survey also shows that the budgets for Enterprise Search are low or even non-existent, and that the majority of the organisations in the survey do not have even one full time member of staff supporting the search application. Only 14% of the respondents have a search strategy in place.

## **1.2 Reasons for complexity of ES repository**

Some of the reasons for the growth in the scale and complexity of enterprise information repositories include:

### ***The employee as publisher***

Employees spend a substantial part of their working day creating text information not only through office productivity applications but also through email and through social media

### ***Compliance and regulatory requirements increasing***

The financial crisis in 2008 showed up inadequacies in compliance and regulatory control and companies are now very concerned that they will not be able to find important documents in response to enquiries from industry regulators

### ***Growth and complexity of individual enterprise repositories***

Organisations now make increasing use of streaming media for internal and external communications, and need to be able to find architectural and engineering drawings, photographs and other image-rich documents, and video and sound recordings. At present these are tagged with text metadata descriptors, and outside of specialised applications in law enforcement and media little attention has been paid by enterprises to search rich-media through image analysis technologies.

### ***Multiplicity of enterprise repositories***

Organisations would like having a single search solution which provides access to multiple repositories, including structured databases. This requirement emerges both from a need to reduce search complexity for the user and the costs of supporting multiple applications.

The situation is also made more complex by organisations not adopting and implementing good practices in removing redundant information. The concept of information lifecycle management emerged during the late 1990s. For the purposes of business records, there are five phases identified as being part of the lifecycle continuum. These include: Creation and Receipt; Distribution; Use; Maintenance; Disposition.

The result is that when a search is undertaken by an employee a significant amount of information will be retrieved, which is relevant in terms of the subject but may not be regarded as good practice by the organisation. It is also quite likely that older information may not be in a file format that is

---

<sup>5</sup> The results are available at <http://www.slideshare.net/findwise/enterprise-search-and-findability-survey-2013> and [www.findwise.com](http://www.findwise.com).

currently supported, and so enterprise search applications have to be able to 'translate' content in these file formats into a format that can be read by the employee.

Considering the concept of 'enterprise search', it is a challenging one for enterprise search vendors and for users of the applications they offer. It is common to consider an 'enterprise' as a commercial business, but in the context of enterprise search it could be a government department, a university, a charity or a non-governmental organisation. An enterprise can also be of any size. Even an organisation of 250 employees will have substantial amounts of information in email folders, shared drives and administrative applications such as finance and personnel.

The term 'enterprise search' is commonly used as a generic description for information retrieval applications that use a range of different core technologies to search enterprise repositories. For the purposes of this report the definition of enterprise search proposed by Hawking is used: (11)

- search of the organisation's external website;
- search of the organisation's internal websites (intranet);
- search of other electronic text held by the organisation in the form of email, database records, documents on file shares and the like.

Overall, this is a market opportunity and a sector of the IT business that has not been researched in depth. The enterprise search industry is tracked by a number of IT market analysis companies (See Appendix) but the focus of these companies is on the software and the technical direction of the company rather than on market size analysis.

A further difficulty in establishing reliable and comparable information on this sector is that many of the companies are privately held and do not disclose any information about revenues and profits. Larger publicly held companies do disclose revenue and profit information but not at a level of individual products such as their enterprise search offerings.

### **Summary**

1. The rate of growth of enterprise repositories of unstructured information is increasing very rapidly because of the growth of social media, increased compliance, regulatory requirements and a lack of resources to remove redundant information.
2. Based on research in the USA large companies (i.e. with over 1000 employees) have accumulated over 100 terabytes of information, and many have more than a 1 petabyte
3. Enterprise search includes the search of the organisation's external web site, intranets and other electronic text held by the organisation in the form of email, database records, and documents on shared drives. This is often referred to as 'unstructured' information
4. Senior managers are aware of the importance of unstructured information but very few are taking action to improve the ability of employees to access this information.

### **Resources**

1. Marshall McLuhan and Quentin Fiore, War and Peace in the Global Village, Gingko Press 1968
2. Big Data: The Next Frontier for Innovation, Competition and Performance. McKinsey Global Institute, 2011. [http://www.mckinsey.com/mgi/publications/big\\_data/](http://www.mckinsey.com/mgi/publications/big_data/)
3. Mind the Search Gap. SmartLogic, 2011. <http://www.smartlogic.com>
4. [http://www.rioh.es/Images/IDC\\_Executive\\_Insights\\_January2011\\_t\\_69-4420.pdf](http://www.rioh.es/Images/IDC_Executive_Insights_January2011_t_69-4420.pdf)
5. Donald Marchand et al. Competing With Information. 1999. John Wiley and Sons.

6. The Information Opportunity Report. CapGemini. [http://www.uk.capgemini.com/insights-and-resources/by-publication/the\\_information\\_opportunity\\_report\\_2008/](http://www.uk.capgemini.com/insights-and-resources/by-publication/the_information_opportunity_report_2008/)
7. Market data provided by International Data Corporation
8. ERP Market To Grow To \$50.3 Billion. <http://enterpriseapplications.cbronline.com/news/erp-market-to-grow-to-503bn-in-2015-forrester-060511>
9. IBM Global Chief Information Officer Study 2009 [http://www-935.ibm.com/services/uk/cio/pdf/implications\\_cie03063-usen-02.pdf](http://www-935.ibm.com/services/uk/cio/pdf/implications_cie03063-usen-02.pdf)
10. Big Data 2.0: The Post-Relational Reality Sets In. Mark Logic <http://www.marklogic.com/news-and-events/press-releases/2011/marklogic-webcast-explores-the-post-relational-big-data-reality.html>
11. David Hawking. Enterprise Search in Ricardo Baeza-Yates and Berthier Ribeiro-Neto, Modern Information Retrieval. Chapter 15. Second Edition. Addison Wesley. 2011

### 1.3 The technology of enterprise search

The 'index text files and find relevant documents in response to a query' functionality dates back to the work of Engelbart and Bourne at the Stanford Research Institute in 1963. (1) This early work was focused on finding scientific and engineering research articles from major secondary services such as Engineering Index to meet the needs to researchers working on the NASA space programmes. In this period IBM was actively investing in information retrieval and random access disk storage technologies, two technologies considered essential for high speed retrieval of information from text sources.

By the late 1960s most of the basic concepts of information retrieval had been established<sup>6</sup>.

#### **Structured and unstructured information**

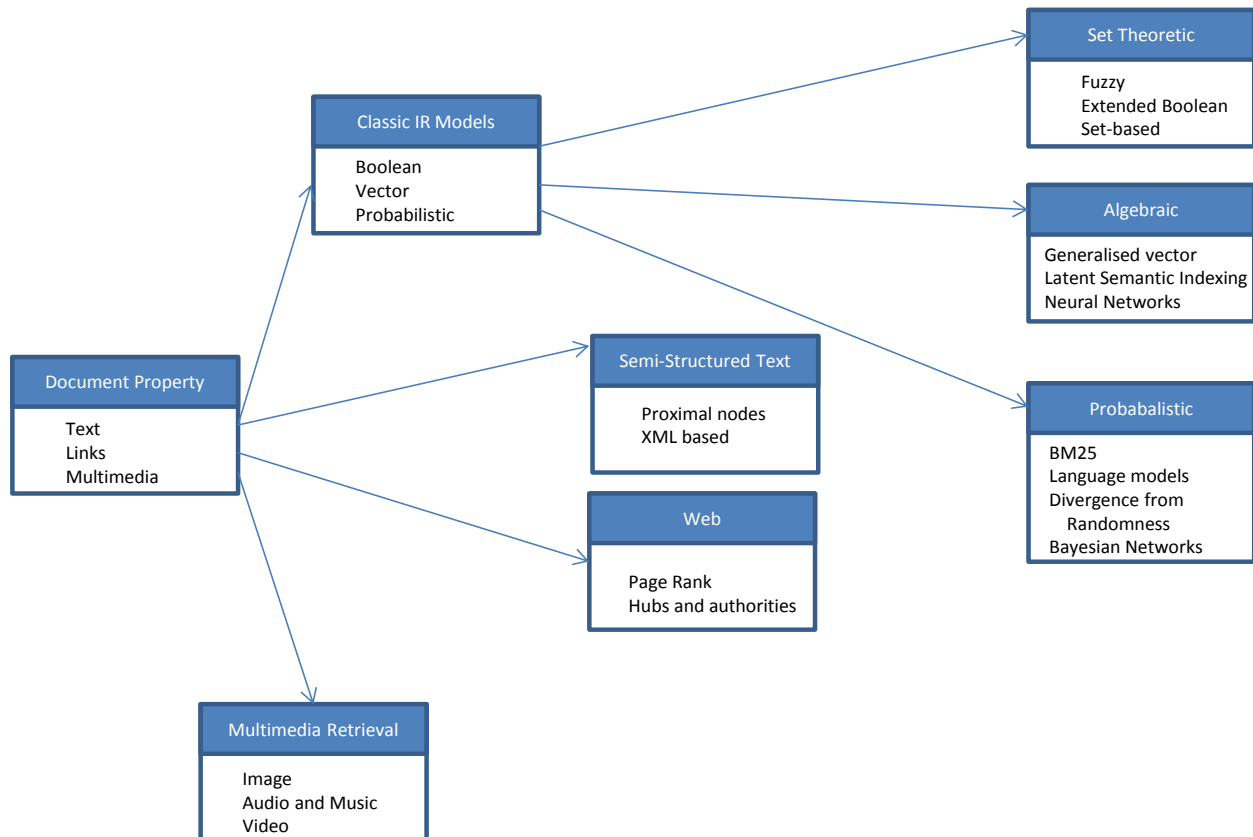
There are some fundamental differences between searching unstructured text-based content and databases containing numeric information and short text information in defined fields, such as a personal name, address or product name. These include

Unstructured content	Structured content
The query can be a set of words in any order, often including the ability to present the query in natural language	Only defined queries can be used which match the syntax and structure of the database
Because of the very close similarity to searching the internet very little, if any training, is required to use the search application	Usually considerable training is required to ensure that the correct syntax is used in a query
A query will return both exact matches and also information that is not an exact match but which may be of relevance	Only returns exact matches to a query
A query will identify and display the content item (a report or document) so that the user can assess the context of the information and make a judgement on its credibility and value	Only defined reports are presented, usually in a tabular format that matches the syntax of the database
Additional repositories can be added to the search scope with only limited development	Searches cannot be extended across databases with different syntax structures

<sup>6</sup> In 1969 the US Department of Justice filed an anti-trust suit against IBM, and to be able to respond to this suit IBM accelerated the development of the ACQARIUS online retrieval application, which in 1970 was renamed STAIRS. One of the many innovations in STAIRS was the way in which the relevance of a search result could be calculated from the occurrence of terms in the source record and their proximity to each other. STAIRS remained an IBM product until 1994.

A very important development over the last few years has been the extension of search technology developed for unstructured information to be used to search structured databases and then integrate the results. This is discussed in more detail in Chapter 5.1.

This schematic shows a taxonomy of information retrieval models and illustrates the range of technical solutions that are available (2)



The three core enterprise search technologies (the Classic IR Models in this schematic) can be summarised as follows:

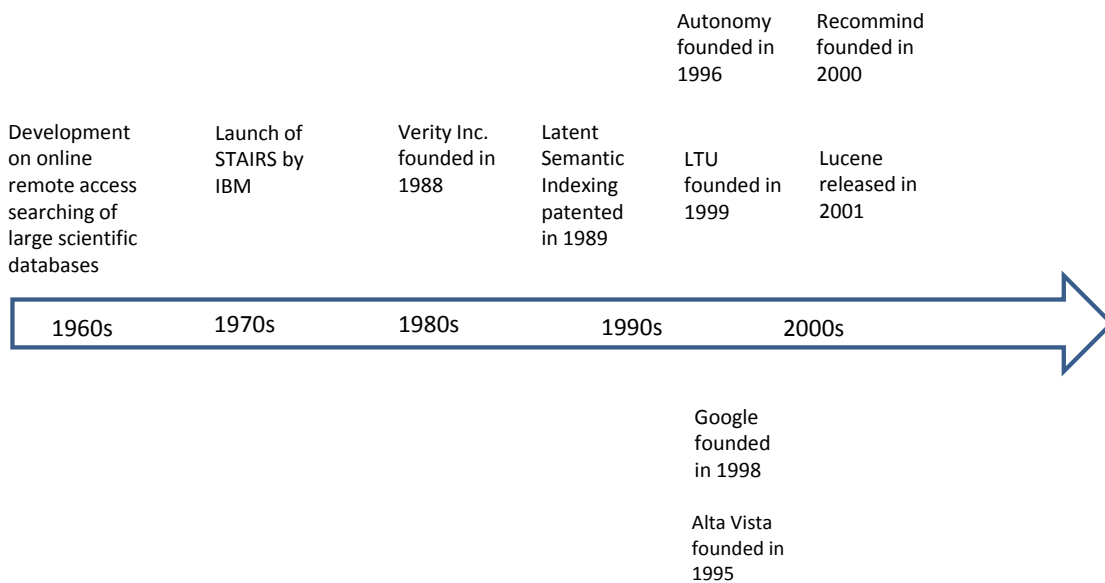
- **Boolean search** involves searching through an index of all words and their proximity to other words using the Boolean operators AND, OR and NOT. This requires relatively little processing and so searches can be carried out very quickly. Boolean searching also scales fairly well. However this approach is not effective at dealing with words which have multiple meaning (the use of 'gas' for petrol in American English) and words which are different but refer to the same concept (car and automobile). Boolean logic does not allow for partial matches. Rather like a query against a relational database a document is either relevant or is not and at its basic level all terms have an equal weight, which means that results cannot be listed by relevance. A significant amount of development has gone into various approaches to term weighting. Another important factor is document length, as a long document may meet the search criteria just because of its length. As a result document length normalisation is an important step but this has to be executed without a loss of relevance.
- **Vector space** model search engines were developed by Gerard Salton in the early 1960s to overcome the problems of Boolean search that were becoming very evident through the rapid development of online scientific database services such as Dialog and SDC Orbit. Vector space models convert text into numeric vectors and matrices and then use matrix

analysis techniques to make 'connections' between documents in the collections being searched. This is achieved by assigning non-binary weights to index terms so that a degree of similarity between two documents can be calculated.

- In **probabilistic search models** documents are ranked by the probabilistic odds of relevance, the ratio of the probability that the document is relevant divided by the probability that it is not relevant. One approach is based on Bayesian statistical inference. This has been developed by Autonomy in its IDOL inference engine which was launched in 1996.

As the schematic shows, **each** of these core technologies **can be extended** to improve search effectiveness and efficiency. For example a development of vector space models is Latent Semantic Indexing, the aim of which is to find topics which are latent rather than specific in a document.<sup>7</sup>

A timeline for the major developments and company launches in enterprise search is shown below:



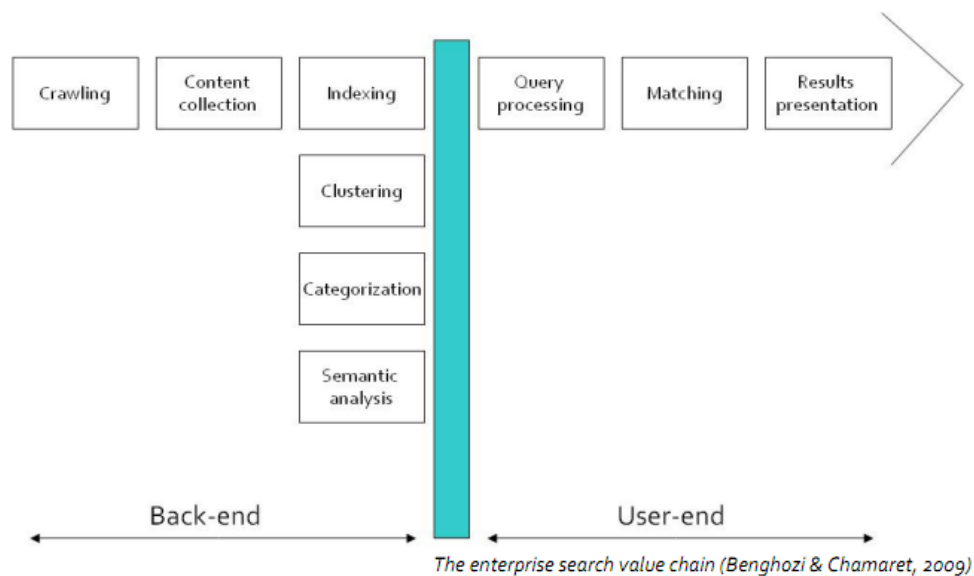
With the exceptions of Autonomy and Recommind search, vendors generally do not disclose the exact technological basis of their search application.

In fact the architecture of the aforementioned models is similar:

- Crawling repositories to identify content that has been added since the previous crawl that needs to be indexed;
- Indexing the content in a way that ensures that queries can be run in a way that generates responses with an acceptable latency;
- Taking the query and matching it against the index using one of the four models described above;
- Ensuring that only information that a user is permitted to see is delivered;
- Providing an effective user interface so that the user is able to undertake a dialogue with the search engine

<sup>7</sup> A document on "solar power" could then be presented in response to a query about "low-carbon technologies" without needing to use the specific term "solar power". The patents for LSI were granted in 1989. (3) There is also a combination of vector, latent and probabilistic approaches used in the Probabilistic Latent Semantic Analysis model developed by Hoffman and Puzicha the late 1990s and which forms the underlying technology of the Recommind search software.(4)

- Present the information in a sequence determined by the user, which is usually in descending order of relevance;
- Facilitating access to this information through links to the relevant repositories.



The processes of content collection and indexing involve the use of connectors and document filters. So-called "connectors" provide a 'translation' of the file management systems of applications that need to be searched, such as a document management application. These are often specific to a release of the application and problems can arise when the application is upgraded, leading to a reduction in search quality or performance. Document filters work on the structure of the information, identifying metadata tags and eliminating codes and control characters that are not required in creating an index of the information item.

Enterprise search technology is more than not just implementing one of three core indexing technologies. It requires specific implementation of information retrieval sub-topics, such as:

- Retrieval Evaluation,
- Relevance Feedback and Query Expansion;
- Text Classification
- Indexing and Searching
- Parallel and distributed information retrieval
- Web Retrieval
- Web Crawling
- Structured Text Retrieval or Multimedia Information Retrieval. (2)

As a result, the development of enterprise search applications requires a wide range of specialised skills, in particular mathematical approaches to set theory, probability and computational linguistics. There are high research and development costs for any company developing search applications. In the case of Autonomy (the only vendor to disclose employee numbers), over 1,000 of the 1,600 employees were categorised as technical staff in the 2010 Annual Report. In the press statements arising out of the HP offer for Autonomy in August 2011 there was an indication that the core development group consisted of 400 engineers.

The target is to provide relevant documents/information as quickly as possible, whilst recognising that the users almost certainly will not know exactly what they are looking for. Search is a dialogue, and a lot of development work goes into user interface design that supports a dialogue between the

user and the search application. Ideally a search interface should have the following characteristics (5):

- Offer informative feedback
- Support user control
- Reduce short-term memory load
- Provide short-cuts for skilled users
- Reduce errors and offer simple error handling
- Strive for consistency
- Permit easy reversal of actions
- Design for closure
- Avoid empty results sets
- Support users with a wide range of subject knowledge.

Each of these is a very substantial challenge because the range of content searched on the initial implementation of the search application will almost certainly change over time, and these changes have to be accommodated in a way that has a minimal impact on the user interface and the user experience.

There are a number of ways to defining **retrieval quality**. The most frequently used are

- *Precision* –the fraction of the retrieved documents which are relevant,
- *Recall* – the fraction of the relevant documents that have been retrieved.

Both are stated in terms of relevance, which is a very subjective metric. In addition both assume that all the documents in a defined collection (repository) are known, which is rarely the case. For compliance purposes, very **high recall** is important to make sure that no documents remain undiscovered that could prejudice the successful outcome of a court case, for example. High precision would mean that a user did not have look at a lot of documents (or at least titles) that were not relevant, so speeding up the search process.

A number of search vendors talk about “accurate search”, but this adjective has no value in assessing search performance, nor has “accurate information”, as the search engine is not able to assess what content is accurate and what is not.

Moreover, it is important to note some of the differences between enterprise search and public web search. These differences are related to the following aspects.

**Security management** – access to certain information may need to be restricted to specific individuals, and managing the Access Control Lists that are used to define access permissions is a major challenge for enterprise search software and subsequent maintenance. There are two different approaches to security management. Either the search can be restricted at the outset to information that an employee is able to see (often called ‘early binding’) or a search is conducted across all repositories and then results are filtered so that only those the employee is entitled to see are displayed (‘late binding’). In both cases the need to check the security profile of the employee can introduce a delay into the presentation of the search results.

**Co-existence of web content, databases and documents** – intranets in particular contain both web pages and text documents. The relevance of web pages can be determined by linkages to the pages, an approach that cannot be used with text documents or structured databases.

**Taxonomy management** – organisations have specific terms for procedures, departments, policies and products, and managing synonyms in particular is an important requirement.

**Multiple languages** – international organisations often use corporate English as a default but in reality there will be many documents in other languages. UN agencies as a matter of policy have to support the six defined languages of the United Nations.

**Dates and other metadata** – dates are of great importance in enterprise search because e.g. it can be very important to limit a search to content published after a defined date. However, often content does not have a defined date because there is no formal requirement for metadata to be added. There are also differences in date formats between countries. 5/4/2011 will be 5 April in Europe but 4 May in North America.

**Document specificity** – in a web search a user is looking for at least some relevant information, but in an enterprise search context, especially in a compliance context, a user may be seeking a specific document, or even ideally all relevant documents (100% recall).

Search technology has progressed slowly and almost invisibly over the last forty years. The three models discussed above (Boolean, vector, probabilistic) still form the basis for all current enterprise search engines. It is however not possible to state *a priori* which technology is the best for a given set of enterprise requirements because of the other components of the search engine.

It is likely that most IT managers will have had little formal teaching about information retrieval technologies and may not be aware of how to evaluate search applications, and to plan adequately for implementation and subsequent optimisation.

Another aspect should be noticed. Enterprise search applications are used not only for searching enterprise repositories but also to provide high quality search features on organisational web sites, and many organisations use the same search software for both internal and web search. One example is the Open University in the UK which uses Exalead for both applications. (6)

It is possible to foresee that over the period to 2015 there will be no major innovations in core enterprise search technology beyond the three approaches summarised above. However, enterprise search vendors are continually adding **new features** to improve search performance and to offer enhancements to the user interface. These include:

Being able to search multiple languages with a single query	Especially in the EU the ability to conduct a query in French but to have the search engine look for content in other languages is important. This requires the search engine to recognise which languages the content is written in and also to understand the semantic structure of the content.
Cloud-based search	As organisations begin to make increasing use of cloud-based applications for enterprise information management, the migration of enterprise search will be a requirement. It seems likely that initially this will be a hybrid model.
Enhanced federated search	One of the longstanding challenges for enterprise search has been to provide a way in which users can search across multiple applications and repositories with a single search query and be presented with an integrated set of results ranked by relevance.
Enterprise mobile search applications	As mobile devices, both tablets and smartphones, are increasingly used to access enterprise information there will be a need to provide search interfaces that can be used on small-format screens with at best a virtual keyboard. Search vendors are gradually beginning to offer these applications
Expert search	When a repository is searched for a specific topic algorithms in the search application recognise names of employees associated with this content as a way of identifying employees with specific expertise
Integrated search of structured and un-structured content	This is a specific case of federated search, but has an additional degree of complexity in indexing relational databases which may be being updated with (for example) new customer data on a very frequent basis
Multimedia search	With the rapid growth of enterprise collections of images, drawings, video and other non-text media search applications need to be able to search this content even when there is minimal or no text metadata associated with the content.



Search as an integration platform (unified access platforms)	Google has transformed web search by providing a simple query box which then provides access to global web content. One vision of the Digital Workplace is that search replaces menus and links in locating relevant information.
Search incorporated into business intelligence applications	Business intelligence applications have traditionally focused on reporting numeric information. This is changing rapidly as organisations recognise the value of unstructured information on assessing business performance, such as the types of complaints handled by a call centre
Search-based applications	In these applications search supports task-specific operations rather than being used just to identify documents and other information.
Semantic search	The underlying concept of semantic search is to identify relevant information from the context of the language of the document even if a specific keyword is not present. For example locating information on high-risk projects when the project reports themselves do not consider the project to be high risk.
Tagging of search results	This technology enables users to comment on the documents retrieved from a search. Positive comments on the value of a document would result in the document being promoted in a relevance-ranked list of search results conducted by other users with similar interests
Text mining	Text mining uses statistical approaches to look for patterns of text in sets of documents which may identify underlying commonalities about the documents. The results are often presented visually. It can be seen as search in reverse – from the pattern analysis keywords can emerge which can then be used as search terms.

### **Summary**

1. The development of enterprise search technologies dates back to the late 1960s. There are three main technical approaches; Boolean, vector space and probabilistic. It is very unlikely that any new technical approaches will be commercially available in the period to 2015.
2. There will continue to be many technical advances to improve the performance of enterprise search software and to increase the range of applications of the software in the enterprise.
3. The development of enterprise search applications requires a wide range of specialised skills, in particular mathematical approaches to set theory, probability and computational linguistics.
4. There are some significant differences between the requirements of searching web sites and searching other enterprise applications, primarily around security management.
5. Nevertheless is it possible to use enterprise search applications for both internal and public web site applications with appropriate customisation.
6. It is likely that most IT managers will have had little formal teaching about information retrieval technologies, and may not be aware of how to evaluate search applications, and to plan adequately for implementation and subsequent optimisation.

### **Resources**

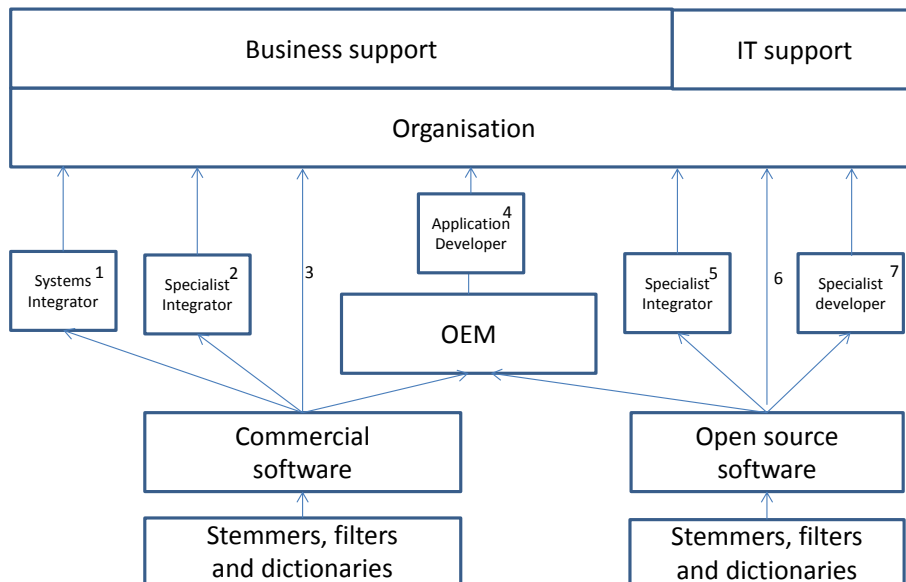
1. Charles P. Bourne and Trudi Bellardo Hahn. The History of Online Information Services 1963-1976. 2003. Massachusetts Institute of Technology.
2. Ricardo Baeza-Yates and Berthier Ribeiro-Neto, Modern Information Retrieval. Chapter 15. Second Edition. Addison Wesley. 2011
3. Thomas K. Landauer, Danielle S. McNamara, Simon Dennis and Walter Kintsch. Handbook of Latent Semantic Analysis. 2007. Psychology Press

4. Thomas Hofmann. Probabilistic Latent Semantic Indexing.  
<http://www.cs.brown.edu/~th/papers/Hofmann-SIGIR99.pdf>
5. Marti A. Hearst. Search User Interfaces. 2009. Cambridge University Press
6. <http://www.open.ac.uk/>

## Part 2: Market Considerations

### 2.1 The value chain for enterprise search

In order to discuss structural issues regarding the enterprise search business it is important to consider the value chain of the business.



From the perspective of an organisation there are seven ways in which an enterprise search application can be procured and implemented, and these are numbered in the diagram above

For **commercial** products:

1. An incumbent systems integrator which is already providing support for enterprise applications (such as IBM or Accenture) could also provide an enterprise search application.
2. An integration company that specialises in search applications could be asked to select and implement an enterprise search application.
3. The organisation could work directly with the enterprise search vendor, who would then provide not only the software but also professional services support for the implementation
4. The search application may be embedded in another application on an OEM basis, where often the company developing the search software is not identified.

In the case of **open source** products:

5. The open source application is purchased as a product from a specialist integrator, where much of the development work has already been carried out and in some cases proprietary code has been integrated with the open-source code (e.g. Attvio, IntraFind).
6. The organisation could decide to develop the application using only internal staff resources.
7. A specialist developer could be used to provide a fully customised application, though this is often carried out in conjunction with internal resources.

Within the organisation there will be a requirement for both IT support to ensure that the hardware and software applications are working to agreed technical performance standards, but there will be a much greater requirement for a search support team.

### ***Stemmers, filters and dictionaries***

Although search vendors have significant skills in software development, supported by expertise in computational linguistics, applied mathematics and enterprise architecture management they also make use of specialised modules from other companies.

One example is the development and supply of document filters. These break down binary files into a plain text that a search engine can index. (1) Document filters are important in the value chain because poor filtering may result in retrieving documents with meaningless titles, garbled summaries and irrelevant results. Hawking provides a summary of the **issues** around document filters (2) which include:

- The use of proprietary document formats (Microsoft Project and Microsoft Visio were good examples until the current release of Microsoft Office);
- The loss of text semantics when a document is encoded in a presentation-oriented format such as PostScript or PDF;
- The poor representation of metadata, with the result that documents which are textually similar but are in different document formats cannot be recognised and filtered out;
- Indexing the content of documents which have been scanned.

The main independent supplier of document filters is Isys-Search. There are also a number of small open-source developers of document filters.

Teragram (owned by SAS) and Basis Technologies are market leaders in a different type of text analysis challenge and that is the extraction of indexable text from Asian, European, and Middle Eastern languages, and related linguistic technologies. Google uses technology from Basis Technology to develop its Chinese, Japanese and Korean search engines. The licence costs charged to vendors are passed on to customers in the price of the search software.

Documill, a Finnish company, provides a visualisation of the documents in a set of results that facilitates the review of the results.

Commercial relationships between search vendors and specialised suppliers are usually not disclosed by either party. This is because the vendor is reluctant to disclose that the technology they deploy is not developed by internally.

### ***Direct sale and implementation***

Smaller search vendors will often work directly with clients, especially where the software has been designed to work out-of-the-box. There may be a need for two-three days of support, mainly around the installation of the software on the server, sorting out disaster recovery options and testing them, and setting up the crawl routines.

### ***Specialist integrator***

Sometimes the integrator has been engaged by the customer to define the search requirements and manage the process of product selection and then implementation. In other cases the vendor may feel that the implementation process is too complex for them to support, especially in countries where they may have little or no local office support, or where there are particular technical issues to be overcome.

### ***General purpose systems integrator***

Often a company has outsourced its IT services, or uses a systems integrator to provide support for the implementation of new applications. Search implementation usually only represents a very small revenue opportunity for systems integrators, and so there may not be many staff who can manage a search implementation. For this reason systems integrators work with a small number of search vendors who can provide back-up support to their consultants.

## **OEM applications**

The OEM (Original Equipment Manufacturer) business is of major importance to many vendors. The version supplied to the customer may well have a reduced functionality compared to the current version of the product and, indeed, may not be subject to the same upgrade roadmap as the stand-alone search product. For many vendors, especially Isys-Search and Autonomy, this is an important sector of their business, but often the vendor of, for example, a customer management application will not disclose the origin of the search application.

A twofold observation could be made: on the one hand, when vendors quote an installed base for their search product, this may well include the number of third-party installations that use the product; on the other hand, when installed in a third-party product, the search application cannot be used to search other applications in the enterprise.

### **Summary**

1. The enterprise search business has a complex value chain. In the past commercial products dominated the market but there is now increasing interest in the use of open-source software.
2. Although search vendors have significant skills in software development, supported by expertise in computational linguistics, applied mathematics and enterprise architecture management they also make use of specialised modules from other companies.
3. From the perspective of an organisation there are seven ways in which an enterprise search application can be procured and implemented, and this makes it very difficult to make any broad generalisations about the future development of the enterprise search business in the EU
4. Within the organisation there will be a requirement for both IT support to ensure that the hardware and software applications are working to agreed technical performance standards, but there will be a much greater requirement for a search support team.

### **Resources**

1. Where Have All the Filters Gone <http://www.ideaeng.com/tabId/98/itemId/136/Where-Have-All-the-Filters-Gone.aspx>
2. David Hawking. Enterprise Search in Ricardo Baeza-Yates and Berthier Ribeiro-Neto, Modern Information Retrieval. Chapter 15. Second Edition. Addison Wesley. 2011

## **2.2 The enterprise search business structure**

The global search business is a collection of **around 200** mainly small specialised companies offering a range of search products and services. Only five of these companies, Autonomy, Google, IBM, Microsoft and Oracle, have sales of enterprise search software in excess of \$200 million. With the exception of Exalead all the other companies are estimated to have software sales of less than around \$20 million. (1)

Appendix A provides a list of over **60** enterprise search vendors whose total sales of search software probably account for over 95% of the **EU market**. In total there are eleven countries represented in this list. Around half of the vendors on this list have their headquarters in the USA.

Enterprise search vendors based in the EU are listed in the following table:

Country of location of corporate headquarters	Representative companies
Austria	Fabasoft
Denmark	Ankiro Surfray
France	Exalead OpenSearchServer Polyspot Sinequa Temis
Germany	Intrafind SAP
Italy	Expert System
Spain	Inbenta
UK	Active Navigation Autonomy Concept Searching Simplexo Smart Logic Sophia Systems

Of the companies based in the EU Autonomy and Exalead have a global reputation for technology, and Sinequa is fast approaching this level. FAST Search was acquired in 2008 by Microsoft and the company keeps a substantial R&D effort in Norway.

The business structure can be clustered following different criteria, such as the company's turnover, its business area, or its technical competence.

Two main features could be observed with regard to the ESs market structure: the diversification in the strategies of ESs providers) and its expected consolidation.<sup>8</sup>

For the purposes of this report the search industry can be categorised into three Types.

**Type 1**

These are all major IT companies selling a range of applications, one of which is enterprise search. They are all public quoted companies and operate internationally. The top six Type 1 companies are Google, HP, Lexmark, Microsoft, IBM and Oracle.

HP, Microsoft, IBM and Oracle primarily sell search software as a module in a suite of products. Lexmark focuses on enterprise content management applications through its subsidiary Perceptive Software. Google is rather different in offering a search appliance; in effect search software ready installed on a server for rapid implementation, but this is only a small element of the company's business interests.

Some of the features of these companies include:

- An extensive network of sales offices, either direct or through partners;
- Significant research and development resources;

---

<sup>8</sup> See P-J. Benghozi and C. Chamaret *Economic Trends in Enterprise Search Solutions*, JRC Scientific and Technical Reports, EUR 24383EN-2010: while the enterprise search market is dominated by US industry, European firms are/were in many cases the first entrants in innovating fields of search, thanks to their R&D capabilities; however, one of the main threats for the European firms is the lack of financial power and the risk to be easily acquired by bigger competitors (p. 5, 49).

- High brand reputation and visibility through high-profile marketing campaigns;
- A global customer base of large multinational companies and other organisations;
- Solid financial position as a stock quoted companies.

Extended profiles of these companies are provided below.

## **Type 2**

The primary characteristic of Type 2 companies is that they have developed proprietary search software and mostly funded by venture capital and private equity placements. Because of their small size there are no requirements to publish detailed accounts of revenues and costs, and most of these companies have revenues of less than \$20 million. During 2012 Attivio has gained an investment of \$37 million and Coveo an investment of \$11 million.

## **Type 3**

Type 3 companies provide enterprise search solutions based on open-source software. Technically IBM would also fall into this category because its Omnifind search solution is based on Lucene/Solr. It is important to differentiate companies providing these kinds of products, such as LucidWorks from Lucid Imagination, from developers who use Lucene/Solr to create a customised solution for a customer.

There are a number of different product/solution **strategies** that have been adopted by search vendors, and these are listed below. The companies listed are representative of the business strategies and a number offer their solutions through several different product/solution strategies

### **Analysis of product/solution strategies**

<b>Business strategy</b>	<b>Characteristics</b>	<b>Examples</b>
Appliance	The search application is supplied installed on a server for direct introduction into a server rack	Thunderstone Google
Cloud	The search application is provided as a cloud service	Autonomy Exalead
Desk-top	The software is installed on individual PCs or on shared drives	Coveo Isys-Search
E-Discovery	Although in principle any text retrieval application could be used for e-discovery a number of vendors offer specialised products for this requirement	Zylab Recommind
OEM	Search software is provided to be embedded in a solution	Isys-Search Autonomy
Open Source Product	Open source software is used as the basis for a search solution but has additional functionality added to the software	IBM IntraFind, Lucid Imagination SearchBox
Pure play	Virtually all revenues arise from the sale of search software	Coveo dtSearch Isys-Search
Suite	Some of the major IT vendors offer search functionality as one of many elements in their enterprise software suites	IBM Websphere Oracle Web Logic SAP Microsoft
Utility	Search offered for free as a means of attracting customers to upgrade to an enhanced solution	Coveo Microsoft

A number of different contractual models for enterprise search are set out in the following table.

**Analysis of contractual models**

Contractual model	Basis	Example
Document volume	The contract combines a licence fee and a fee per document. Usually there is an initial document limit which can be extended if required. The duration of the licence fee for Google is either two or three years, and then a further licence fee has to be paid	Google
Free	There is no charge for the software but there is usually a limit to the number of documents that can be indexed	Microsoft Coveo
Microsoft SharePoint 2010	The pricing for SharePoint is based on the number of servers and on per user Customer Access Licence (CAL). See below	
Open source	There is no fee for the software licence but there will be development costs that have to be met by the organisation.	Apache Lucence
Open source product	Although the basis of the application may be open-source additional features have been added, and this enables a licence fee to be charged by the company	Intrafind Lucid Imagination
Software license	The software is provided under a licence for the core software and specific specialised modules. A maintenance fee is charged for service support and software upgrades, typically at 15-20% per year starting on installation.	Isys-Search
Suite module	IT companies offering enterprise suites will offer search as a module, with the cost often being a function of the total contract value for the suite	IBM Oracle SAP

An indication of how important the leading vendors see the EU market can be gauged from the extent of their office **locations in Europe**. Autonomy has offices in eight EU countries (Belgium, France, Germany, Italy, the Netherlands, Spain, Sweden and the UK), Exalead in six (FR, DE, IT, NL, ES, UK), Sinequa in two (FR, UK), and Recommind in two (DE, UK)

Many of these offices are small sales offices. In the case of Exalead this table excludes Dassault offices in the EU. The limited amount of resources in the EU of the Type 2 companies is an indication of the relatively low levels of business they currently gain from the EU market. This low level of sales and support indicate that with the exception of Autonomy (which in any case offers a wide range of other applications as well as enterprise search) these companies are not making significant sales in the EU at the present time.

An additional element in considering the structure of the industry and the importance of the EU market is that enterprise search is often acquired in one country and then rolled out across the international enterprise. The sale may be made to a US corporate headquarters and the sale booked to the US sales force but in effect the sale is also providing a solution to any operations in the EU. There may be some localisation requiring local development resources, but in effect (excluding internal cost transfers) the EU offices have an enterprise search licence at no direct cost.

There have been very few acquisitions of search companies over the last years.



The major acquisitions since 2005 are:

Company	Purchaser	Date	Business case
Verity	Autonomy	November 2005	Technology and client base Deal value \$500 million
iPhrase	IBM	November 2005	Technology Deal value not disclosed
Convera	FAST	April 2007	Technology and client base Deal value \$23 million
Mondosoft	Surfray	October 2007	Client base
Meridio	Autonomy	November 2007	Client base Deal value not disclosed
Teragram	SAS	March 2008	Broaden SAS product range Deal value not disclosed
FAST	Microsoft	April 2009	Technology and client base Deal value \$1.2 billion
Exalead	Dassault Systems	June 2010	Broaden Dassault product range Deal value \$185 million
Clearwell	Symantec	June 2011	Broaden product range Deal value \$390 million
inQuira	Oracle	August 2011	Broaden product range Deal value not disclosed
Autonomy	HP	August 2011	Broaden HP services range Deal value \$10 billion
I2	IBM	September 2011	Content analytics Deal value not disclosed
SchemaLogic	SmartLogic	September 2011	Broaden client base and extent product range Deal value not disclosed
Brainware	Lexmark	March 2012	Broaden product range Deal value not disclosed

Most of the usual business cases for acquisition do not apply in the search market. Few vendors have a large client base, so there is little to be gained from acquiring a company in order to persuade its customers to switch to the acquirer's search application. Only a large IT-based company would consider buying a company just for technology and in the case of Oracle and IBM their research is substantially ahead of most of the commercial vendors.

This is why the purchase of Autonomy is of special significance and it looks as if HP was willing to pay a substantial premium to ensure that no other company made a competitive bid.

There are a number of other players in the market, many of which would be categorised as systems integrators. The relationship between systems integrators and search vendors is complex. In general the source of revenues for any integrator is the number of person-days that can be sold to a client. As a result these companies are not all interested in search software products that can be setup and run with very little external support, of where the vendor undertakes the initial installation, either on site or by remote management. There are four categories of systems integrators:

- **Specialised integrators:** These companies specialise in search engine implementation and optimisation. They will typically work with a range of search vendors. Among companies offering these services in the EU are Raytion, Findwise, Search Technologies, Infosys, CapGemini and Forward IT ApS.
- **Microsoft Channel Partners:** These companies will have expertise in SharePoint 2010 search and in the implementation of Fast Search Server for SharePoint 2010. In most cases they will not have the expertise to manage the implementation of FAST Search ESP edition.

- **General purpose systems integrators:** These companies include Logica and Accenture, and usually have close alliances to a small group of high-end applications such as Autonomy and FAST Search.
- **Open-source search developers:** With the exception of Lucid Imagination these are mainly small companies often specialising in a particular vertical market and also working mainly for customers in just one country.

### **Summary**

1. The global enterprise search business probably has no more than 200 companies providing enterprise search software. There are around 60 enterprise search vendors who are either based in the EU or have made sales in the EU.
2. Six vendors, all of them multinational IT companies, have a major impact on the development of not only search technology but also the development of the search market through their promotional activities. These can be termed Type 1 vendors, with the other companies forming a large Type 2.
3. The primary characteristic of Type 2 companies is that they are mostly funded by venture capital and private equity placements and, because of their small size, there are no requirements to publish detailed accounts of revenues and costs.
4. The Six Type 1 companies are Google, HP, Lexmark, Microsoft, IBM and Oracle.

### **Resources**

1. International Data Corporation estimate for IPTS

## **2.3 The EU market for enterprise search**

In this section the EU market is considered from three perspectives

- Market revenues from sales of enterprise search licences
- The installed base of enterprise search applications
- The market potential for enterprise search

### **Market revenues**

It is very difficult to assess the size of the EU market for enterprise search in terms of revenues. There is very little published information even on the global market for enterprise search and none that provides an analysis of the EU market. According to IDC the worldwide revenue for the search and discovery market was \$2.1 billion in 2009. (1). The most recent estimate comes from Frost and Sullivan , which finds that the global market earned revenues of over \$1.47 billion in 2012 and estimates this to reach \$4.68 billion in 2019. (2) None of these companies sets out the revenues of individual vendors. This is to be expected as the large IT vendors do not publish revenues for specific products and virtually all other vendors are privately owned and there is no regulatory requirement for these vendors to disclose financial performance.

Among the leading vendors of enterprise search software would be IBM (Omnifind and Vivisimo), Oracle (Oracle Secure Search and Oracle Endeca), Google (Enterprise Search Appliance) and Microsoft.

To put these estimates into the context of global software markets Gartner forecasts that the global enterprise application software market in 2012 would be €120 billion and the largest single segment is enterprise resource planning software with forecast revenues of €24.9 billion (3). If for the purposes of a first approximation the global sales of search software are around \$2 billion then

this represents around 10% of those for enterprise resource planning software and 1.8% of the global market.

The only publicly quoted search vendor was Autonomy up to the date of the acquisition by HP. However the company set out in Note 5 to the Consolidated Financial Statement for 2010 a range of reasons why the company was not able to provide an analysis by product line. Following the acquisition by HP of Autonomy Oracle Corporation released trading information that had been supplied to the company by Autonomy during discussions about a potential acquisition by Oracle. (4) However this disclosure was not audited and so it is unwise to base any assumptions on this information. If Autonomy closed two large search contracts a week worth €1million across the world then that would be worth perhaps €100 million of sales but it is currently unclear how Autonomy realised the value of these sales in its financial reporting.

The largest EU search software company is Exalead, which was acquired by Dassault Systems in 2010. In 2010 the sales of SIMULIA, DELMIA and 3DVIA totalled €198 million and in 2011 the sales of SIMULIA, DELMIA, 3DVIA and Exalead totalled €283 million.(5) This would indicate that Exalead sales were of the order of €60 million (taking into account growth in the other software products). For a period of time Sinequa published sales revenues on its web site. In 2009 total sales of the company were \$12.8 million in 2009, of which over 90% came from sales in the EU (6). Since 2009 the company has only released annual growth figures which have been around 20% so that would put the 2011 revenues below \$20 million. The company does not provide a split of revenues by region. There are a number of other small EU search vendors, notably Polyspot (France) and Q-Sensei (US/Germany) but these companies are privately held and do not disclose revenues. A best guess for the total sales of EU search vendors would be no more than €100 million excluding Autonomy.

The other significant vendor in this market is Google, which publishes no analysis by product. Enterprise services are listed under Other Revenues in the Google annual report (7) and in 2011 sales were \$1.4 billion. These revenues arise from a wide range of enterprise services, such as enterprise licences to Google Aps.

There are a substantial number of small and medium-sized US enterprise search vendors who have some degree of representation in Europe. These include BA-Insight, Coveo, dT-Search, Isys (now owned by Lexmark) and Recomind. The size of the EU offices of these companies suggests that the number of customers is low. IBM, Oracle and Google strongly target the EU market.

The conclusions that can be drawn from this limited amount of information are

- An estimate for the global market for enterprise search software based on the published market information from leading analysts would be €2 billion. This represents less than 2% of global enterprise search sales
- The total revenues of EU-headquartered search vendors are between €100 million and €200 million depending on the view taken of Autonomy search revenues.
- Sales by IBM, Oracle, Google and Microsoft in the EU could be of the order of €300 million
- This would give a total market in the EU of at most €500 million, or 25% of the global market.
- There is no dominant supplier of search software in the EU.

These numbers have to be taken as indicative because of the lack of publicly disclosed information from search software vendors.

### ***Installed base***

It is equally difficult to establish an installed base figure for enterprise search in the EU, for basically three reasons

- Many large multinational EU companies may be using enterprise search applications that have been purchased in the USA or another country outside of the EU;

- Only a few companies disclose even a broad indication of the installed base of their software;
- Many companies may be using enterprise search applications embedded in enterprise application suites such as Microsoft SharePoint and Oracle Web Logic/Web Center.

The Exalead website (8) records that in total the company has around 350 customers. Sinequa claims to have over 250 customers worldwide. Autonomy has for some years indicated that it has an installed base of 20,000 customers. (9) A substantial number of these customers are based in the USA and many of these are using Ultraseek, which was a popular and low-priced enterprise search application acquired by Autonomy through its acquisition of Verity in 2005. It also seems likely that Autonomy includes sales of OEM versions of IDOL in this figure.

A number of companies quote a seat use figure. An example of this is PolySpot which reports that several million users are using the company's products for enterprise search requirements. (10) This seems to be a high figure but if for the purposes of illustration the total number of users is taken at 4 million and the average customer has 10,000 users then this would be an installed base of 400 customers. This figure should not be taken as an estimate of the PolySpot installed base, but only as an illustration that even a seat user based of millions only translates to an installed base of a few hundred customers. PolySpot profiles 30 customers on its web site.

A strong indication of the low installed base of enterprise search applications in the EU comes from the small scale of the vendor support organisations in the EU, with only Autonomy and Exalead having significant customer support operations. Probably the dominant vendor in terms of installed base is Google with its Enterprise Search Appliance but the support for this is provided mainly through e-mail and through a telephone help desk. Support is also provided by systems integration companies.

The software license fees for search software vary widely, from perhaps €100,000 up to over €1 million. If, for the purposes of an example, the average licence cost is €400,000 then to generate sales of €200 million would require 500 licenses to be sold, and for €500 million license sales would be 1250. Extrapolating an installed base from this level of supposition can only give an order of magnitude estimate for an installed base of enterprise software but this would suggest that the installed base over the last four years would be between 2000 and 3000 customers.

### **Market potential**

All the evidence from the surveys carried out by AIIM, Findwise and MarkLogic are that companies have failed to invest in the technologies needed to manage large unstructured information repositories. A survey sponsored by MarkLogic in 2012 indicated that only 25% of respondents were certain that they had the right infrastructure and database systems to manage the data volumes they expect in three years' time.

For the purposes of making an estimate of the market potential it might be reasonable to assume that a company with more than 500 employees would have a reasonable business case for an enterprise search application. In the EU no information is published on the number of companies by size other than those with 250 or fewer employees. In the USA there are 18,000 companies with 500 or more employees out of a total of 5 million. (11) In Europe there are 21 million companies of all sizes, but scaling up number of US businesses with 500 or more employees to 72,000 would not be a realistic comparison. For the purposes of this analysis the assumption is made that there are 20,000 companies in the EU with 500 or more employees to provide a very conservative market potential.

To this corporate market potential have to be added in central and local government and agencies, hospitals, public services and charities and the academic sector. For example there are around 4000 universities and colleges in the EU (12) and approximately the same number of hospitals (13) As a conservative assumption the potential installed base in the public sector might be a further 20,000 installations to give a total of 40,000. The six largest countries by GDP in the EU (Germany, France,

UK, Italy, Spain and the Netherlands) account for 65% of total GDP so might represent 26000 of these potential sites.

To provide a very conservative EU potential market we might assume that only 1 in 4 of these organisations will implement enterprise-wide search, to reduce the potential to 10,000. Comparing this to what is probably a high figure for the installed base of 3000 would suggest that there is still a substantial market for enterprise search vendors, and for suppliers of open-source applications, in the EU of 7000 establishments.

The purpose of the estimates that have been made is to suggest that:

- There is no dominant supplier of search applications in the EU
- There is a substantial potential market for the sale of commercial and open source search applications.

## **References**

1. International Data Corporation estimate for IPTS
2. <http://www.siliconrepublic.com/strategy/item/31182-global-enterprise-search-ma>
3. <http://www.gartner.com/newsroom/id/2055215>
4. <http://www.oracle.com/us/corporate/press/503343>
5. Annual Report 2011, Dassault Systems <http://www.3ds.com>
6. <http://arnoldit.com/wordpress/2010/08/09/sinequa-hits-13-million-in-revenues/>
7. <http://investor.google.com/financial/tables.html>
8. <http://www.3ds.com/products/exalead/customer-stories/>
9. <http://www.autonomy.com/content/Customers/index.en.html>
10. <http://www.polyspot.com/en/customers/polyspot-customers.html>
11. <http://www.census.gov/econ/smallbus.html>
12. [http://europa.eu/legislation\\_summaries/education\\_training\\_youth/lifelong\\_learning/c11067\\_en.htm](http://europa.eu/legislation_summaries/education_training_youth/lifelong_learning/c11067_en.htm)
13. [http://www.euro.who.int/\\_data/assets/pdf\\_file/0004/98401/E74486.pdf](http://www.euro.who.int/_data/assets/pdf_file/0004/98401/E74486.pdf)

## **2.4 Making a business case for enterprise search**

Because most business prospects for search vendors are buying an enterprise search application for the first time these customers face a number of challenges, such as:

- How to make a business case for implementing enterprise search,
- How to specify and select a search application,
- How to get the best return on the investment.

These topics are covered in the next chapters of this report.

Before a company selects or upgrades a search application there will be a need to prepare a business case for approval.

Given that most organisations already perform search in various ways, mostly through other enterprise applications and only for specific processes, the value may not be fully appreciated.

These include:

- Business Intelligence,
- Customer Relationship Management,
- Desktop Productivity Software,
- Document Management,

- Email Archiving,
- Enterprise Resource Planning,
- Web Content Management.

Most of these applications are either based on business processes and/or the need to comply with regulations and are also 'owned' by an individual department that is well prepared to make a business case for an upgrade. The challenge in the case of enterprise search is that it is of potential **value to most**, if not all, **employees**, but no single department wishes to take responsibility for making a business case.

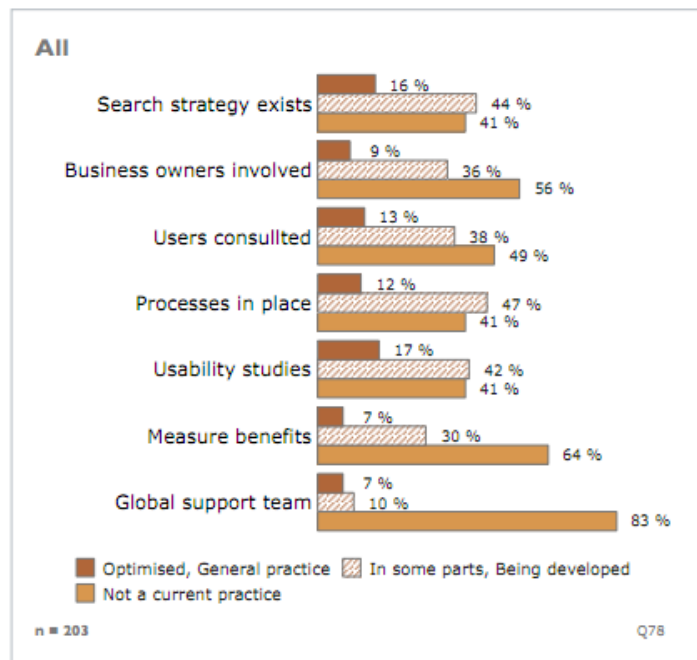
Organisations usually have no research which identifies the most important tasks carried out by employees. There are research techniques that can be used, such as those developed by Customer Carewords (1)

Among the tasks carried out in a multinational manufacturing company appear:

- Finding documents about policies, procedures, standards and guidelines,
- Finding technical drawings and diagrams,
- Find people by name,
- Locate information on pay, salaries and benefits,
- Checking on news about the company and its activities,
- Managing projects,
- Team working,
- Finding product information,
- Finding staff with specific expertise.

All of these tasks would benefit substantially from enterprise search and the scale of this list suggests that the IDC figure of 8 hours per week spent finding information is probably close to reality. Even if this work was halved by investing in an enterprise search application in a company with 30,000 employees this is the equivalent of increasing the workforce to almost 33,000 without any additional employment costs.

In 2009 the Global Intranet Trends survey (2) included some questions about the management of search.



This chart from the 2009 survey shows that only 16% of the organisations surveyed had a search strategy, and only 7% made a practice of measuring the benefits of the search implementation. In the Findwise study carried out in 2012 14% of organisations had a search strategy.

Much has been made of the value of productivity for making a business case, taking a view (as with the IDC research quoted above) that time saving can be calculated on a \$ value to the business. There are some fundamental **issues** with using **productivity as a business case** and these have been summarised by James Robertson, (3) namely:

- It is hard to quantify the time saved as it is rarely possible to compare search tasks under 'experimental' conditions. One vendor may claim that their application will reduce search time, but it will not be possible for its customer to undertake a range of carefully monitored searches on the current application, and then undertake a directly equivalent set on the new search engine to show productivity gain. There are too many variables.
- A productivity metric focuses solely on the time needed to find information, not the task or activity itself. Much more valid productivity gains can be argued from improving end-to-end performance, not just the time taken to find one item of information as part of an overall decision process with perhaps a dozen further steps
- Financial benefits must be "realised" before they count. In other words, if enterprise search could be shown to save \$1.7 million it is possible to get this as cash or equivalent? In practice, there are only two real ways of concretely realising the benefits: increase the number of tasks that can be done by the same number of staff, or reduce headcount. Neither are useful approaches.

One of the important uses of enterprise search is to locate information that has been internally generated about good practice that can then be used for a current project. Any employee writing a good practice document is likely to have been with the organisation for a number of years, and had some specific training as well as gaining significant practical experience. If the good practice information cannot be found then the costs incurred by the organisation in providing the training and the opportunity to gain the expertise are also wasted, and indeed may well be duplicated.

Quantitative approach is to review risk registers and show that search can reduce impact and probability scores. One major hospital in the UK was able to make a business case for a new enterprise search solution on the basis that it would enable clinicians to find treatment protocols

more easily and so reduce the professional indemnity insurance premiums that the hospital was paying.

The ability to speed the integration of an acquisition of another company, or a merger between two companies, through providing the employees of both companies with equal access to the information resources of the other, should be a powerful business case, especially for companies with a strategy for growth through acquisition.

A potential business case is the ability to track down the expertise of employees through name entity extraction. Name entity extraction enables a set of documents (e.g. on corrosion of aluminium in seawater) to be matched against the names of employees mentioned in the documents. There is no requirement to tag the names with metadata as software modules within the search application can recognise names from a set of rules about how names are presented.

This type of functionality would not be implemented in other enterprise applications, and is not available on public search sites such as Google and Bing so it is not unlikely that IT managers and business managers are unaware of the potential of enterprise search as a knowledge identification application.

From the views of the respondents and from information obtained from other sources it seems highly likely that many companies in the EU are unaware of the potential of enterprise search on business performance. One factor that may change this situation is the adoption of Microsoft SharePoint 2010 as the search functionality of this product, especially the FAST Search Server for Sharepoint version is substantially better than in Microsoft Office SharePoint Server 2007.

In the Delphi survey conducted for this report participants were asked for their views on what they regarded as the main barriers for vendors to selling an initial enterprise search application.

	All	Vendor	Analyst	Integrator
There is a lack of awareness of the functionality of enterprise search applications	1	1=	3	1
Organisations find it difficult to make a business case	2		1=	2
The view is that Google is the best search application and offers what is needed	3		1=	3
There is a lack of internal expertise to define requirements	4=	1=		
There is a lack of internal expertise to support the implementation	4=		4	4
The view is that SharePoint offers the required enterprise search functionality	4=	4		
Enterprise search is regarded as a low-priority investment		3		
Information is not seen as a business asset				

Note that the fact of enterprise search being given a low priority and that information is not seen as a business asset is not regarded as barriers by the respondents. Based on research carried out by a search vendor in 2011 (4) on the applications that were under consideration for investment in 2012 enterprise search was 14<sup>th</sup> in importance.

The responses in our Delphi-type study are ranked in decreasing order of an averaged viewpoint of all the respondents. It is of note that the Integrators' responses are significantly higher than the other two groups of respondents for the following issues:

- There is a lack of awareness of the functionality of enterprise search applications,



- Organisations find it difficult to make a business case,
- The view is that Google is the best search application and offers what it is needed.

Often these integrators are a first point of call for an organisation which is uncertain about which search application to deploy and also does not have the resources to carry out the implementation. These results support a view that overall there is no lack of awareness of information as an asset but there is a significant lack of knowledge about what an enterprise search engine can do and how best to make the business case. Google is seen as having a strong reputation for search excellence, making it a risk free procurement from a technical perspective.

Another reason for a company wishing to select a new search application is dissatisfaction with the current application. The reasons for this dissatisfaction are discussed in more detail in chapter 4.2.

The views of the respondents of the Delphi survey indicate that this dissatisfaction is a driver for replacing the current search application.

	All	Vendor	Analyst	Integrator
User dissatisfaction with the performance of the search application	1	1	1	1
The need to reduce on-going development and maintenance costs	2	2	2	2
Acquisition of the current vendor or concern about their long-term stability	3	3	4	3
Action taken by a business competitor	4	4	3	4

In the survey conducted by MarkLogic, cost is the major issue, but since none of the search vendors (with the exception of Google) publish licence fees for their products the view that managers are taking about the high cost of search may be based on a comparison with other enterprise-level applications. Cost is perceived by the respondents (56%) as one of the main barriers to improving unstructured data management, followed by human resources skills and by the lack of awareness of levels of unstructured data.

Respondents to our Delphi study were asked what would become increasingly important in the selection of an initial enterprise search application.

	All	Vendor	Analyst	Integrator
Ease of implementation of the application	1	1	1=	1
Brand strength and visibility of the vendor	2	3	1=	2
Technical features linked to business requirements	3	2	3	4
The ability to predict the total cost of implementation at the outset	4=	4		
A recommendation by a systems integrator	4=		4	
Minimal support needed to support the application after implementation	4=			3
A recommendation by a business colleague				
Quality of the information on the vendor web site				

This is quite a surprising result as the ease of implementation cannot be verified until the search engine is selected and installed. The quality of information on the vendor web site is ranked as being of low importance, even though evidence in Chapter 7 suggests the initial source of information on search products is obtained from a search of the internet by managers.

### **Summary**

1. The challenge in the case of enterprise search is that it is of potential value to most, if not all, employees, but no single department wishes to take responsibility for making a business case.
2. The main barriers to making a business case are that there is a lack of awareness of the functionality of enterprise search applications and the benefits that effective search can have on the enterprise.
3. Organisations usually have no research which identifies the most important tasks carried out by employees and the extent to which enterprise search would improve operational effectiveness.
4. Based on survey research carried out by a search vendor in 2011 on the applications that were under consideration for investment in 2012, enterprise search was 14<sup>th</sup> in importance.

## **Resources**

1. <http://www.steptwo.com.au/columntwo/25-reasons-why-saving-time-on-your-intranet-is-a-bad-metric>
2. <http://www.customercareswords.com/>
3. <http://www.digital-workplace-trends.com/>
4. <http://www.customercareswords.com/>



## **Part 3: The choice of enterprise search solutions**

### **3.1 Selecting and implementing enterprise search applications**

Either an organisation has a search engine but is not satisfied with its performance, or a search engine is being acquired for the first time. Each of these two situations has substantial challenges for search vendors in gaining new business.

Although an organisation may not be satisfied with the performance of a search application there is usually little quantitative information available to understand what the problems are. Often the decision to upgrade is made on the basis of anecdotal information about a document that could not be found or a search seemed to be 'slow'. Slow searches can be the result of many variables, but one that is usually overlooked is that the search engine uses a late-binding route to security management, and the person concerned only has limited access privileges. The result is that the search engine has to check each result before presenting the complete list of results.

Another common problem is that the quality of the information found through search is poor, and this is blamed on the search engine rather than on the failure of the organisation to maintain high information quality standards.

The process of selecting a search application is quite different between commercial products and open source applications, and the first section of this chapter deals with the selection of commercial products

Where a search engine is being selected for the first time there is often no relevant experience to call on within the organisation. This is unlikely to be the case with most of other enterprise applications where there is already a very large installed base. A senior business manager and/or a senior IT are likely to make the decision in principle to upgrade or buy a new search engine. The next step will be to bring in a business analyst to define user requirements and to develop a list of functional requirements that will meet the user requirements.

One of the challenges that an organisation faces in procuring a search application is to identify potential vendors. A search of Google on the topic of [enterprise search] will result in over 2 million results being identified, of which over half have been published in the last year. For technical reasons these numbers are extrapolated from the index files and may not represent an exact number of results. However the results do show that a) there is a significant increase in information about enterprise search over the last year and b) looking through even the first few pages of results will not result in a comprehensive list of vendors

There are five primary sources of information on enterprise search software

- The list of vendors covered in the Forrester report on enterprise search (1)
- The MarketScope analysis from the Gartner Group (2)
- The list of vendors covered in the detailed product reviews undertaken by the Real Story Group (3)
- The list of vendors profiled in the Ovum report Enterprise Search and Retrieval 2011/2012 (4)
- The list of vendors listed in the Wikipedia entry on enterprise search software. (5)

A comparison of the coverage of these five sources in 2011 is set out below. Some vendors listed in the Wikipedia entry do not provide general enterprise or e-discovery software applications.

	Forrester	Gartner	RSG	Ovum	Wikipedia
AskMeKnow					√
Active Navigation				√	
Apache Project			√	√	√
Attivio	√		√		√
Autonomy	√	√	√	√	√
Brainware				√	√
Clearwell				√	
Concept Searching					√
Coveo	√		√	√	√
Dieselpoint			√	√	√
dtSearch			√		√
EMC					√
Endeca	√	√	√	√	√
Exalead	√	√	√	√	√
Expert System		√		√	√
Fabasoft	√	√		√	√
Flax				√	√
Funnelback				√	√
Google	√	√	√	√	√
IBM	√	√	√	√	√
Inbenta					√
InQuira					√
ISYS		√	√		√
Lucid Imagination				√	√
Mark Logic				√	√
Microsoft	√	√	√	√	√
Nefonie					√
Omniure			√		√
OpenText			√		√
Oracle		√	√	√	√
Polyspot		√			√
Q-go			√		√
Recommind			√	√	
SAP			√		√
Simplexo				√	√
Sinequa	√	√	√	√	
SLI Systems					√
Sphinx					√
Stored IQ				√	
SurfRay			√		
TeraText					√
Thunderstone			√		
Vivisimo	√		√	√	√
Xapian					√
X1 Technologies				√	√
ZyLab					√
<b>Total</b>	<b>11</b>	<b>12</b>	<b>22</b>	<b>25</b>	<b>39</b>

However even the Wikipedia list is far from complete. BA-Insight, Sematext and Sophia are just three of the missing vendors and several (such as Q-Go) are arguably not enterprise search vendors.

It could be interesting to note that some research carried out in the USA by a search vendor provides an insight into the range of **ways in which search vendors are identified**. In decreasing order of importance the approaches used were:

- Search the internet to find what is available
- Read reviews on the internet
- Ask colleagues

- Ask IT providers and systems integrators
- Read industry analyst reports
- Look for relevant advertisements on IT-related web sites and magazines
- Attend IT webinars
- Attend IT events and conferences

These ways to identify search vendors are, for different reasons, substantially inefficient, as described below.

Approach	Inefficiency
Search the internet to find what is available	This is quite time-consuming, and there is no guarantee that all potential vendors will be identified
Read reviews on the internet	There are very few reviews of any quality on the internet. Many are largely anecdotal and as with all reviews the bias of the reviewer may not be immediately obvious
Ask colleagues	In SMEs it is probably unlikely that colleagues will have a broad experience of search, and even in large organisations there may be only limited experience of search applications
Ask IT providers and systems integrators	Systems integrators will have a good overview of the market but usually focus on the implementation and support of only a few search applications
Read industry analyst reports	The main providers of these services are the Gartner Group, Forrester, Ovum and the Real Story Group.
Look for relevant advertisements on IT-related web sites and magazines	Relatively few search companies, with the notable exception of Autonomy, use advertising as a means of developing business leads
Attend IT webinars	Only a few companies provide webinars, a good example being BA-Insight
Attend IT events and conferences	There are very few conferences either devoted to enterprise search or offering a track on enterprise search applications.

Larger companies in particular tend to rely on the reports from the major industry analysts, in particular the **Gartner Magic Quadrant**, although Gartner warns against the Quadrant as a quick evaluation tool.

The promotion of the benefits of enterprise search by vendors is very often based on features and functionality and not on presenting a business case that would help potential customers to make a case for investing in search.

It should be noted, furthermore, that until the announcement in March 2011 of the Enterprise Search Europe conference (7) there had been no regular event specifically for the industry at which the vendors could promote their products. Even in the USA the Enterprise Search Summit and Enterprise Search Summit Fall conferences are the only search-specific events, though the ESS Fall event is in fact one stream of the KM World Conference.(8) For some years a number of vendors have participated in the Online Information Conference and Exhibition held annually in London in early December.(9) There is also the Gartner Portals and Collaboration event in the USA and Europe but the search element in this event is quite small

Attendance at industry-specific events, especially in the legal sector, is another option, but many vendors have found it difficult to maintain the momentum after the initial contact is made at this type of event as the contact is usually only with the line-of-business manager.

Direct mail is unsuccessful as it is not possible to buy good lists of people who have search budget or search procurement roles in organisations. These people may be IT project managers or business analysts who are only involved with the requirement development and procurement for a short period of time.

The steps in the procurement process and the typical time taken to complete each step are:

Step	Scope	Time to complete (illustrative)
Initial decision taken to install a new/replacement search engine	N/A	N/A
Determination of user requirements	This may involve surveys, face-to-face interviews and group discussions	2 months
Preparation of the RFP	This will go through several iterations, and as well as IT and business input there will also be input from procurement specialists Development of list of potential vendors	2 months
Receipt of proposal	The vendor will need time to contact the company for clarification, and then write the document itself	1 month
Evaluation of the proposal	This is usually a more complex process than the company expected because there will be a core of requirements that all vendors can meet, and then many which are vendor-specific	1 month
Setting up an evaluation meeting	The vendor may either be given a three-hour demonstration slot or asked to demonstrate their application against a test collection.	1 month
Conducting the evaluation	See above	1 month
Contract negotiations	Because of the complex pricing models used by the vendors and the need to agree on an implementation partner this can be a time-consuming process	1 month
Initial installation	The initial installation and testing	1 month
User acceptance testing		1 month
	Total elapsed time	11 months

Although this is not untypical for enterprise solution procurement business managers are often frustrated that it will take this long for a solution to be implemented. This time delay is a major factor in deciding to implement either the Google appliance or use the search application in SharePoint 2010 extended to an enterprise scale just to provide a more immediate solution.

This long procurement cycle is similar to that for any enterprise application but there are some particular issues that have an impact on the enterprise search business.

Most search vendors have less than 250 staff, so writing a proposal in response to a specification can take a significant amount of time and expertise. The initial payment to offset this work will not be gained until perhaps 8 or 10 months later.

When it comes to open source applications the process is somewhat different, because there are no barriers at all to downloading any of the open source products, and in principle a company could download a number of them, develop them in parallel and then chose which provides the required functionality and performance. In practice companies do not have the resources to adopt this



approach, and more importantly may not have the resources to develop even one open source application. Although the code for these products, such as Lucene/Solr, is free and there are some books available on how to develop a search application, doing so requires a combination of software skills, a knowledge of how search engines work and a knowledge of business requirements.

To meet these requirements, companies will team up with development companies who specialise in open source applications, either outsourcing the development completely or just using the development company to supplement in-house resources. In effect each implementation is custom-built. Supporting these development companies is a global community of developers.

One of the challenges in open source development is being able to feel confident in the skills of the development team. Recently, LucidWorks, among the leaders in Lucene/Solr development, introduced a **certification scheme for developers** which will potentially have a significant impact on the confidence that companies will have in adopting Lucene/Solr based applications. (10)

The course covers:

- Lucene Background
- Indexing
- Searching
- Debugging Solr
- General Solr knowledge
- Architecting a Deployment
- General search and web application environment topics.

Whether the application is commercial or open source, a major challenge lies in testing the application to ensure that it meets technical and user requirements. The challenges are unique to enterprise search because of the unstructured nature of the repositories being searched. Other enterprise applications will be built around well-established workflows and data structures, as would be found in employee databases and customer databases.

A significant amount of work has been carried out over the last decade in the **development of test collections** which can be used to evaluate search applications on a common basis. Although these test collections are valuable in the development of search technology, the collections that are available (mainly through the TREC programme) are not suitable for testing the performance of in-situ enterprise deployments. (11) It is usually only when the initial index is built that issues around information quality, metadata availability or consistency and file formats come to the surface.

Solving these issues is not a one-time process, as the addition of new content to a current repository, or adding a new repository, can have significant impacts on search performance and this is why continuing support for an enterprise search application is so important in achieving consistently high search satisfaction amongst users.

#### **Summary**

1. There are no reliable and complete lists of search vendors operating in the EU market.
2. When procuring an enterprise search application the primary source of information used by IT managers is the web sites of vendors located through a search of the internet. This is quite time-consuming, and there is no guarantee that all potential vendors will be identified.
3. Because of the low installed base of enterprise search applications few IT departments have the expertise and experience to write tender documents for these applications
4. One of the challenges in open source development is being able to feel confident in the skills of the development team. Recently LucidWorks, among the leaders in Lucene/Solr development, introduced a certification scheme for developers which will potentially have

a significant impact on the confidence that companies will have in adopting Lucene/Solr based applications.

### **Resources**

1. <http://www.forrester.com>
2. <http://www.gartner.com>
3. <http://www.realstorygroup.com>
4. <http://www.ovum.com>
5. [http://en.wikipedia.org/wiki/List\\_of\\_enterprise\\_search\\_vendors](http://en.wikipedia.org/wiki/List_of_enterprise_search_vendors)
6. World Bank Request for Proposals for Enterprise Search RFP 10-0116, 3 September 2009
7. <http://www.enterprisearch europe.com/2011/>
8. <http://www.enterprisearchsummit.com/Spring2011/>
9. <http://www.online-information.co.uk>
10. <http://www.lucidworks.com/support-services/lucidworks-university>
11. <http://ilps.science.uva.nl/trec-entity/>

### **3.2 Search implementation and user satisfaction**

Research carried out by NetStrategy JMC in 2009 for the Global Intranet Trends survey report (1) indicated that search implementation is poorly executed.

More precisely, 60% of organisations reported that search is business-critical to them, but in only 10% of organisations are users very satisfied with the performance of the search application. One possible reason for this is that only 20% of the organisations have more than one person supporting the search application. This is not adequate to not only analyse the considerable amount of information in the search logs but also to have the skills and resources to make enhancements to the search application.

The survey was repeated in mid-2011 and showed that search engine satisfaction has dropped slightly with regard to 2 years ago. (2) In 2011, 10% are "very satisfied" and 40% are "moderately satisfied" for both the search engine features and the quality of the search results. In 2009 the figures for search engine features were approximately 15% "very satisfied" and 50% "moderately satisfied". Regarding content returned by the search engine the figures were 5 and 55%. 65% of the companies have "less than one person" dealing with search. The 2009 figure was 69%. These figures are in line with a subsequent study carried out in 2012 by Findwise.

Every user is an individual who has their own context for undertaking a search. Two clinical trials managers in a pharmaceutical company may search using the same query terms but have a very different take on the value of the results they obtain. The business is also changing day by day. A competitor launches a new product and as a result users need to locate a new range of internal information to counteract the threat. Search query terms may be used that have never been used before, especially if they are acronyms or specialised terms.

When users complain about the quality of search there is anecdotal evidence that the decision is to 'upgrade' the search application, on the basis that clearly the current search implementation is not adequate. Since the underlying issue is one of a lack of support post-implementation the results from the replacement search application are usually no better.

Organisations seem reluctant to invest in a search support team, almost certainly because there is little awareness of the importance of monitoring the performance of the search application. At a search workshop held at the Enterprise Search Summit in New York in May 2011 (4) the size of the search teams represented by the delegates ranged from 15 (a leading investment bank) to 0.5 (a

global top 10 pharmaceutical company). From an IT perspective the role of the department is to select and install the search application, and at most to ensure that all relevant repositories are indexed. The success of the search application can only be determined by business departments, and no individual department would be able to support a full-time member of staff just to monitor search satisfaction.

The issue then becomes one of which department should take the responsibility of assessing and enhancing the extent to which search is meeting business requirements, a problem that the organisation has already had to cope with in the case of intranet management.

One test of the commitment of the organisation to quality enterprise search is the extent to which a search team can be identified and funded at the stage when the requirements for the search application have yet to be fully developed. Enterprise search is not just a technology project but requires a very clear understanding of the information requirements of staff. No IT Director would dream of installing such a system without a cadre of business analysts delving in to every element of the process workflows but the same IT Director assumes that enterprise search just requires technical support for hardware maintenance. In practice, it is essential that a search support team is set up.

The **typical roles** within the search support team would be:

- ***Search (Information Discovery) Manager***

This role is not an IT role but requires a very good understanding of how information is used in the business, with a particular emphasis on unstructured information. The Search Manager might have a background in business intelligence applications but the key success factor is to understand the language of the business.

- ***Search Technology Manager***

This is an IT role and the person concerned will be responsible for assessing server and network performance, crawling schedules, load balancing, back-up and disaster recovery. In a multi-national company this may require treading on the operations of national IT managers. Typically an ERP or CRM application is country or at most regional specific, but enterprise search will be global from the outset and requires 24/7 availability. This may require an investment in hardware from the centre which cannot be justified by a national IT operation. As a result this is a 'management' role and not a technical role as the person concerned has to have the experience and the authority to ensure that things happen in operations over which they have no direct control. Another important responsibility of the Search Technology Manager is to manage information security, user authentication and user permission. It is usually not until an enterprise search application is fired up that all sorts of 'confidential' information is found in shared drives. Finally, this role should take responsibility for API management and documentation. Effective enterprise search across multiple applications will require some complex APIs which have to be kept under review as the individual applications are upgraded or re-structured.

- ***Information Specialist***

Good search needs good consistent metadata, and yet metadata management is not given the priority it needs in an enterprise search implementation. Relevance ranking invariably places more weight on words and concepts in the title of the document. If the title is either missing or not well written then the relevance of that document may be decreased even if in fact the value to the user of the content of the document is high. The Information Specialist ideally needs to have a background in information science or in librarianship so that they have a fundamental training in metadata management and in the benefits and challenges of taxonomies. A good taxonomy can be of considerable value in enhancing the search dialogue, but the development of taxonomies requires specialist skills, especially where a company is working in more than one language.

- ***Search Analytics Manager***

One of the critical success factors for enterprise search is the quality interpretation of the search analytics. One of the most important tasks for the Search Analytics Manager is to work through the searches which resulted in zero hits being found. If the assumption is made that only 0.1% of searches failed to find anything then this still represents a total of 500 searches a month, or around two each working day. Finding out why this search has failed may require some detective work, and certainly some feedback to the search user. A good case can be made for an analytics specialist for each business area in a highly diversified global corporation. The search terms used for aero engines will be very different for those from financial services.

- **Search Support Manager**

This person acts as the user-facing member of the team, doing training and usability testing, and providing feedback from surveys on the performance of the application. Although in theory search applications claim to need only minimal training the reality is that this is not the case, especially where federated searching is being carried out. Users may not fully appreciate the provenance of the various information repositories being searched and will need good guidance notes and suitable Help documentation on the search application.

In smaller organisations there would not be a need to appoint one person to each role, but for large multinational enterprise applications a team of five or more people would be required

The support requirements become greater when enterprise search is rolled out globally. There is likely to be a need for an Information Specialist for each major content language, especially in the case of German (where word length and complexity can raise some novel issues) and of course in ideographic languages such as Chinese, Japanese and Korean. These and other languages (Finnish is the classic example) will need attention paid to stemming and lemmatization and to seemingly simple issues such as the way that organisational names (such as OECD) appear differently in French (OCDE). This may not be a full time position but certainly the expertise needs to be available to the search team.

At the specification and selection stage not all of these are required full time. In principle it might be thought that there is no requirement for the Search Analytics Manager but given the importance of analytics they need to be involved in ensuring that the analytics requirements are fully specified and are tested in the Proof of Concept stage. Even at the early stages of implementation the team may be able to cope on a 'part time' basis but the evidence is that this approach is not sustainable for very long. It is important to remember that search touches everyone in the organisation who has access to a desktop. It is of worthy of note that only in mid-2011 was a book on site search log analytics published.(5)

### **Summary**

1. In a recent survey 60% of organisations reported that search is business-critical to them, but in only 10% of organisations are users very satisfied with the performance of the search application. One possible reason is that only 20% of the organisations have more than one person supporting the search application.
2. When users do complain about the quality of search there is anecdotal evidence that the decision is made to 'upgrade' the search application, on the basis that clearly the current search implementation is not adequate. Since the underlying issue is one of a lack of support post-implementation the results from the replacement search application are usually no better.
3. From an IT perspective the role of the department is to select and install the search application, and at most to ensure that all relevant repositories are indexed. The success of the search application can only be determined by business departments, and no individual department would be able to support a full time member of staff just to monitor search satisfaction.

4. For a large enterprise search implementation the advisable minimum search team comprises five persons with different profiles (an Information Discovery Manager, a Search Technology Manager, an Information Specialist, a Search Analytics Manager and a Search Support Manager)

### **Resources**

1. <http://netjmc.com/>
2. Personal communication from NetStrategyJMC
3. Personal communication
4. <http://www.enterprisesearchsummit.com/Spring2011/>
5. Louis Rosenfeld. Search Analytics for Your Site. 2011. Rosenfeld Media.

### **3.3 Enterprise search skills availability**

In 2011 a large multinational with headquarters in the EU took over six months to find a candidate to join the company as head of the team in charge of implementing FAST ESP. The company found that even recruitment agencies specialised in IT-related positions had no understanding of the requirements for enterprise search support. Two systems integration companies interviewed in the course of this project reported difficulty in finding staff with the appropriate qualifications, as did two EU-based search vendors.

The problem lies in the fact that the development of search software requires skills that are unique to the area of unstructured information. Although no surveys have been undertaken about the supply and demand for candidates for search-related positions the indications from a variety of sources is that there is a shortage of candidates with a good understanding of the science and technology of enterprise search.

One of the reasons for this situation is the low number of graduates in information retrieval. However there seem to be no full (i.e. three year) degree courses in information retrieval. There are some modules on information retrieval taught in more general undergraduate courses (for example at University College, London and the University of Sheffield) and the University of Twente teaches a course in data warehousing and data mining.

The workshop conclusions and discussion with professionals in the field seem to indicate that:

- Although there is a substantial amount of high-quality research into new information retrieval technologies in the EU, students (with a PhD) in information retrieval may not having the appropriate experience to work on enterprise search implementation within an organisation;
- There is little academic research being carried out on the specific issues of enterprise search;
- There seem to be no full time (3 year) undergraduate courses in information retrieval and search. In contrast, there seems to be more specific information retrieval courses in the USA, such as the undergraduate degree in Information Science and Technology at the University of Milwaukee (2). This University sees the characteristics of an IST Graduate as:
  - Highly skilled in software applications and Web 2.0 technologies
  - Cognizant of the role that information plays in society
  - Dedicated to providing a link between people and the information resources they need
  - Active in research that is pushing information in new directions
  - Specialist in the ethics and theory of information dissemination
  - Adaptable to the information needs of businesses, organizations, and institutions

In general the iSchool movement in the USA does seem to place a higher premium on information retrieval and research, but only five EU institutes<sup>9</sup> are members of this group (3)

In particular, the University of Sheffield iSchool (4) has developed a module on information retrieval for undergraduate teaching in 2011/2012.

It may be noted that Google already employs about 5,000 people in Europe and it announced recently expects to hire more than 1,000 new employees across Europe, at least half of them with engineering or computer science backgrounds. Given the potential value of having worked at Google and the leading edge research that this company is undertaking, a position at Google would be widely sought after by young graduates with the appropriate expertise. Given the rate of growth of Google, there are likely to be good opportunities to progress careers in the company and smaller enterprise search vendors will struggle to compete with Google and other large IT companies with research interests in information retrieval and enterprise search.

An innovative approach to ensuring good availability of graduate students with relevant skills is being taken by Findwise, a Swedish-based search integration company with 60 employees. The company is working closely with the University of Gothenburg and Chalmers University of Technology, and invites students on computer technology and engineering courses to undertake Master's projects with the company.

Companies seeking staff with the skills to support the implementation and management of enterprise search are also likely to struggle to find suitably qualified candidates. In most EU member states there are companies providing training in a wide range of enterprise application implementation, especially enterprise resource planning and Oracle database management. The majority of the courses on enterprise search are either workshops being run by vendors, or training courses on the implementation of Microsoft SharePoint Enterprise Search and the Google search appliance.

### **Summary**

1. The EU seems to have a significant lack of academic institutions that are offering taught courses in information retrieval.
2. There are less than 30 institutions in the EU undertaking research into enterprise search applications, but there are no full-time three year undergraduate courses.
3. As a result there is a shortage of skilled professionals to join search vendors as development and implementation engineers, and to join enterprise search support teams
4. In general the iSchool movement in the USA does place a higher premium on information retrieval and research, but only five EU institutes are members of this group.

### **Resources**

1. This list is based on [www.facebook.com/topic.php?uid=2243460765&topic=2185](http://www.facebook.com/topic.php?uid=2243460765&topic=2185)
2. <http://www4.uwm.edu/sois/about/mission.cfm>
3. <http://www.ischools.org/>
4. <http://www.shef.ac.uk/is/about>

---

<sup>9</sup> These are: Humboldt-Universität zu Berlin School of Library and Information Science; Royal School of Library and Information Science, Denmark; Rutgers, the State University of New Jersey School of Communication and Information; University; College Dublin School of Information and Library Studies; University College London Department of Information Studies; University of Sheffield, Information School.

## Part 4: Analysis and policy considerations

### 4.1 Technology Assessment and Forecast

The science behind enterprise search is information retrieval. The term was first used by Calvin Mooers in 1950. In 1959 he coined the term 'Mooers' Law', which states:

“An information retrieval system will tend not to be used whenever it is more painful and troublesome for a customer to have information than for him not to have it.”

This has been the foundation of the work on the development of search systems which are **efficient** and **effective** over the last fifty years.

Information retrieval is one of the disciplines within information science, which was formalised by the founding of the Institute of Information Scientists in the UK in 1958 and the American Society for Information Science (formerly the American Documentation Institute) in 1968. The development of the principles of information retrieval since the early 1960s has taken place in the USA and the UK in parallel, though with substantial contributions from researchers in many other countries of the world. Much of this work was undertaken at Cambridge University, resulting in the establishment of a number of enterprise search companies including Autonomy, Muscat and Flax. Another major contribution was made by Cyril Cleverdon, the Librarian at the Cranfield Institute of Technology. Cleverdon and his colleagues established the basic principles of the evaluation of retrieval systems using test collections, and set up a series of conferences which were of great importance in bringing together research teams from around the world to discuss common issues in information retrieval. These meetings formed the basis for the development of the Special Interest Group for Information Retrieval (SIGIR) of the Association for Computing Machinery (ACM) in the USA. The SIGIR has focused mainly on the science of information retrieval. From 1996 to 2010 the annual Search Engine Meeting, founded by the UK company Infonortics, became the focus for cross-fertilisation of information retrieval and enterprise search.

Another important series of conferences are the Text REtrieval Conferences (TREC), co-sponsored by the National Institute of Standards and Technology (NIST) and the U.S. Department of Defence, which started in 1992 as part of the TIPSTER Text program. These conferences aim to support research within the information retrieval community by providing the infrastructure necessary for large-scale evaluation of text retrieval methodologies. The TREC workshop series has the following goals:

- to encourage research in information retrieval based on large test collections;
- to increase communication among industry, academia, and government representatives by creating an open forum for the exchange of research ideas;
- to speed the transfer of technology from research labs into commercial products by demonstrating substantial improvements in retrieval methodologies for real-world problems; and
- to increase the availability of appropriate evaluation techniques for use by industry and academia, including the development of new evaluation techniques more applicable to current systems.

The TREC conferences were largely initiated and sponsored by the US Government's intelligence services, though the papers given at the conference are published. There is no similar workshop series in the EU.

A major issue in developing new enterprise search technologies is that it is **impossible** to construct, or use with permission, enterprise-type **collections** of information. Companies are not willing to provide access to what is regarded as confidential information and, even if a collection could be constructed, the range of queries that would be suitable to use as test queries would be constrained by the information content.

Within the EU there is a strong information retrieval community and the European Commission has actively supported information retrieval research for many years. However, little research has been carried out on the development of enterprise search.

One project that focused on enterprise requirements for information management was the Nepomuk project on the social semantic desktop. This project ran from January 2006 to December 2008 and the final report noted:

“Access to market and potential exploitation paths are some of the most crucial evaluation criteria for European projects, and it is often one of the weakest points of excellent researchers that they can’t bring their results into practice. NEPOMUK gathered almost all big international software companies besides Microsoft, thus creating a critical mass of extraordinary significance”[...] “Apart from access to market and potential exploitation paths, the project benefits from the presence / experience of the big industrial partners: they could provide the proof of concept that open source basic infrastructure can go well together with relevant proprietary systems and with commercial interests. It was important to have both IBM and HP in the boat, two major international software players which have a convincing open source strategy [...]”(1)

Many of the leading IT companies also undertake research into information retrieval and enterprise search, but understandably little of this is in the public domain. Among the companies that do publish information retrieval research are Fuji Xerox, Xerox Parc and HP.

During the research for this report, the respondents to the Delphi study were asked which technologies from a set developed by the IPTS Project Team they regarded as important to the demand and adoption of enterprise search.

	Total	Vendor	Analyst	Integrator
Integrated search of structured and un-structured content	1	2=	2=	1=
Search as an integration platform (unified access platforms)	2			1=
Search-based applications	3=	1		1=
Text mining	3=	2=	1	
Cloud-based search	3=			
Enterprise mobile search applications	3=		2=	
Search incorporated into business intelligence applications	3=	4	4=	4
Multimedia search	3=		4=	
Semantic search	4		4=	
Expert search				
Tagging of search results				
Enhanced federated search			4=	

The technology incorporated in enterprise search software allows these applications to be used in the enterprise for more than just searching through text information.

The reasons for this include

- Search engines are able to manage a wide range of structured and unstructured information in multiple file formats and in both internal and external repositories through the use of document filtering technologies.
- Search engines can adapt to changing data structures and terminological changes without the need to rebuild a Master Data schema, and are not restricted to queries defined by the column and row definitions of a relational database.
- Search engines can build clusters of content through categorisation routines that provide a context to the information retrieved, and facilitate a step-wise process of information discovery.



- Search engines are able to query multiple repositories using federated search, and the user is in a position to determine which repositories are included in the search.
- Search technology can be embedded in a wide range of other applications, such as customer relationship management and enterprise resource planning applications, on a third-party OEM basis.
- In this Section, we summarize briefly the main areas of technical development.

### ***Business intelligence***

BI tasks often require searching through different types of data within the enterprise. For example, a salesperson who is preparing for a meeting with a customer would like to know relevant customer information before the meeting. This information is often siloed into different sources: CRM databases, email, documents, and spread-sheets, both in enterprise servers and also on the user's desktop. Increasingly, a large amount of valuable data is present in the form of text: for example, product catalogues, customer emails, annotations by sales representatives in databases, survey responses, blogs and reviews. In such scenarios, the ability to retrieve and rank the required information using the key-word search paradigm is valuable for BI. (2)

One of the main drivers of enterprise mobile applications is the provision of business intelligence information, and was the reason why SAP acquired Sybase in June 2010. As a result, this demand could also be a major stimulus to the enterprise search market in the next couple of years.

The text of the workshop on 'Using Search Engine Technology for Information Management', held at the Very Large Database Conference in Lyon in 2009, summarises the situation well.(3)

"For information management, databases offer precise, controlled access to data. But, they do not offer the easy-to-use search capabilities that most knowledge workers manipulate daily on sites such as Google. Access to information contained in databases is more difficult, and more restricted. One solution to this information bottleneck is to let search engines take the brunt of the work, by offloading information from the database into alternative infrastructures, such as that provided by search engine technology. Many business applications such as search, report generation and data analysis might be performed more efficiently on the replicated data without involving the native database technology, e.g. transactions. These offloaded databases retain some of their structure, and can be recombined and mashed up, creating one-off, possibly disposable databases, while the primary data is kept safe in the original database."

### ***Cloud-based search***

For some years, a number of companies have been offering hosted search (Search as a Service) for web sites, but the use of these applications for enterprise search has been minimal because of concerns about security management. There is now considerable interest in cloud-based search but overall concerns about cloud security still remain. There have also been some high-profile outages of both the Amazon and Microsoft Office 365 cloud applications. Enterprise search presents some particular challenges to cloud service vendors in maintaining indexes which are subject to constant change and providing low latency to search users. (4)

### ***Expert search***

Entity extraction is an important component of enterprise search applications as it presents the user with information such as names, locations, geographical coordinates, phone numbers, dates, and other entities that can enhance the dialogue with the search application. In particular the approach is used to identify employees in an organisation who may have specific expertise because they are cited either as authors of a document or are mentioned in the document.

To this scope, three core technologies are available (5):

- Statistical models highlight names and probable best answers when words can have multiple meanings. (For example analysing the correlation with the other words helps

identify the correct context such as deciding when the word “gas” is used as a synonym for petrol in US usage.)

- Regular expressions define entities with standard patterns like telephone numbers and email addresses.
- Gazetteers—lists of entities—are used for entities which are well-defined words with little ambiguity about their meaning like the names of nations.

An example of the technique is provided by Basis Technologies.

All the highlighted terms have been identified by the entity extraction technology. For some years the entity extraction track at TREC (6) has been one of the most popular tracks, an indication of the potential value of this technology to customers.

### ***Federated search***

Federated search enables users to search across multiple repositories, each with its own search application, and be presented with an integrated and ranked list of results. Much of the research and development in federated search has been in the context of digital library services where there is a requirement for users to search across multiple databases. However these databases are highly structured and the solutions are often not suitable for enterprise.

Federated searching clearly has benefits to users but the technology issues are complex, requiring:

- Considerable flexibility in combining results in different ways;
- Maintaining and mapping security credentials across applications which may not have a common identity management application;
- Advanced detection and removal of duplicate documents, which has long been a major challenge for search vendors;
- Combining results-list and page navigators;
- Translating queries into different search syntaxes.

Most enterprise search vendors offer some degree of federated search but the performance of federated search applications is very dependent on content quality and metadata quality.(7)

### ***Mobile search and enterprise search***

The status and the prospects of mobile search was described in a recent IPTS report (80) in which it was explained to what extent the particularities of the mobile device (e.g. small screen size, preferred user actions of interaction, etc.) determined the range of applications that were more or less suitable. One of the limitations is that multiple-level architectures common for PC-based intranets are not suitable for mobile smartphone access. It is interesting to compare Amazon on a PC and the mobile version of Amazon to see how the architecture has been reduced to a much smaller number of levels. This could be a substantial problem for intranet managers as the CMS may not be able to support two architectures, one for desktop applications and one for mobile devices. The situation can be even more complex in the case of portal applications, where the font sizes can be very small indeed.

The Delphi respondents were asked for their views on the impact and direction of mobile search technology:

	Total	Vendor	Analyst	Integrator
Companies will prefer to extend the capabilities of their current search application to mobile devices	1	1	1	1
Mobile search applications for enterprise use will stimulate new vendors to enter the market	2	3	4	2
Innovation in mobile search will result in innovation in workplace enterprise search	3		2	3
Companies will place mobile search requirements as a justification for replacing their current search application	4=	4	3	4
The requirement for mobile enterprise search is overstated	4=	2		

The issue of mobile search is especially challenging given the almost ubiquitous use of filtering/facets in enterprise search applications to help manage the very long list of results returned by the search application. On the small screen of a smartphone, all the information presented in anything other than the central results pane is not going to be visible. Even with a reduced amount of metadata associated with each result (file size, for example), it may only be possible to display ten results at most, and probably fewer. As almost any query of an enterprise intranet produces a list of results that can quickly run into hundreds, this will be a significant challenge to manage. Scrolling through long documents is also a very difficult exercise on a smartphone, so the user interface will need to enable the user to step through each occurrence of the search term(s), rather than just scroll through the document.

One interesting outcome of the Intranet Focus Ltd survey (carried out in April 2011) was that almost half of the respondents had not yet considered the issues around mobile search. (9) Certainly in early 2011 there was little in the way of mobile search development from any vendor other than Isys-Search (10). However, a number of vendors are now developing mobile search products, notably Coveo (11) and Fabasoft (12).

### **Multi-language search**

Cross language information retrieval (CLIR) describes the retrieval of information written in one language based on a query expressed in another, e.g. typing a query in English to retrieve documents written in Finnish.(13) For such a process to succeed, translation of the user query written in the source language (e.g. English) and/or of the documents written in the target language (e.g. Finnish) must occur.

The extent to which vendors support multiple languages and cross language information retrieval varies widely, with Autonomy claiming that its IDOL technology supports 112 languages and that it is trivial to add support for more because the technology is fundamentally language independent. However, in most international companies, the corporate language is English and the percentage of content that is available only in a local language and is not a translation of an English original is quite small. At the other end of the spectrum, the EU Member States have 23 different languages, but as with the United Nations (which has six official languages), the working languages are English and French.

One of the important issues with multi-language search is the need to have stemming algorithms that are specific to the language. The aim is to improve search performance by bringing under a single heading variant forms of a word which share a common meaning. Stemmers are generally only applicable to Indo-European languages, and not to languages such as Japanese and Chinese.

One of the most well-developed and supported stemmers is Snowball, developed by Martin Porter in 1980.

### **Multimedia search**

Companies use Digital Asset Management applications to manage collections of images and videos. These applications are used by small teams of professionals within communications and marketing groups, and they have gained considerable expertise in adding metadata to this multimedia content that can then be searched in an embedded search application. The extent to which these collections need to be searched by a 'find a similar image' query is quite small outside of large media and publishing companies.

For a number of years, the EU has funded research into multimedia search, but the emphasis of this research has been in supporting multimedia search on the internet. Little account seems to have been taken of the opportunities and challenges of enterprise multimedia search. In a review published by the Chorus Project in 2010 "Cross-disciplinary Challenges and Recommendations regarding the Future of Multimedia Search Engines", no reference was made to enterprise multimedia opportunities. (14) However, the reality is that the enterprise market is relatively small.

The world leader in image search is LTU, which is based in Paris and was founded in 1999. Other companies offering image retrieval applications include IQ Engines, Moodstocks, Ookaboo, Imagga and TinEye, but, of these, only TinEye, developed by the Canadian company Idée Inc., is a close competitor to LTU. Autonomy specialises in searching video files with its Virage product, which is able to index close-captioning and also voice through voice-recognition to provide additional search metadata.

### **Search-based applications**

The line dividing unified information access from search-based applications is blurred. This is reflected in the almost identical scores in the Delphi survey for these two technologies, with the higher score from the vendors being a reflection that some of the respondents promote their technology as offering search-based applications. Search-based applications aim to offer precise, multi-axial information access and analysis that is virtually indistinguishable at a surface level from database applications, yet is endowed with the usability and massive scalability of Web search. Exalead defines a Search-Based Application (SBA) as any software application built on a search engine backbone rather than a database infrastructure, the purpose of which is not classic information retrieval, but rather mission-oriented information access, analysis or discovery.(15)

SBA platforms will be able to

- Collect and process unstructured, structured and semi-structured data,
- Consolidate data via an open API and connector framework (with an option to aggregate,
- content through federated search, mash-ups and metasearch as well),
- Use semantic technologies to effectively analyze and enrich source data,
- Automatically categorize and cluster content to support faceted search, navigation and reporting,
- Provide a search API or built-in dashboard tools for information visualization and analysis,
- Offer high performance and unlimited scalability (not all search engines are created equal!).

Virtually every Type 1 and Type 2 vendor is now in the process of extending their technology to be able to provide UIA and SBA capabilities.

However, this technology is not a replacement for transactional databases. Although enterprise search engines are able to create visual representations of database information, they are not able to offer the interactive modelling capabilities of current generation business intelligence applications.

## **Text mining**

Text mining concerns the analysis of concept co-occurrence patterns across documents in a collection. A significant difference between text mining and enterprise search is that, in text mining, the information is not provided by any one specific document but through the statistical analysis of the information in all the documents in the collection. (16)

For example, there could be a large collection of documents on the use of mobile smartphones and tablets for enterprise use. Using an enterprise search application to identify issues that might impact the future market would require using a query such as [barriers and/or growth and/or market]. Because text mining can examine the changes of patterns of words in the documents over time, it is quite likely that using text mining would identify issues around spectrum availability and ownership of handsets from the frequency analysis.

Text mining applications make extensive use of interactive visualisation of the results from successive passes across the document set, primarily to look at clusters of related concepts that can then be progressively refined.

Among the companies that provide text mining software are:

- AeroText (Lockheed Martin),
- Attensity,
- Clarabridge,
- Endeca Technologies,
- Expert System S.p.A. (Italy),
- IBM,
- SAS
- Temis (France).

This is not a difficult area for enterprise search companies to move into, but at present it is quite a specialised market sector. (17)

A market forecast in 2009 suggested that the global market was worth around \$350 million, less than 25% of the enterprise search market. As document collections get progressively larger it is likely that the requirements for text mining are likely to increase substantially.

## **Unified information access**

As a result a number of search vendors are now planning to expand the scope of their products and services beyond searching through unstructured data through unified information access (UIA) applications. Unified information access platforms provide a single point of access to multiple heterogeneous sources of information.

According to the International Data Corporation, (18) they are:

- highly scalable and typically include tools for semantic understanding, including fuzzy matching and a range of search and text analytics routines, as well as structured data and analytics operations.
- designed to work with real-time or near-real-time updating and analytics.
- combine elements of database, business intelligence, and search technologies to make information access dynamic and ad hoc for business users.

Unified access platforms are not just a combination of business intelligence and enterprise search applications, nor are they just an extension of federated search applications.

The International Data Corporation has suggested the following uses of these platforms:

- For quick-changing data, and for ad hoc access to current data, often in real time or near real time.
- For better, more complete understanding of the organization and its business across information silos of both structured and unstructured information.
- For better customer understanding and service, by merging external social media information streams like Twitter or Facebook with internal customer records.
- In healthcare, to find a unified view of the patient.
- For faster decision support and strategic decision making.
- For more flexibility in delivering information depending on the question or context: customers and employees receive more relevant information because these systems understand who they are, what their role is, and what their question is.

Among the search vendors have been at the forefront of creating unified information access platforms, including Attivio, MarkLogic, Palantir, Perfect Search and Sinequa, and to a more limited extent (Microsoft-only platform) BA-Insight.

### **Summary**

There are a number of technological developments that are being introduced into enterprise search applications to provide better search performance, broaden the value of the search application to businesses and provide solutions to emerging business requirements.

The most important of these developments are business intelligence, expert search, federated search, mobile search and enterprise search, multi-language search, multimedia search, search-based applications and text mining and unified information access.

### **Resources**

1. <http://nepomuk.semanticdesktop.org/xwiki/bin/view/Main1/Deliverables>
2. CapGemini Monthly Technology Briefs – The new meanings for the term Business Intelligence. October 2009. <http://www.capgemini.com>
3. <http://vldb2009.org/?q=node/30>
4. <http://cloudcomputing.sys-con.com/node/1952501>
5. <http://www.basis.com>
6. <http://ilps.science.uva.nl/trec-entity/>
7. <http://federatedsearchblog.com/>
8. J. Luis Gomez-Barroso, R. Compano et al. Prospects of Mobile Search, IPTS- JRC Scientific and Technical Report 2010.
9. Survey carried out by Intranet Focus Ltd and Intrateam A/S in April 2011. Unpublished.
10. <http://www.isys-search.com/technology/isysanywhere/index.html>
11. <http://www.coveo.com/en/technology-platform/coveo-platform-version-6-5/search-navigate/Mobility>
12. <http://www.foliocloud.com/en/features/everywhere>
13. <http://www.clef-campaign.org/>
14. [http://cordis.europa.eu/fp7/ict/netmedia/publications\\_en.html](http://cordis.europa.eu/fp7/ict/netmedia/publications_en.html)
15. Gregory Grefenstette and Laura Wilber. Search-Based Applications: At the Confluence of Search and Database Technologies. 2010. Morgan and Claypool Publishers

16. Michael W. Berry and Jacob Kogan (Editors) Text Mining: Applications and Theory. 2010. Wiley-Blackwell
17. <http://www.slideshare.net/SethGrimes/text-analytics-past-present-future>
18. Susan Feldman and Hadley Reynolds. Unified Access to Information; Less Seeking, More Finding. International Data Corporation Report #227790, April 2011

## 4.2 SWOT analysis

In this section, a SWOT analysis is presented of the overall EU enterprise search business and its markets.

Strengths	Weaknesses
<p>World-class search vendors based in the EU</p> <p>Significant information retrieval research capability</p> <p>Understanding of cross-language retrieval</p> <p>EC (FP7) support for information retrieval research</p> <p>Active information retrieval community</p>	<p>Difficulty in identifying search vendors and integrators</p> <p>No research being carried out into the benefits of enterprise search</p> <p>No enterprise search community and until Oct 2011 no specialised ES conference</p> <p>EU vendors finding expansion into USA a challenge</p> <p>Enterprise search has a low organisational priority</p> <p>Lack of reliable market data</p> <p>Public sector procurement procedures</p>
<p>Significant market potential in all Member States</p> <p>EU Member States committed to Open Data</p>	<p>Small scale of the EU enterprise search industry may restrict the exploitation of very large data sets ("Big Data") by EU companies</p> <p>Enterprise search licences bought by US companies for global use (including EU subsidiaries) reduces the EU market potential for large search contracts</p> <p>Shortage of candidates with appropriate training could impact the development of search vendors and the rate of implementation of enterprise search</p> <p>Market penetration by Google and Microsoft</p>
Opportunities	Threats

### **Strengths**

There are a number of world-class enterprise search vendors based in the EU, including Autonomy, Exalead, Fabasoft and Sinequa. There is a strong information retrieval research capability in the EU, which has been built on a strong funding commitment from the European Commission under Framework Programmes going back many years. Much of this research has been into multi-language and multi-media retrieval research to meet the requirements of EU institutions. There is an active information retrieval community, especially in the UK where the British Computer Society has taken a leading role.

### **Opportunities**

There is a significant market opportunity in the EU for enterprise search applications. There is a market potential of perhaps 200,000 organisations, though fewer than 20,000 organisations in the EU currently use an enterprise search application. There are no dominant enterprise search vendors

in Europe in terms of market share because of the scale of the potential market. All EU Member States are committed to releasing data sets of information that have been collected on business, social and economic activities.

### ***Weaknesses***

There is no reliable directory of enterprise search vendors. Research indicates that IT managers seeking to purchase or replace an enterprise search application tend to turn first to the internet. However the web sites of vendors tend to highlight the functionality of their products and not the benefits of search to the organisation, making it more difficult for an IT manager to develop a business case, especially as search seems to have a low priority in terms of IT investment. Although there is a significant amount of information retrieval research, no funding has been allocated to understanding enterprise search user requirements and implementation issues. Up until now, there has been no conference in the EU devoted to exploring the benefits and challenges of enterprise search, which could have formed a community of interest around the topic. The first conference of this kind will take place this year. The majority of search vendors are small companies with revenues of less than €20 million. This means that they are not in a position to undertake extensive market promotion of enterprise search, and entry into the USA has a high commercial risk. There is very little reliable market data on the installed base and potential markets for enterprise search. Public sector organisations have to use an open tender process for procurements which precludes discussions with prospective vendors about changing the RFP on the basis of a further refinement of user requirements.

### ***Threats***

Although there are no dominant enterprise search vendors, the wide-scale adoption of Microsoft SharePoint Server 2010 with FAST Search Server for SharePoint 2010, and the strong promotion by Google of the company's search appliance present a considerable threat to the ability of smaller vendors to capitalise on potential market opportunity as these two vendors are seen as 'low risk' by IT managers who are unfamiliar with enterprise search technology. Many of the larger EU companies with offices in the USA may be taking advantage of global licence contracts that have been purchased in the USA. It is probable that there is already a lack of suitably trained candidates with a background in information retrieval and its application in enterprise search to fill vacancies within the search business and in organisations seeking to implement and develop enterprise search applications.

## **4.3 Opportunities for EC support actions. Some policy briefs.**

The SWOT analysis conducted for this study shows that there are some significant weaknesses and threats in the EU enterprise search sectors. In this chapter, an indication is given as to potential actions that might be taken by the European Commission in conjunction with the enterprise search business and with organisations seeking to gain the best possible return from the wealth of information that they have created and stored. Policy challenges emerging from our techno-economic analysis of enterprise search in Europe relate mainly to the need to stimulate the potential of ES and to further increase European presence in the global context. A transversal issue regards whether and to what extent the EU might propose and intervene with ad-hoc policy options. With this ambition in mind, our report examined opportunities and challenges for the ES market, trying to envisage future potential development and techno-economic impacts of ES applications.

The policy considerations that follow are also based on the key findings of the "Exploring the future of Enterprise Search" validation workshop that took place on 13 and 14 October 2011 at the Institute for Prospective Technological Studies (IPTS) in Seville (Spain), see Annex E for the agenda.



### ***Identification of research requirements***

There is a significant amount of research being carried out into new information retrieval technologies in the EU, funded by both the European Commission and national funding agencies. Proposals for this research are generated within academic institutions who will have only limited, if any, relationship with either enterprise search vendors or end-users of enterprise search applications. The result is that exploitation plans are not as well developed as they might be, and issues relating to the intellectual property of outcomes of the research are not taken into account at the outset of the project or programme.

The Commission should consider inviting representatives of search vendors and integrators to attend a meeting at which the research interests of the industry could be discussed. Although there is considerable competition between search vendors, there are also common interests in some areas of research, especially in semantic analysis and in the design and evaluation of search user interfaces. The advent of mobile enterprise search will require a careful assessment of user interfaces. In the USA, the TREC conferences have not only provided the US search industry with the ability to test out new retrieval technologies but also provided a forum for discussion about fundamental and applied research requirements. There is currently no similar forum in the EU, and the proposed meeting could discuss the benefits and challenges of such a forum.

### ***Skills availability***

Developing enterprise search applications requires substantial investment in some sophisticated areas of applied mathematics, linguistics, database design and user interface design allied to an understanding of the organisational environment into which these applications will be deployed. Once deployed, a different set of skills is required by search support teams that enables the technical and retrieval performance of enterprise search applications to be monitored and enhanced to meet changing user requirements.

A review of the major information retrieval departments in the EU shows that there are almost no taught undergraduate or masters courses in applied information retrieval and enterprise search. The research orientation of these courses may well mean that faculties do not appreciate the potential requirements of both search implementers (especially in the open source sector) and of enterprises seeking trained support staff.

Experts consider that Europe, unlike the US, lacks trained people in the field of ES and they regret the low number of specific academic programmes on ES in European Academia. Though there is a general agreement amongst stakeholders on the skill gap in ES, opinions diverge on what kind of teaching and skills should be provided by academia to suit the industry's needs. While some experts consider forming "information retrieval specialists" in fully fledged master courses on ES to be the best option, others consider that training good IT experts on the particularities of search-related applications in dedicated courses would suffice.

Another bottleneck is the knowledge transfer from academic research to commercial products. While relevant European research projects in ES do exist, there is insufficient support and emphasis on bridging the transfer gap between research and industry. A closer relationship would also favour a better strategic approach in research in the area. Experts observe that – if further research is needed on enterprise search solutions – it is not always clear to the industry and professionals in which directions it should go (i.e. novel approaches to tackle the semantic gap, interfaces between modelling blocks, data-sets and collections for benchmarking, or linguistic filters, etc.).

This bottleneck, rather than being based on a divergence on content, seems rooted in a lack of exchange of views amongst practitioners and academics. There are very few platforms where they can meet. The Chorus+ activities like the Think Tanks or the specific thematic workshops are amongst those few. While there are several academic conferences on information retrieval, there is practically no industry-driven event in Europe.

The European Commission should assess whether existing tuition on information retrieval and associated disciplines is adequate to support the rate of growth of the EU enterprise search business and the effective implementation of enterprise search applications that are essential to enhancing the competitive position of EU companies.

**Enterprise Search helps to manage the increasing data growth and contributes to effectiveness and economic development, but awareness of its benefits is still low.**

One reason for adopting Enterprise Search is the growth in data generation. Right now, nearly all information is born (or immediately turned) digital; the amount of data is expected to reach 35 Zettabytes in 10 years. As explained above, even more worrying than the huge amount of information is its structure: it is estimated that about 80% of the information stored is either unstructured or has no adequate metadata for the needs of employees.

The major challenge for enterprise search solutions is still, therefore, to provide an integrated and value-adding retrieval of both structured and unstructured information and merge them into a sort of 'Hybrid structured data'. The vision is thus of Unified Information Access, which would provide the end-user with a user-friendly interface capable of retrieving heterogeneous data sources and providing added value by making use of semantic modelling.

One of the main barriers to making a business case for ES is the lack of awareness of the functionality of enterprise search applications and the benefits that effective search can have for the enterprise. However, awareness is steadily percolating into the minds of decision makers in IT-intensive companies and large companies first, and will do so in medium-sized companies, later.

Thus, European enterprises are beginning to realize the importance of ES for improving the internal work process and the efficiency of the company in general. However this has not yet resulted in making a straightforward business case for ES.

More information about what an ES solution can offer (how it differs from web search, and what the benefits are for potential clients) is needed, particularly for those companies that have no internal search experts or IT development groups, which are the vast majority of European companies. Organisations usually have no research which identifies the most important tasks carried out by employees and the extent to which enterprise search would improve operational effectiveness. More effort in this direction, possibly supported by the Commission, would be beneficial.

Many client companies still harbour wrong perceptions of the services that ES can or cannot offer. Important differences between the requirements of searching web sites and searching other enterprise applications, primarily to do with security management, need better explanation. Clarification is also needed on the extent to which the same enterprise search application can be used for both purposes. A realistic assessment of and guidance on ES tools is the key to improved efficiency for customer enterprises.

**The growth potential for ES is good but the market remains largely unexplored.**

Unfortunately there are neither reliable data nor revenue analyses for the EU enterprise search market, but we carried out some indicative estimates ourselves by triangulation of different sources:

- An estimate of the global market for enterprise search software based on the published market information from leading analysts would be €2 billion. This represents less than 2% of global enterprise search sales.
- The largest EU search software company is Exalead, which was acquired by Dassault Systems in 2010, but there is no dominant supplier of search software in the EU.
- There is a substantial number of US enterprise search vendors who have representation in Europe. IBM, Oracle and Google strongly target the EU market.

- The total revenues of EU-headquartered search vendors are between €100 million and €200 million depending on the view taken of Autonomy search revenues. Sales by IBM, Oracle, Google and Microsoft in the EU could be of the order of €300 million.
- This would give a total market in the EU of €500 million at most, or 25% of the global market.

From a policy point of view, the message is that ensuring a solid European approach on ES is vital for the competitiveness of ES industry, particularly in these times of market consolidation.

### **Europe needs to improve the supply of ES into business.**

Several studies show the economic value of rapid and reliable access to internal and external company information. This value is both in terms of opportunity loss (e.g. time spent on searching rather than on more productive work) and of added value (e.g. identifying useful relationships – for example to make a better offer to clients, detecting inconsistencies in internal data, etc.). A higher use of search applications and other state-of-the-art business process tools is likely to result in higher productivity of the individual companies. Given its systemic nature, there are generalized benefits across all industrial sectors. The impact of ES will be higher, the more IT-intensive the industry is, and these are generally the most dynamic knowledge intensive sectors. Consequently, a wider use of effective ES tools is likely to increase competitiveness of European companies as a whole.

### **The need to furnish industry-specific applications and complexity to provide these to customers are obstacles to the take-off of Enterprise Search business.**

With respect to the web search, the value chain of enterprise search business is more complex. In addition, established commercial products are facing the pressure from open-source software.

From the perspective of an organisation, we distinguish seven different ways in which an enterprise search application can be procured and implemented. In the realm of commercial products, ES can be provided by a) a system integrator, b) a specialist integrator, c) a commercial software company (which may also provide service support for implementation) and d) an application developer. In the case of open source, products can be provided either by a) system integrator or b) specialist integrator.

The business models for the above cases differ, so that this distinction is indispensable to making sound assumptions about the future development of the enterprise search business. Unfortunately, such fine tuning is beyond the scope of this report, but broader generalizations about the future development of the enterprise search business in the EU still have value.

### **More awareness and trust needed for adopting ES Applications.**

When procuring an enterprise search application, an IT manager's primary source of information is the web sites of vendors on the Internet. Given the lack of a complete and reliable compendium of vendors, this favours large players who can afford publicity. Many European ES companies, however, are small and often only known in specialist circles in their own country. Despite their dynamism and expertise, they cannot promote their products efficiently throughout the EU27. On the other hand, potential customers would find it quite time consuming to identify these providers, and even if they were to do so, they might lack confidence in small ES providers operating in a different country.

Confidence and trust in the solution provider is generalized amongst all demand side, particularly for newcomers that can hardly assess the cost-to-benefit ratio of the ES investment or the quality of the product provided. The latter is even more aggravated in the case of open source solutions where customers must feel confident in the skills of the development team. Therefore, it is understandable that *Lucid Imagination*, amongst the leaders in Lucene/Solr development, recently introduced a certification scheme for developers. This way they expect to increase significantly the confidence that companies have in Lucene/Solr-based applications.

A related issue is that few IT departments have the expertise and experience to write tender documents for ES applications. Here, a potential European Commission measure could be to provide guidelines on how to define and procure ES applications.

### **Further ES technology development is desirable but not the main obstacle to the success of ES companies**

According to our Delphi-type study, the technologies that are regarded as most important to the demand and adoption of ES are **Search-Based Applications** and integrated search platforms (unified access platform). Integrated platforms and search-based applications (SBA) are key solutions because these hold the capability to provide semantic linking (combining structured and unstructured data) and semantic search (allowing intelligent analysis of query).

SBA can rely on faceted search (facets are visualizations of semantics that users can understand), on semantic databases and on search engines. Given that search engines can handle the semantics of databases, that facets allow for Business Intelligence type reporting, SBA can use the power of search engines (intuitive, scaling, agility) to extract and merge information from databases and text<sup>10</sup>.

Despite the need to improve technological building blocks, it is not the absence of effective technologies that prevents ES from taking off. Stakeholders and experts do not consider a lack of specific ES technology to be the limiting factor for the take-up of enterprise search. This does not mean that there are no technological challenges. On the contrary, improvements in re-use of components or better interoperability are examples of missing elements that would be beneficial for the development in the short to medium term. In the long-term, the growth of semantic web applications and of the 'Internet of things' will impact on ES, both in terms of new applications and in terms of innovative business models to be developed. Finally, the demand for visual enterprise search, though still quite limited and applied only to specific scopes, is growing, as enterprises have more and more multimedia assets.

One of the future trends and challenges for ES we identified is in the "domain-specific semantics" - that is, the development of the intelligent query interpretation (intelligent analysis of query to extract relevant terms).

The use of Enterprise search is likely to be propelled by mobile applications. In our technology forecast 2011-2015, based on the Delphi study, respondents considered mobile as the key enabling technology in the future of enterprise search.

Particularly, ES vendors believe that customer companies will prefer to pay for extending the capabilities of their current search applications to mobile devices. They consider that mobile components would be particularly attractive for those customer companies that have high numbers of employees on the move (e.g. salesmen). Offering employees off-site with updated information (e.g. to submit instantaneously an offer to customers) will give these innovative companies a competitive advantage.

### **ES in the future: cloud-based and user-demand approach, open data models, interoperability.**

Cloud computing is a driver in many IT applications and it is worth discussing the opportunities and limitation for ES. Cloud-based approaches are where the world is heading: it is supposed to be gradual transitions and not necessarily a generalized phenomenon across all enterprises. Some pioneering companies will quickly transfer most of their data assets to the cloud for reasons of costs and accessibility. And the more companies transfer value (sensitive) data to the cloud, the more the search function will be useful. Currently, most companies, however, are still dependent on their legacy of data systems. Likewise, enterprise search vendors' businesses rely on the current situation that systems need to be deployed on-site at the customer premises, and integrated with

---

<sup>10</sup> G. Grefenstette, L. Wilber, *Search-Based Applications*, Morgan & Claypool, 2011

existing legacy systems. The world is changing in this respect and a scenario could be envisaged whereby valuable data will reside simultaneously in legacy repositories and in new cloud-based repositories. This future scenario will require technology options with open data models, interoperability and standards, which are different from current closed systems and proprietary interfaces.

Other socio-technical challenges relate to improved user interfaces and feedback from user experience delivered. Today there are already some good techniques and enterprise search technologies, but the user experience delivered is unfortunately not yet in the focus area. So far, user interfaces are driven by “what technology can offer” (technology-push), and too little by user-demand. End-users want reliable and useful results, delivered in a user-friendly way, where the efficiency gain becomes evident. This can be achieved only in a cross-disciplinary manner where best practices, use demands and design aspects are integrated in research projects right from the start.

### ***Need to stimulate the potential of Enterprise Search***

There are positive messages with regard to opportunities to further stimulate the potential of ES. First, our analysis confirms a solid base of world-class vendors in the EU.

Second, there is significant information retrieval research capability in Europe, both academic and industrial. There is the potential to exploit these research and development results as products and services.

EU-based vendors have a competitive advantage in that they understand the implications of multiple language issues and provide added-value when implementing software. This has allowed the co-existence and a symbiosis of large and smaller players.

An additional strength in Europe is the extensive knowledge base with regard to Semantic Web technology and Linked Open Data, and also open source. The latter –although currently small in market terms– may evolve as a stable niche especially for small companies in need of low-cost solutions. Many of the required technologies are available as open source software and therefore companies may be attracted by opportunities to start new open source projects in this area.

Our study has collected statements from European customer enterprises indicating that some are more willing to contact vendors based in the US than vendors of the same provider in Europe. Worse still, they may prefer to talk to US subsidiaries of EU headquartered vendors. This appears to be a recurrent pattern rather than anecdotal and the question arises as to why this happens. One explanation might be – as explained before – the general difficulty in identifying search vendors and integrators in Europe. Another reason may be the limited interaction between IR and ES communities on this side of the Atlantic. Moreover, it seems that SMEs perceive US companies to be more trustworthy than companies in a different EU Member State.

Another –more worrying– explanation is the perception that innovative products are first deployed in the US. The underlying argument is that when developing a new ES solution, the first version is targeted to the US market, given that the market is big and homogenous and investments are more quickly capitalized than in a fragmented European market. Once the first products have been deployed in the US, they are then progressively introduced into the EU. The fragmentation of the European market and the lack of confidence in EU-developed software products are not specific to the ES domain, though they add to ES specific organizational and operational obstacles. In particular, enterprise search needs to find its place in a wider market context, namely in an interaction and integration perspective with the rest of the enterprise system.

On the one hand, some industry representatives complain that EU support to ES projects is too academic, and lacks proper commercialization. They consider that EU-supported ES projects ought to strive for clearer and more measurable and verifiable business benefits, and possibly, have an enterprise sponsor from the beginning. This could be ensured by requiring that the enterprise

customer become more involved: they should be, encouraged to contribute financially to the solution to be developed and participate from the beginning in the development of the project. The enterprise's level of interest in development projects is a measure of their importance for the market. On the other hand, some ES experts consider that EU-supported projects, so far, have delivered many useful research results based on established retrieval concepts but have contributed very little to the development of radically new search concepts. They consider governments should invest more in high-risk / high-return projects, rather than in conservative projects with more predictable returns.

ES professionals acknowledge networking as an enabler for better understanding "the user perspective" (what does the market need and expect?) and for establishing partnerships. Given the limited occasions at European level they suggest maintaining and expanding existing platforms. During the ES workshop in Seville it was suggested that a permanent industrial and user community of actors should be established in the field.

One of its roles would be to give guidance and support to benchmarking and dataset creation. This would increase the number of reliable and extensive data collections, which are sadly missing for benchmarking research results and prototypes. Existing datasets are limited; benchmark exercises (like TREK) are few and mostly coordinated by experts outside the EU. One way to gather such datasets could be the establishment of 'guinea pig' organizations for testing enterprise search applications or the acquisition of data from bankrupt companies. In addition, a permanent platform of this kind would help to advance user testing and user interface.

Another objective would be to promote interoperability standards. This can only be achieved with strong support from the industry. One challenge to be faced is the industry structure: many ES providers are SMEs and can hardly afford to invest in standards, so the risk is that the few large players may impose de facto standards.

## **Outlook**

The *raison d'être* for ES remains the struggle that having to manage ever increasing amounts of data represents for companies and their need to retrieve relevant data efficiently. ES offers a solution to this struggle and will therefore accompany us for a long time.

From an industrial policy point of view, the prime interest is to propel adoption of efficiency-enhancing technologies in all European enterprises in order to increase Europe's economic competitiveness as whole. ES is certainly an "enhancing technology" and questions of whether to take policy measures - and if so, which ones - arise. Here we have to differentiate between two levels of potential activity. One is to assess the importance of ES as an industrial sector *per-se*; the other is the impact of ES on other industrial sectors.

With regard to ES as an industrial sector, we have observed two major waves of market consolidation and it seems that the mergers and acquisitions are not yet complete. If we extrapolate the trajectories of recent years, it seems likely that many more dynamic smaller players will disappear, because they are acquired by larger companies (type 1) or they cannot resist market pressure. The question arises whether it is better to let the market consolidate on its own or take active steps and specific actions to influence the consolidation. The second case would be justified if ES would be considered highly strategic for Europe's economy. One option would be to promote specific research programs for smaller ES companies (specialized on segment of ES or on special tools) to develop further their technologies and research capacity in order to be able to compete with larger players, or being able to collaborate with these in more favourable terms. This might reduce the probability being acquired by non-European 'big fishes', and eventually to operate in collaboration with bigger European companies (or integrated within them).

To sum up:

- 1) Enterprise Search helps to manage the increasing data growth and contributes to effectiveness and economic development, but there is still low awareness about its benefits.
- 2) The growth potential for ES is good but the market remains largely unexplored.
- 3) Europe needs to improve the supply of ES into business.
- 4) The need to furnish industry-specific applications and complexity to provide these to customers are obstacles to Enterprise Search business taking off.
- 5) Skill gaps, divergent opinions amongst practitioners and academics, and low investments limit the speed of ES adoption.
- 6) More awareness and trust is needed if ES Applications are to be adopted.
- 7) Further ES technology development is desirable but its lack is not the main obstacle to the success of ES companies.
- 8) ES in the future will be cloud-based and the focus will be on a user-demand approach, open data models, and interoperability.





## Appendix A: List of enterprise search vendors – alphabetical

Company	Country	URL
Active Navigation	UK	<a href="http://www.activenav.com">http://www.activenav.com</a>
Amazon	USA	<a href="http://a9.com/">http://a9.com/</a>
Ankiro	Denmark	<a href="http://www.ankiro.com">http://www.ankiro.com</a>
Apache (Lucene/Solr)	Community	<a href="http://lucene.apache.org">http://lucene.apache.org</a>
Applied Relevance	USA	<a href="http://www.appliedrelevance.com">http://www.appliedrelevance.com</a>
Attensity	USA	<a href="http://www.attensity.com">http://www.attensity.com</a>
Attivio	USA	<a href="http://www.attivio.com">http://www.attivio.com</a>
Autonomy (HP)	UK	<a href="http://www.autonomy.com">http://www.autonomy.com</a>
BA-Insight	USA	<a href="http://www.bainsight.com">http://www.bainsight.com</a>
Basis	USA	<a href="http://www.basistech.com">http://www.basistech.com</a>
Brainware	USA	<a href="http://www.perceptivesoftware.com">http://www.perceptivesoftware.com</a>
Clearwell	USA	<a href="http://www.clearwellsystems.com/">http://www.clearwellsystems.com/</a>
Cognition Systems	USA	<a href="http://cognition.com">http://cognition.com</a>
Commvault	USA	<a href="http://www.comvault.com">http://www.comvault.com</a>
Concept Searching	UK	<a href="http://www.conceptsearching.com">http://www.conceptsearching.com</a>
Constellio	Canada	<a href="http://www.constellio.com">http://www.constellio.com</a>
Coveo	Canada	<a href="http://www.conveo.com">http://www.conveo.com</a>
Dieselpoint	USA	<a href="http://www.dieselpoint.com">http://www.dieselpoint.com</a>
Digital Reasoning	USA	<a href="http://digitalreasoning.com">http://digitalreasoning.com</a>
dtSearch	USA	<a href="http://www.dtsearch.com">http://www.dtsearch.com</a>
EMC	USA	<a href="http://www.emc.com">http://www.emc.com</a>
Elastic Search	Community	<a href="http://www.elasticsearch.org">http://www.elasticsearch.org</a>
Endeca	USA	<a href="http://www.oracle.com">http://www.oracle.com</a>
Exalead	France	<a href="http://www.3ds.com/products/exalead">http://www.3ds.com/products/exalead</a>
Exorbyte	USA	<a href="http://www.exorbyte.com">http://www.exorbyte.com</a>
Expert System	Italy	<a href="http://www.expertsystem.net">http://www.expertsystem.net</a>
Fabasoft	Austria	<a href="http://www.mindbreeze.com/en">http://www.mindbreeze.com/en</a>
Forward Search	Denmark	<a href="http://www.forwardsearch.dk">http://www.forwardsearch.dk</a>
Funnelback	Australia	<a href="http://www.funnelback.com">http://www.funnelback.com</a>
Google	USA	<a href="http://www.google.com/enterprise/search/products_gsa.html">http://www.google.com/enterprise/search/products_gsa.html</a>
IBM	USA	<a href="http://www.ibm.com">http://www.ibm.com</a>
Inbenta	Spain	<a href="http://www.inbenta.com">http://www.inbenta.com</a>
Infosys	USA	<a href="http://www.infosys.com">http://www.infosys.com</a>

InQuira	USA	<a href="http://www.oracle.com">http://www.oracle.com</a>
Intelligenx	USA	<a href="http://www.intelligenx.com">http://www.intelligenx.com</a>
Intrafind	Germany	<a href="http://www.intrafind.de">http://www.intrafind.de</a>
ISYS	Australia	<a href="http://www.perceptivesoftware.com">http://www.perceptivesoftware.com</a>
Karmasphere	USA	<a href="http://www.karmasphere.com">http://www.karmasphere.com</a>
Lexalytics	USA	<a href="http://www.lexalytics.com">http://www.lexalytics.com</a>
LTU	France	<a href="http://www.ltutech.com">http://www.ltutech.com</a>
Lucid Works	USA	<a href="http://www.lucidworks.com">http://www.lucidworks.com</a>
Mark Logic	USA	<a href="http://www.marklogic.com">http://www.marklogic.com</a>
MaxxCAT	USA	<a href="http://www.maxxcat.com">http://www.maxxcat.com</a>
Microsoft	USA	<a href="http://www.microsoft.com">http://www.microsoft.com</a>
Omniture	USA	<a href="http://www.omniture.com">http://www.omniture.com</a>
OpenSearchServer	France	<a href="http://www.open-search-server.com">http://www.open-search-server.com</a>
OpenText	Canada	<a href="http://www.opentext.com">http://www.opentext.com</a>
Oracle	USA	<a href="http://www.oracle.com">http://www.oracle.com</a>
Perfect Search	USA	<a href="http://www.perfectsearchcorp.com">http://www.perfectsearchcorp.com</a>
Polyspot	France	<a href="http://www.polyspot.com">http://www.polyspot.com</a>
Q-Sensei	USA	<a href="http://www.qsensei.com">http://www.qsensei.com</a>
Recommind	USA	<a href="http://www.recommind.com">http://www.recommind.com</a>
SAP	Germany	<a href="http://www.sap.com">http://www.sap.com</a>
SearchBlox	USA	<a href="http://www.searchblox.com">http://www.searchblox.com</a>
SearchDaimon	Sweden	<a href="http://www.searchdaimon.com">http://www.searchdaimon.com</a>
Sinequa	France	<a href="http://www.sinequa.com">http://www.sinequa.com</a>
Smart Logic	UK	<a href="http://www.smartlogic.com">http://www.smartlogic.com</a>
Sophia Systems	UK	<a href="http://www.sophiasearch.com">http://www.sophiasearch.com</a>
Sphinx	Community	<a href="http://sphinxsearch.com">http://sphinxsearch.com</a>
Stored IQ	USA	<a href="http://www.storediq.com">http://www.storediq.com</a>
SurfRay	Denmark	<a href="http://www.surfray.com">http://www.surfray.com</a>
Synaptica	USA	<a href="http://www.synaptica.com">http://www.synaptica.com</a>
Temis	France	<a href="http://temis.com">http://temis.com</a>
Teragram	USA	<a href="http://www.teragram.com/oem">http://www.teragram.com/oem</a>
TeraText	USA	<a href="http://www.teratext.com">http://www.teratext.com</a>
Terrier	UK	<a href="http://terrier.org">http://terrier.org</a>
Thetus	USA	<a href="http://thetus.com">http://thetus.com</a>
Thunderstone	USA	<a href="http://www.thunderstone.com">http://www.thunderstone.com</a>
Vivisimo	USA	<a href="http://www.vivisimo.com">http://www.vivisimo.com</a>
Wand	USA	<a href="http://wandinc.com">http://wandinc.com</a>

Xapian	Community	<a href="http://xapian.org">http://xapian.org</a>
X1 Technologies	USA	<a href="http://www.x1.com">http://www.x1.com</a>
ZyLab	USA	<a href="http://zylab.com">http://zylab.com</a>



## **Appendix B: Corporate profiles of selected enterprise search vendors**

In this Appendix more detailed profiles are provided of companies that illustrate the range of businesses that make up the enterprise search business<sup>11</sup>.

### ***Autonomy***

Autonomy was found in 1996 by Dr. Michael Lynch and Richard Gaunt as a spin-off from Cambridge Neurodynamics, a company formed to develop the use of Bayesian probabilistic approaches to information retrieval. The company was floated in 1998. Over the last ten years the company has grown to become the largest provider of search software outside the four major IT companies in this list of five companies. At the heart of Autonomy's infrastructure software is the Intelligent Data Operating Layer (IDOL) Server. The IDOL Server collects indexed data from connectors and stores them in a proprietary structure, optimized for fast processing and retrieval of data. It is an advantage for a very scalable search application but it comes with a high degree of complexity from an implementation and administration perspective.

Autonomy developed through a wide range of acquisitions, notably the purchase of Verity/Ultraseek in 2005. The result was a very complex network of operational entities, which increasingly saw Autonomy moving away from being a pure-play search vendor.

The announcement in August 2011 that HP intended to acquire Autonomy for a substantial premium over the share price changes the enterprise search market considerably. The justification by HP for the purchase was to enable the company to provide a wider range of business services, especially through cloud-based applications. At the same time as announcing the acquisition, HP also signalled its move out of webOS mobile applications and out of the PC business. HP indicated that one of the reasons for the acquisition was to capitalise on the OEM business that Autonomy had developed.

Even though HP have committed to manage Autonomy as a separate business, this seems to be in conflict with the need by HP to have a more integrated services business along the lines of IBM, Oracle and Microsoft. Overall this is a good deal for Autonomy shareholders, but the future benefits to HP or to Autonomy and its customers are far from clear at this time and a number of market analysts are quite sceptical about the extent to which the HP/Autonomy deal is going to be a future success story. The reaction from other search vendors is that the acquisition will significantly open up the market for them, as potential Autonomy customers may defer making a commitment until the positioning of Autonomy within HP becomes clear.

### ***Google***

Google is seen by the public and by many senior managers as defining search excellence. The Google product is a search appliance, which comes as a sealed standard server box that can be plugged directly into a server rack-space. Google is not unique in using a sealed search appliance route. Thunderstone, based in Cleveland, OH, also uses this approach, but much of the business of Thunderstone is in the provision of OEM search applications for large publishing companies.

Much of the search performance of Google comes from the use of the Page-Rank algorithm and the scalability and performance of its Big Table database technology. In an enterprise application the Page-Rank approach is not valid as there are relatively few web pages that can be used as reference points. It is interesting to note that the Google GSA search appliance is not very much promoted by Google on the site that offers business applications.

The link to enterprise solutions is given relatively little prominence and even the search page itself promotes a range of Google search products

---

<sup>11</sup> More extended profiles of Autonomy, Exalead, Google, Microsoft and Vivismo can be found in The Landscape of Enterprise Search by Stephen Arnold, published by Pandia in 2011. Details of this report can be found at <http://www.pandia.com/enterprise-search/>

There are two important aspects of the Google enterprise search marketing strategy. The first is that the license is based on the number of pages indexed and is only a two year agreement. This means that although the initial costs seem very low over a six year period the true cost of ownership may end up being higher than those of competitor products.

The second is that there is very limited technical support for the product. Google supply a locked-down appliance-based application. Although over the last few years Google have provided better administrative interfaces it is still very much a 'black box' application. Communications with Google support staff are predominantly through email, and this means that even quite simple problems may take time to resolve.

Google already employs about 5,000 people in Europe and announced recently that it expects to hire more than 1000 new employees across Europe. Much of its work across its facilities in Europe is focused on content ID for YouTube, search (including product search and insights for search) and Android, the mobile smartphone operating system.

### **Lexmark**

Lexmark was created in 1991 when IBM divested its printer business. In 2010 Lexmark acquired Perceptive Software, a US company providing document management and business process management applications. In 2012 Lexmark acquired Brainwave (which had developed data capture technology) and Isys-Search, the Australian-owned enterprise search vendor. It seems likely that Perceptive Software will acquire additional businesses in the information management sector.

### **Microsoft**

Microsoft owns the corporate desktop, and given the wide use of SharePoint 2007 and 2010 it will be for many companies their first experience of an enterprise search application. For corporations looking for an enterprise search solution, Microsoft offers plenty of options: Bing, SharePoint Foundation 2010 Search, Search Server Express, Search Server 2010, SharePoint Server 2010 and FAST Search Server 2010 for SharePoint. The various search alternatives vary in capabilities, sophistication and price.

The pricing is based on the concept of a Customer Access Licence. In addition to the software server licence, customers need to purchase a CAL for each user (of around €80). This licence is for the Search Server 2010 application only. The FAST Search Server for SharePoint 2010 offers enhanced functionality, but requires the purchase of an Enterprise CAL per user, which is an additional €70. In addition a FAST server licence is required.

The FAST ESP product, which is a very-high functionality search application is not a component of SharePoint 2010. There is very little information on the FAST ESP product available on the Microsoft web site, which might indicate that this sector of the business is not one that Microsoft has a long-term commitment to developing.

To gain market share, Microsoft is following well-established pattern of finding ways to commoditize products in markets first developed by other companies. Historically search systems have been complex, costly and difficult to implement. Consequently, their use has been limited to large enterprises. Rather along the lines of Google the approach that Microsoft has adopted has been to lower the cost of enterprise search by an order of magnitude and push search procurement decisions down to enterprise departments and smaller companies.

One market force that is difficult to assess is the impact of Microsoft SharePoint 2010 search products on the market. The penetration of SharePoint seems to be substantial in larger companies in Europe, and Microsoft is promoting quite heavily the benefits of the FAST Search Server for Microsoft SharePoint 2010. In a presentation at the Microsoft SharePoint Conference 2011 in Anaheim on 3-5 October 2011 Microsoft announced that 65,000 organisations were using SharePoint, including 80% of Fortune 500 companies.

A recent survey by AIIM<sup>12</sup> indicates that only 7% of respondents were using SharePoint 2010 search both within SharePoint itself and also as an enterprise search application. In practice using Fast Search Server for SharePoint 2010 to search content outside of the SharePoint repositories requires considerable development.

To index what are often referred to as Line-Of-Business (LOB) applications connectors have to be used to convert the file formats used in the LOP application into a format that FAST Search for SharePoint 2010 can then query. Setting these connectors up and optimising them requires significant experience of both BCS and the LOB application.

Microsoft has campuses in the UK, Ireland, Denmark, Norway, Poland, Russia, Germany and France, working on product development and research to support its IT consulting services business.

### **IBM**

IBM was one of the pioneers in full-text retrieval technology with in the 1960s with the STAIRS search application. Over the last few years IBM has moved its core Omnifind search application to a Lucene/Solr code base that the company initially tested as the IBM Omnifind Yahoo Edition. With effect from April 2011 IBM OmniFind Enterprise Edition V9.1 and IBM Content Analytics V2.2 are packaged into a single bundled offering the so called IBM Content Analytics with Enterprise Search. The IBM product web site contains very little information about either the enterprise level product or the Enterprise Starter Edition which is optimised for Lotus Domino and Lotus Notes applications. Although in principle both products are available on a stand-alone product basis the majority of sales come through enterprise contracts with large organisations.

IBM has substantially reduced its interests in information retrieval research over the last few years. It is now the responsibility of the IBM research centre in Haifa, Israel to conduct research.

### **Oracle**

Oracle offers its Enterprise Secure Search application as a standalone product at €28000 per processor. As with IBM comparatively little information about the product is available from the Oracle product web site, but the product can be downloaded from the site. However it is more usually provided to customers through an enterprise agreement with other Oracle products and services.

Oracle has an Information and Machine Learning research group with interests in:

- Core Relevance: Determining what the most relevant set of documents are for a given query.
- Application-Specific Search: Using the properties of a given application, such as email, to affect how search is implemented for that application.
- Document Classification: Learning how to assign documents to classes.
- Named Entity Extraction: Extracting the names of people, organizations, places, etc. from text.
- Query Classification: Deciding how to respond to a query by trying to understand what the user is asking for.

This group has also been responsible for the continuing development of the Java-based Minion search engine that was initially developed by Sun before its acquisition by Oracle.

In October 2011, Oracle acquired Endeca, a company that specialises mainly in e-commerce search using faceted navigation. The impact on the enterprise search sector will be minimal.

---

<sup>12</sup> <http://www.aiim.org/sharepoint/research/State-of-the-ECM-Industry-2011>.

## **Apache Foundation**

The Apache Lucene project develops open-source search software, including:

- Apache Lucene Core (formerly named Lucene Java) provides a Java-based indexing and search implementation, as well as spellchecking, hit highlighting and advanced analysis/tokenization capabilities.
- Apache Solr is a high performance enterprise search server, with XML/HTTP and JSON/Python/Ruby APIs, hit highlighting, faceted search, caching, replication, distributed search, database integration, web admin and search interfaces.
- Apache PyLucene is a Python port of the Lucene Core project.
- Apache Open Relevance Project is a subproject with the aim of collecting and distributing free materials for relevance testing and performance.

Lucene was integrated into the Apache Software Foundation's Jakarta family of open source Java products in September 2001 and became its own top-level Apache project in February 2005.

A list of applications based on Lucene is published by the Foundation



### Appendix C: Delphi summary tables

For the aim of this report, a Delphi-type survey was undertaken, with the questionnaire being sent to all the participants in the Workshop as well as to some other experts in the UK and the USA. Eighteen replies were received, eight from search vendors, five from integration companies and five from experienced industry analysts.

In a number of cases there was a significant difference between the scores from vendors, analysts and integrators. In general the integrator viewpoint was a very practical one, coping with the challenges of helping companies select and implement enterprise search applications. Vendors tended to take a more middle-distance view. Understandably their responses aligned to their business strategy. The analysts took a 'helicopter view' probably seeing a bigger picture but not perhaps based on direct implementation experience.

Respondents to the survey were asked to rank the extent to which they supported each statement by allocating points from a total of 20. The scores were then averaged between each of the three categories, and also across all the responses. The process of averaging inevitably hides the range of response scores from individual experts. However putting a range on the averages is not appropriate given the small sample size and the different experience of each of the consultants

The responses from the initial round were sent out to all respondents asking if there were any outcomes that they felt were significantly at variance with their own views. None of the respondents raised any concerns about the results.

In the tables a rank of the top four responses for each topic is shown. Below this level the scores were sufficiently low that to continue the ranking beyond this point is not valid.

This approach also shows the differences between the responses of the three groups of respondents.

In the Delphi survey participants were asked for their views on what they regarded as the main barriers for vendors to selling an initial enterprise search application.

	All	Vendor	Analyst	Integrator
There is a lack of awareness of the functionality of enterprise search applications	1	1=	3	1
Organisations find it difficult to make a business case	2		1=	2
The view is that Google is the best search application and offers what is needed	3		1=	3
There is a lack of internal expertise to define requirements	4=	1=		
There is a lack of internal expertise to support the implementation	4=		4	4
The view is that SharePoint offers the required enterprise search functionality	4=	4		
Enterprise search is regarded as a low-priority investment		3		
Information is not seen as a business asset				

The views of the respondents of the Delphi survey indicates that this dissatisfaction is a strong driver for replacing the current search application

	All	Vendor	Analyst	Integrator
User dissatisfaction with the performance of the search application	1	1	1	1
The need to reduce on-going development and maintenance costs	2	2	2	2
Acquisition of the current vendor or concern about their long-term stability	3	3	4	3
Action taken by a business competitor	4	4	3	4

Another question asked what the main future direction would be of enterprise search.

	Total	Vendor	Analyst	Integrator
Current market players will extend their product range outside of enterprise search	1=	2=	2	2=
Open source applications will take an increasing proportion of the market	1=	2=	4=	1
Some smaller vendors may not be able to remain in business	1=	4	1	2=
There will be increasing vertical specialisation by search vendors	4	1	4=	
Software and systems houses will buy a search vendor to increase their breadth of offering				4
There will be new entrants into the pure-play sector offering novel technologies			3	
There will be market consolidation through acquisition by search vendors				
Business intelligence vendors will move into enterprise search				
There will be no significant changes to the market				

Respondents were asked what would become increasingly important in the selection of an initial enterprise search application.

	All	Vendor	Analyst	Integrator
Ease of implementation of the application	1	1	1=	1
Brand strength and visibility of the vendor	2	3	1=	2
Technical features linked to business requirements	3	2	3	4
The ability to predict the total cost of implementation at the outset	4=	4		
A recommendation by a systems integrator	4=		4	
Minimal support needed to support the application after implementation	4=			3
A recommendation by a business colleague				
Quality of the information on the vendor web site				

Enterprise mobile applications are emerging rapidly, and respondents were asked for their views on the impact and direction of mobile search technology.

	<b>Total</b>	<b>Vendor</b>	<b>Analyst</b>	<b>Integrator</b>
Companies will prefer to extend the capabilities of their current search application to mobile devices	1	1	1	1
Mobile search applications for enterprise use will stimulate new vendors to enter the market	2	3	4	2
Innovation in mobile search will result in innovation in workplace enterprise search	3		2	3
Companies will place mobile search requirements as a justification for replacing their current search application	4=	4	3	4
The requirement for mobile enterprise search is overstated	4=	2		

The final question asked for a ranking of technologies from a set developed by the IPTS Project Team that respondents regarded as of importance to the demand and adoption of enterprise search. Respondents were also invited to suggest technologies that were not in the original list but no additional technologies were highlighted.

The survey also asked for a view on the year when these technologies would be widely offered by most search vendors, but there was no consensus at all on a forecast date.

	<b>Total</b>	<b>Vendor</b>	<b>Analyst</b>	<b>Integrator</b>
Integrated search of structured and un-structured content	1	2=	2=	1=
Search as an integration platform (unified access platforms)	2			1=
Search-based applications	3=	1		1=
Text mining	3=	2=	1	
Cloud-based search	3=			
Enterprise mobile search applications	3=		2=	
Search incorporated into business intelligence applications	3=	4	4=	4
Multimedia search	3=		4=	
Semantic search	4		4=	
Expert search				
Tagging of search results				
Enhanced federated search			4=	
Being able to search multiple languages with a single query.				

The high position of text mining was of interest, as was the greater conviction of analysts that search-based applications would be an important development.



## Appendix D: Enterprise Search Industry Analysis Consultancies

The enterprise search business and its markets are tracked by a number of consulting companies.

A list of these is given below:

Consultancy	URL
Accenture	<a href="http://www.accenture.com">http://www.accenture.com</a>
ArnoldIT	<a href="http://www.arnoldit.com">http://www.arnoldit.com</a>
Findwise	<a href="http://www.findwise.com">http://www.findwise.com</a>
Forrester	<a href="http://www.forrester.com">http://www.forrester.com</a>
Gartner	<a href="http://www.gartner.com">http://www.gartner.com</a>
Gilbane	<a href="http://gilbane.com">http://gilbane.com</a>
International Data Corporation	<a href="http://www.idc.com">http://www.idc.com</a>
Intranet Focus	<a href="http://www.intranetfocus.com">http://www.intranetfocus.com</a>
New Idea Engineering	<a href="http://www.newideaeng.com">http://www.newideaeng.com</a>
Ovum	<a href="http://www.ovum.com">http://www.ovum.com</a>
Real Story Group	<a href="http://www.realstorygroup.com">http://www.realstorygroup.com</a>
The 451 Group	<a href="http://the451group.com">http://the451group.com</a>



## Appendix E: Workshop "Exploring the future of enterprise search"

A validation workshop on Enterprise Search prospects took place on the 13 and 14 October 2011 at the Institute for Prospective Technological Studies (IPTS), in Seville (Spain).

### Aim of the workshop

The aim of the workshop was to gain insights into the techno-economic and socio-economic trends in enterprise search market in the EU, and explore the impact of Enterprise Search on the European Economy and Society.

The workshop was organised to discuss and analyse the following aspects of enterprise search:

- **Market Dynamics:** Painting the landscape of enterprise search including current and emerging business models of providers of enterprise search solutions and services and their respective value chains.
- **Future Prospects:** Identifying emerging techno-economic trends, discussing likely developments and the future market structure in the domain of enterprise search
- **SWOT Analysis:** Exploring the strengths, weaknesses, opportunities and threats (SWOT) for the Europe Union with respect to enterprise search
- **Policy Options:** Is there any need of policy changes or new policies, or support actions in the field of enterprise search in the EU and if yes, what should these be?

### Agenda of the workshop

<i>Thursday 13 October</i>	
13:30 – 14:00	<b>Registration</b>
14:00 – 14:25	<b>Welcome and Introduction</b> <i>David Broster (IPTS, EC), Ramón Compañó (IPTS, EC), Stavri Nikolov (IPTS, EC)</i>
14:25 – 15:00	<b>Invited keynote speech</b> Enterprise Search and Search Based Applications <i>Gregory Grefenstette (Exalead)</i>
15:00 – 15:50	<b>Session 1: Technologies, solutions and research issues (50min)</b> Semantic Enterprise Search and Content Intelligence <i>Matthieu Jonglez (Smartlogic) (10min)</i> Open Source Techniques Push Enterprise Search & Search Driven Applications, Especially the Application of Text Analytics <i>Christoph Goller (IntraFind) (10min)</i> An Approach to Overcoming Obstacles to Enterprise Search Research <i>David Hawking (Funnelback) (10min)</i> Visual Search for Enterprise Documents <i>Mika Könnölä (Documill) (10min)</i> Session 1 Q&A (10min)
16:20 – 17:10	<b>Session 2: Services, applications and business models (50min)</b> Cloud-based Enterprise Search: Future or Reality? <i>Kathrine Hammervold (Microsoft) (10min)</i> Google as a Role Model for an Enterprise Search Team <i>Harald Kirsch (Raytion) (10min)</i> Visual Asset Search: Application Areas and Business Models <i>Ray Owens (LTU Technologies) (TBC) (10min)</i>

	Technologies for Mobile Application Development <i>Andrew Terry (Toumetis) (10min)</i> Session 2 Q&A (10min)
17:30 – 18:20	<b>Session 3: Techno-economic aspects of enterprise search (50min)</b> Enterprise Search – A Gap Analysis? <i>Martin White (Intranet Focus) (10min)</i> Search and the Contraction of Markets - Lessons Learned from the Catholic Church, Railroad Tycoons, and Oil Magnates <i>Hendrik Speck (University of Applied Sciences Kaiserslautern) (10min)</i> The Future for Enterprise Search in Europe Based on Altered Demographics <i>Henrik Strindberg (Findwise) (10min)</i> Visioning Enterprise Search in Future Enterprises <i>Man-Sze Li (IC Focus) (10min)</i> Session 3 Q&A (20min)
18:20 – 18:30	<b>First Day Concluding Remarks</b> <i>Stavri Nikolov (IPTS, EC), Henri Gouraud (INRLA)</i>
21:00 – 23:00	<b>Enterprise Search Think Thank dinner (by invitation only)</b> <i>Pieter van der Linden (Technicolor)</i>
<b>Friday 14 October</b>	
9:00 – 9:15	<b>Summary of First Day</b> <i>Henri Gouraud (INRLA), Stavri Nikolov (IPTS, EC)</i>
9:15 – 10:40	<b>Workshop and Think-Tank Roundtable discussion of the future trends and directions of Enterprise Search (covering technological and techno-economic aspects)</b> <i>Stavri Nikolov (IPTS, EC), Henri Gouraud (INRLA) &amp; Pieter van der Linden (Technicolor), and all previous speakers</i> What have we learned: technologies and business models
11:00 – 12:30	<b>Workshop and Think-Tank Roundtable discussion of the future trends and directions of Enterprise Search (covering technological and techno-economic aspects)</b> <i>Stavri Nikolov (IPTS, EC), Henri Gouraud (INRLA) &amp; Pieter van der Linden (Technicolor), and all previous speakers</i> What does this mean for Europe?
12:30 – 13:00	<b>Summary of Second Day/Workshop and Close</b> <i>Henri Gouraud (INRLA), Pieter van der Linden (Technicolor), Stavri Nikolov (IPTS, EC)</i>



## References

- Ambient Findability  
Peter Morville. 2005 [O'Reilly Publishing](#)
- The Answer Machine  
Susan Feldman. 2012 [Morgan and Claypool](#)
- Apache Solr 3.1 Cookbook.  
Rafa Ku. 2011. [PACKT Publishing](#)
- Book of Search.  
Kjetil Halvorsen et al 2006. Fast Search and Transfer (No longer available)
- Designing Search. UX Strategies for eCommerce Success.  
Greg Nudelman, 2011. [Wiley Publishing](#)
- Designing the Search Experience.  
Tony Russell-Rose and Tyler Tate, 2012 . [Morgan Kaufmann](#)
- Economic Trends in Enterprise Search Solutions.  
Pierre-Jean Benghozi and Cécile Chamaret. 2010. [Institute for Prospective Technological Studies](#),  
Joint Research Centre, European Commission
- Enterprise Search  
Martin White, 2012. [O'Reilly Media](#)
- Enterprise Search in Action.  
Mark Teutelink. 2011. [Manning Publications Company](#)
- Enterprise Search Vendor Evaluations.  
Version 4.0. 2011. [The Real Story Group](#).
- Faceted Search.  
Daniel Tunkelang. 2010. [Morgan and Claypool Publishers](#)
- Feature Centric View of Information Retrieval.  
Donald Metzler. 2011. [Springer](#)
- Findwise, Enterprise Search and Findability Survey, 2012, <http://www.findwise.com/about-us/enterprise-search-and-findability-survey>
- Fundamentals of Predictive Text Mining.  
Sholom M. Weiss, Nitin Indurkha and Tong Zhang. 2010. [Springer](#).
- Foundations of Large-Scale Multimedia Information Management and Retrieval.  
Edward Y. Chang. 2011. [Springer](#)
- Information Retrieval; Algorithms and Heuristics.  
2nd Edition. David A. Grossman and Ophir Frieder. 2004. [Springer](#)
- Information Retrieval Evaluation.  
Donna Harman. 2011. [Morgan and Claypool Publishers](#)
- Information Retrieval: Implementing and Evaluating Search Engines.  
Stefan Büttcher, Charles L. A. Clarke and Gordon V. Cormack. 2010. [Massachusetts Institute of Technology](#).
- Information Retrieval: Searching in the 21st Century.  
Ayse Goker and John Davies (Editors). 2009. [Wiley-Blackwell](#)
- Innovations in Information Retrieval: Perspectives for theory and practice.  
Allen Foster and Pauline Rafferty (Editors). 2011. [Facet Publishing](#)

Interactive Information Seeking, Behaviour and Retrieval.  
Ian Ruthven and Diane Kelly (Editors). 2011. [Facet Publishing](#)

Introduction to Information Retrieval.  
Christopher D. Manning, Prabhakar Raghavan and Hinrich Schutze. 2008. [Cambridge University Press](#)

Introduction to Modern Information Retrieval.  
G. G. Chowdhury. 2010. [Facet Publishing](#)

Lucene in Action.  
Michael McCandless, Erik Hatcher and Otis Gospodnetic. 2010. [Manning Publications](#)

Making Search Work.  
Martin White. 2007. [Facet Publishing](#)

Multilingual Information Retrieval: From Research to Practice.  
Carol Peters, Martin Braschler and Paul Clough. 2011. [Springer](#)

Multimedia Information Retrieval.  
Stefan R ger. 2009. [Morgan and Claypool Publishers](#)

Modern Information Retrieval: The Concepts and Technology behind Search.  
Ricardo Baeza-Yates and Berthier Ribeiro-Neto. 2nd Edition. 2011. [Addison Wesley](#)

New Landscape of Enterprise Search.  
2011. Stephen Arnold. [Pandia Search Central](#)

Professional Microsoft Search.  
Mark Bennett, Jeff Fried, Miles Kehoe and Natalya Voskresenskaya. 2011. [Wrox](#)

Prospects of Mobile Search.  
Jose Luis Gomez-Barroso et al. 2010. [Institute for Prospective Technological Studies](#), Joint Research Centre, European Commission

Search Analytics for Your Site.  
Louis Rosenfeld. 2011. [Rosenfeld Media](#)

Search-Based Applications: At the Confluence of Search and Database Technologies.  
Gregory Grefenstette and Laura Wilber. 2010. [Morgan and Claypool Publishers](#)

Search Engines: Information Retrieval in Practice.  
W. Bruce Croft, Donald Metzler and Trevor Strohman. 2010. [Addison Wesley](#).

Search Patterns.  
Peter Morville and Jefferey Callender. 2010. [O'Reilly Publishing](#)

Search User Interfaces.  
Marti A. Hearst. 2009. [Cambridge University Press](#)

Search User Interface Design  
Max Wilson. 2012. [Morgan and Claypool](#)

Semantic Software Technologies.  
2010. Lynda Moulton. [Gilbane Group](#)

Successful Enterprise Search Management.  
Stephen Arnold and Martin White. 2010. [Galatea Publishing](#)

Teaching and Learning in Information Retrieval.  
E. Efthimiadis, J. M. Fern ndez-Lunam J. F. Huete and A. MacFarlane. 2011. [Springer](#)

Text Mining: Applications and Theory.  
Michael W. Berry and Jacob Kogan (Editors) 2010. [Wiley-Blackwell](#)

Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data.  
Ronen Feldman and James Sanger. 2006. [Cambridge University Press](#)

Understanding Information Retrieval Systems: Management, Types, and Standards.  
Marcia J. Bates (editor). 2011. [Auerbach Publications](#)

Working with Microsoft FAST Search Server 2010 for SharePoint  
Mikael Svenson, Marcus Johansson and Robert Piddocke. 2012. [Microsoft Press](#)



European Commission

EUR 26000 – Joint Research Centre – Institute for Prospective Technological Studies

Title: Enterprise Search in the European Union: A Techno-economic Analysis

Authors: Martin White, Stavri G Nikolov

Luxembourg: Publications Office of the European Union

2013 – 87 pp. – 21.0 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN 1831-9424 (online)

ISBN 978-92-79-30493-4 (pdf)

doi:10.2791/17809

#### Abstract

The term '**enterprise search**' (ES) refers to the information retrieval applications that use a range of different core technologies to search enterprise repositories. It includes the search of the organisation's external web site, intranets and other electronic text held by the organisation in the form of email, database records, and documents on file shares. One reason for adopting Enterprise Search solutions (ES) is the **growth** in data generation; however, more worrying than the huge amount of information is its structure. It is clear that quick access to information is of strategic importance for enterprises and in general for the Information Economy. Although the issue is acknowledged as of extreme importance, only a small number of companies benefit from dedicated search technologies.

The **installed base** of enterprise search applications is still **low in the EU** (probably no greater than 10,000 organizations) and no dominant supplier of search applications exists. However, data suggest that there is a **significant market potential** in the EU for enterprise search and we can assume that the potential is considerable particularly for mid-sized companies that could benefit from using ES solutions. Moreover, the market for search applications is significantly larger than just the corporate sector and potential customer sectors would include Government departments and agencies, Hospitals and University.

These are some of the insights emerged from a Delphi study conducted by IPTS and Intranet Focus Ltd in 2011 and from further analysis about ES. This report builds also on the results of the expert workshop organized at the IPTS, Seville, on Enterprise Search in Europe in October 2011. In particular, a number of **challenges** have been identified that need to be addressed at European level. For instance, efforts to meet changing business requirements, the lack of support post-implementation, or the lack of a search support team. **Six important areas** of technical development for enterprise search over the next five years have been identified (e.g., Integrated search of structured and un-structured content) in which, there are evolutions of current search technologies and products. Moreover, **future trends in ES have been envisaged** (for instance, in cloud-based and user-demand approach, open data models, interoperability).

To our knowledge, this report contains the most **complete and detailed techno-economic analysis of the ES market in the EU** and includes an up-to-date list of existing ES vendors.

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new standards, methods and tools, and sharing and transferring its know-how to the Member States and international community.

Key policy areas include: environment and climate change; energy and transport; agriculture and food security; health and consumer protection; information society and digital agenda; safety and security including nuclear; all supported through a cross-cutting and multi-disciplinary approach.

