"Exploring the future of Enterprise search Workshop"

Summary Report

Last Update: November 29, 2011

Distribution Level: PU = Public

Abstract

The term of 'enterprise search' refers to technology enabling information retrieval within organizations. It includes the search of the organization's external web site, intranets and other electronic text held by the organization, such as emails, database record and documents on file sharing, often referred to as 'unstructured information'.

In order to analyze the related market of enterprise search and to understand, under a techno-economic perspective, the opportunities and challenges for Europe, the "Exploring the future of Enterprise Search" workshop was organized by IPTS in the framework of its Chorus+ work-plan 2011 tasks, in Seville, the 13-14 October 2011.

The scope of this document is to provide a summary report of the work done by IPTS in preparation of this workshop, of the workshop contents and of the main issues raised during the discussion.

The deliverable starts with a brief introduction, followed by the executive summary of the "Enterprise Search in the EU, A Techno-economic Analysis" report. This report, resulting from a six-months study on the current state of the enterprise search carried out by IPTS in collaboration with Intranet Focus Ltd and completed in October 2011, aimed at gaining insights of the techno-economic and socio-economic trends in enterprise search and of their likely impact on the European Economy and Society.

This report provides the most updated list of enterprise search companies as well as a comprehensive study of the main techno-economic trends in the field of enterprise search, identifying some of the challenges and barriers to future development. Recent trends were discussed and captured throughout the 2011 IPTS/Intranet Focus Delphi-type survey and during and after the expert workshop on "Exploring the future of Enterprise search" (organised in October 2011 in Seville), both presented in the following sections of this deliverable. They provide a picture of the current state of the enterprise search and discuss on possible prospects of enterprise search solutions.

_____

Table of Contents

1. Referred documents

[1] Project Contract GA 249008

[2] Project Contract – Annex 1 "Description of Work" (DoW)

[3] José Luis Gómez-Barroso, Ramón Compañó, Claudio Feijóo, Margherita Bacigalupo, Oscar Westlund, Sergio Ramos, Ajit Jaokar, Federico Álvarez, Rudy De Waele, Gema Mateos-Barrado and María Concepción García-Jiménez, Prospects of Mobile Search. March 2010, EC-JRC-IPTS report EUR Number: 24148 EN, available online at http://ipts.jrc.ec.europa.eu/publications/pub.cfm?id=3159

[4] S. G. Nikolov, R. Compañó, M. Bacigalupo, J. L. Gómez-Barroso, C. Feijóo, "What is Missing for the Full Deployment of Mobile Search Services? Results from a Survey with Experts" in the Proceedings of the 9th Conference of Telecommunication, Media and Internet Techno-Economics (CTTE) 2010, Ghent, Belgium, 7-9 June 2010

[5] Expert Workshop: "Exploring the Future of Enterprise Search", 13-14th October 2011 (Seville), available online at http://is.jrc.ec.europa.eu/pages/TFS/search/workshop2011.html].

[6] Notes of the Expert Workshop: "Exploring the Future of Enterprise Search", 13-14th October 2011 (Seville), available online at http://is.jrc.ec.europa.eu/pages/TFS/search/workshop2011.html].

[7] 2011 Delphi-type Survey organized by IPTS in collaboration with Intranet Focus, in the period of July-September 2011 (Appendix A)

[8] P-J. Benghozi, C. Chamaret, Economic Trends in Enterprise Search Solutions EC-JRC-IPTS report EUR Number: 24383 EN-2010, available online at http://ipts.jrc.ec.europa.eu/publications/pub.cfm?id=3320

2. Introduction

Enterprise search solutions render business processes more efficient. Being a technology that retrieves information within organizations, enterprise search is a key element in increasing the competitiveness of the digital economy and constitute, therefore, a strategic market for the European Union.

The 'Enterprise Search in the EU, a techno-economic analysis' report, enclosed in this deliverable, presents the results of a study conducted by IPTS, the Institute for Prospective Technological Studies (IPTS)-JRC of the European Commission, in collaboration with Intranet Focus Ltd, during the summer-autumn 2011 on the current state of the enterprise search. It represents the most updated and comprehensive overview on enterprise search market and on its potential.

The "Exploring the Future of Enterprise Search" workshop was organised, in the framework of the Chorus+ project, by the IPTS, the 13 and 14 of October 2011, in Seville, in order to validate the results of the this report and to contribute to better understand some of the following issues of enterprise search:

- Current and emerging business models of providers of enterprise search solutions and their respective value chains.
- Emerging techno-economic trends and future market structure in the domain of enterprise search.
- Strengths, weaknesses, opportunities and threats (SWOT) for the Europe Union with respect to enterprise search
- Need of policy changes or support actions in the field of enterprise search in the EU.

The content of this deliverable develops and updates the results of the 'Economic Trends in Enterprise Search Solutions Report'[3] and of the 'Prospect of Mobile Search Report', both published by IPTS during the 2010 [8].

3. "Enterprise Search in the EU, a techno-economic analysis" Report

As part of the Chorus+ project, this report presents the results of a six-months study on the current state of the enterprise search carried out by IPTS in collaboration with Intranet Focus Ltd within the CHORUS+ project and completed in October 2011. The <u>executive summary</u> of the report is provided below.

The 'Enterprise Search in the EU, a techno-economic analysis' report starts by providing an introduction to enterprise search. It considers the value chain for Enterprise search and the global business. It highlights differences and commonalities with other kinds of search solutions. It analyse the European market of ES and describes the main barriers and challenges for the take off of the Enterprise search, exploring how to make a business case for it. It studies users' selecting and implementing ES application and the general acceptance. Moreover, it presents emerging trends in technology and markets that could shape enterprise search. This foresight was partly based on an analysis of a Delphi study/ questionnaire [Appendix A] designed by IPTS and Intranet Focus and addressed to enterprise search experts.

Drivers, barriers and enablers for enterprise search are summarised in a SWOT analysis. The report concludes with some policy recommendations in view of the likely socio-economic implications of enterprise search in Europe.

Enterprise Search in the EU, A Techno-Economic Analysis
Final Report

Executive summary

The Institute for Prospective Technological Studies commissioned Intranet Focus Ltd. to undertake a techno-economic study of the enterprise search market in the EU. The project was undertaken by Martin White, Managing Director, Intranet Focus Ltd. over the period from 3 June to 16 September 2011. A programme of literature research and interviews with enterprise search vendors and integrators was undertaken, together with a Delphi study among search vendors, integrators and industry analysts.

The value of enterprise search

For the purpose of this project the term '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 file shares. This is often referred to as 'unstructured' information. The development of enterprise search technologies dates back to the late 1960s to meet the requirement 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; Boolean, vector space and probabilistic. 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 some differences between the requirements of searching web sites and searching other enterprise applications, primarily around security management, but it is possible to use the same enterprise search application for both purposes.

Enterprise repositories of unstructured information are growing very rapidly because of the widespread adoption of social media, increased compliance and regulatory requirements and a lack of resources to remove redundant information. 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. Surveys indicate that 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.

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 products dominated the market but there is now 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.

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.

Five 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 search market through promotional activities. These can be termed Type 1 vendors, and are Autonomy, Google, IBM, Microsoft and Oracle. Type 1 companies in general sell enterprise search as part of an overall enterprise application suite, and not as standalone products. The long term implications of the planned acquisition of Autonomy by HP 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 are mostly funded by venture capital and private equity placements. Examples of Type 2 companies with headquarters in the EU are Exalead (France), Fabasoft (Austria) and Sinqeua (France). Type 3 companies build products around open-source software such as Lucene/Solr, with Intrafind (Germany) being an example.

Because there are no comprehensive lists of search vendors operating in the EU market 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. 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 is on Wikipedia but even this has a number of omissions

Specifying and selecting search software

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. 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, and this is because no single business unit is likely to be able to make a business case for enterprise search. 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. IT departments have a role to play in the technical evaluation of enterprise search applications and in the initial installation, but

IT departments will have an important role to play in the procurement process 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 may 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. The installed base of enterprise search applications is so low in the EU that even in larger companies there is unlikely to be any previous experience of specifying, implementing and managing enterprise search applications.

The EU market for enterprise search

There are no reliable revenue analyses for the EU enterprise search market. In the case of the Type 1 companies there is no disclosure of the specific sales of enterprise search software, and most of the Type 2 companies are privately held and do not disclose detailed financial information. In addition open-source software (notably Lucene/Solr) is increasingly being used for enterprise applications. There are no license fees to be paid but there will be development costs. In the case of commercial search applications these development and implementation costs could be at least three times the software license fee, and this is an important element of any assessment of the investment being made by organisations in enterprise search. One of the challenges in open source development is being able to feel

confident in the skills of the development team. Recently Lucid Imagination, 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.

Based on information provided by International Data Corp., sales of commercial enterprise search applications in the EU in 2010 were around $500 million, but in addition organisations will typically be paying a further $1.5 billion in development and implementation costs. Probably 40% of the sales revenues can be attributed to search modules in large enterprise suites from IBM, Oracle and Microsoft. Autonomy's share of the EU market is around $200 million. Most Type 2 companies have sales of less than $20 million with the exception of Exalead. The current installed base estimate is probably no greater than 10,000 organisations. Any company with more than 1000 employees is likely to have enterprise repositories of a size to benefit from an enterprise search application so the potential market could be of the order of 200,000 businesses excluding those in the financial sector. On this basis there is clearly a significant potential market in the EU for enterprise search.

Achieving high search performance

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 a 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 then 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 users 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 less than 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.

Technology forecast 2011 – 2015

The outcomes of the Delphi study and interviews with search vendors indicated that there were six important areas of technical development for enterprise search over the next five years.

- Integrated search of structured and un-structured content

- Search as an integration platform (unified access platforms)

- Search incorporated into business intelligence applications

- Search-based applications

- Text mining

- Enterprise mobile search applications

In all cases these 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.

Information retrieval and enterprise search

There are at least thirty research institutes in the EU undertaking research into information retrieval. However very few offer taught 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.

SWOT analysis

A SWOT analysis of the availability and adoption of enterprise search in the EU highlighted the following topics.

Strengths

- World-class EU search vendors
- Significant information retrieval research capability
- Understanding of cross-language retrieval
- EC (FP7) support for information retrieval research
- Active information retrieval community

Opportunities

- Significant market potential in all Member States
- EU commitment to Open Data

Threats

- Difficulty in identifying search vendors and integrators
- No research being carried out into the benefits of enterprise search
- No enterprise search community

- Enterprise search as a low organisational priority
- Lack of reliable market data
- Public sector procurement procedures

Weaknesses

- Small scale of the EU enterprise search industry may restrict the exploitation of very large data sets by EU companies
- Enterprise search licences bought by US companies for global use reduces the EU market potential for large search contracts
- Shortage of candidates with appropriate training could impact the development of search vendors and integrators, and the adoption of enterprise search
- Market penetration of Google and Microsoft

EU support actions

In order to further support research initiatives in this area the European Commission is recommended to consider taking the following actions

Identification of research requirements

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

The European Commission should assess whether the tuition of 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.

Adoption of good practice in enterprise search within the European Commission

The Commission should ensure that the scope of activities and programs of DIGIT do include the effective specification, adoption and implementation of enterprise search applications for both internal repositories and also for Commission web sites.

The impact of Big Data

The European Commission should assess whether the EU has the technological and professional skills to exploit very large data repositories by bringing together stakeholders from the Commission, Member States with a commitment to Open Data and vendors with a current or potential ability to add value to these datasets.

4. Enterprise Search Delphi-type Survey

During the course of the project, a Delphi-type survey was undertaken with the questionnaire, designed by IPTS in collaboration with Intranet Focus Ltd (see Appendix A) being sent to 30 enterprise-search professionals during the period of July-October 2011. The questionnaire was designed to gather expert opinions on current trends, main challenges and likely future developments of enterprise search.

In particular, the survey inquired about enterprise search trends with regards to technology, business models, market structure and future prospects.

The questionnaire was structured around six main issues:
- How will the enterprise search business develop over the period to 2015?
- What do you see as the main barriers for vendors to selling an initial enterprise search application?
- Which factors will become increasingly important in the selection of an initial enterprise search application/solution?
- Which factors do you think will become increasingly important in the decision to replace an existing enterprise search application/solution?
- What is your assessment of the potential impact of the following technologies in increasing the demand for and adoption of enterprise search over the next five years
- Enterprise mobile search

Eighteen replies were collected. To note that in a number of cases there was a significant difference between the scores from vendors, analysts and integrators.

The results of the survey were analyzed by the IPTS team and fed the discussion at the "Exploring the Future of Enterprise Search" Expert Workshop and at the Think-thank that followed (Seville, October 2011) [6].

5. Exploring the Future of Enterprise Search 2011 Expert Workshop

The "Exploring the Future of Enterprise Search" expert workshop [6] was organized by the IPTS in Seville, Spain in October 2011, aiming at gaining insights into the techno-economic and socio-economic trends in enterprise search market (ES) in EU and how these may impact on the European Economy and Society.
The workshop was attended by 15 of the top enterprise search experts in Europe from both industry and academia. It resulted in a lively discussion about some of some of the main issues and challenges of enterprise search, such as:

> **- Market Dynamics**: including current and emerging business models of providers of enterprise search solutions and services and their respective value chains.
> **- Future Prospects**: 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

Participants debated in particular the results of the **Delphi-like Study [7]** and of the Report on **"Enterprise Search in the EU, A techno-Economic Analysis"** [section 3 of this deliverable], carried out by the IPTS with the collaboration of Intranet Focus Ltd.
The workshop started with the keynote speech of *Gregory Grefenstette,* of *Exalead*, one of the leading enterprise search companies in Europe on "the challenge of Search for Enterprise Search and Search-based applications", followed by three thematic sections:

a) **Technologies, solutions and research issues**
b) **Services, applications and business models**
c) **Techno-economic aspects of enterprise search**.

The complete notes are included in this section. The notes and all the presentations of the workshop are available online on the IPTS web site [5] [6].

Notes of Exploring the Future of Enterprise Search

13-14 October 2011, IPTS, Seville, Spain

Table of contents

1. Objectives of the workshop

The "Exploring the Future of Enterprise Search" workshop was organised by the Institute for Prospective Technological Studies (IPTS), which is part of the Joint Research Centre (JRC) of the European Commission, in the framework of the Chorus+ a coordinated action and a FP7 project. The main objective of the workshop was to gain insights into the techno-economic and socio-economic trends in enterprise search market in the EU, and how these may impact 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?

The workshop was held on 13 and 14 October 2011 in Seville, Spain.

As a complementary event, a roundtable discussion was scheduled to follow in the second part of the workshop to foster debate on the future trends and directions of enterprise search (covering techno-economic aspects). This debate was organized as a Think Tank on Enterprise Search, with the aim to come up with a (possibly consensual) assessment of the most important technological and business developments in the enterprise search area and propose policy options.

Questions the think tank meeting has addressed include:
- What are the main **technologies** shaping enterprise search?
- What is the **state** of the enterprise search market and how is it likely to **evolve**?
- Which are the major **bottlenecks** and the main challenges to be overcome?
- What are the main **business models** today and in the future?
- How is **Europe** placed with regards to the rest of the world?
- Is there a **consensus** on future trends and directions?
- What are the main technological and economic **challenges** ahead?
- What does this mean for Europe (SWOT analysis and policy options)?

The notes of the workshop which follow below are organised chronologically by session and within each session, by topic of discussion. Care has been taken to respect the flow of the argument, which required, in some cases, changing the order of some contributions.

The workshop was opened by Dr David Broster, Head of the Information Society Unit of the JRC-IPTS, Seville, who welcomed the invited speakers and other participants in the workshop and offered an overview on the Institute hosting the event. The IPTS mission is to provide customer-driven support to the EU policy-making process by developing science-based responses to policy challenges that have both a socio-economic as well as a scientific/technological dimension. He briefly described the research activities carried out at the IPTS and at the IS Unit in particular, focused on the socio-economic analysis of emerging Information technologies, in collaboration with several Directorate Generals of the European Commission (DG INFSO, DG JUST, etc.).
Following that, Dr Ramon Compano (IPTS Direction), provided some historical information regarding the previous research activities carried out by IPTS within the Chorus+ and its predecessor, "Chorus" project, about search engines, mobile search and finally on enterprise search. He reminded that IPTS is studying how new search engines are in general being developed as an answer to the citizens and organizations demands of retrieving digital information and suggested some reflections on the results and the main challenges encountered while conducting these researches. In particular he mentioned the risk that, despite the thorough and vast research studies (of socio, economic and legal nature) conducted in order to provide adequate policy options for the EU in this domain, these policies could "arrive too late", when the technological process is already underway (running the risk, for instance, that the little Search Engine companies are eaten by the big 'elephants'). He stressed the very aim of the workshop, i.e., to discuss with the invited experts the thorny issues in the enterprise

search domain and to arrive to a consensus on possible policy options. Dr Ramon also noticed the big absent at this kind of debates: the 'citizens'.

Dr Stavri Nikolov concluded the introductory speeches of the IPTS representatives, providing a brief description of the Chorus+ project (http://www.ist-chorus.org) and presenting the objectives of the workshop and the agenda. He reminded that Chorus+, as a Coordination Action, aims at coordinating national and international projects and initiatives in the area of multimedia search and at extending this coordination and cooperation to non-European countries. A particular emphasis is devoted on setting concrete R&D and industrial objectives on multimedia search in Europe.
Three thematic sessions were planned in the first part of the workshop to gather expert contributions on different aspects of the enterprise search, to be discussed during the TT roundtable.

Dr Nikolov, reminded the audience that the main objective of the workshop was to gain insights into techno-economic trends in the enterprise search, that is to say "where we are today and where we are going tomorrow". He referred to the following three thematic sessions as corresponding thematic areas of enterprise search:

a) **Technologies, solutions and research issues**
b) **Services, applications and business models**
c) **Techno-economic aspects of enterprise search**.

He anticipated that the participants would have been asked to give their opinions in particular about the results of the **"Enterprise Search in the EU, A techno-Economic Analysis"** and about **"the Delphi Study",** carried out by the IPTS with the help of Martin White (Intranet Focus Ltd).

Following the introductory speeches, Dr Nikolov gave the floor to the invited keynote speaker: Mr Gregory Grefenstette of Exalead, a global software provider in enterprise and web search markets (http://labs.exalead.com/).

2. The challenge of Search for Enterprise Search and Search-based applications *Gregory Grefenstette (Exalead): keynote speaker*

The keynote talk focused on the **meaning of Enterprise Search and the difference between this** The keynote talk focused on the Search Based Applications in the Enterprise Search field. Mr Grefenstette firstly presented a brief overview on the concept of enterprise search (ES) as the practice of identifying and enabling specific content across the enterprise to be indexed, searched and displayed to authorized users.

The term *enterprise search* is commonly used to describe the application of search technology to information within an organization, and therefore should be differentiated from other type of search, such as web search.

He passed then to highlight the differences[1] between ES and web search, stressing that:

1) ES is not just "a small Internet";
2) in ES "web pages in" and "web pages out" is not valid, since much of enterprise information to be indexed comes from databases, emails, tables
3) enterprise information is both structured and unstructured information and ES aims to merge them ('Hybrid Structured data');
4) people in ES do not just search information but aim at "getting a job done";
5) people use facets to navigate to known information in ES data after an initial keyword;
6) differently from a web search engine, such as Google, in enterprise data there are few links between "pages" so the page rank algorithm does not work;
7) Users are identified in ES and access rights have to be managed;
8) enterprise users expect to be able to manipulate their search results using sliders, bar graphs, pie charts.

Following to that, Mr Grefenstette addressed the following aspects:

- **Motivators** for enterprise search:
  - o There is increasing information everywhere. 210 billion of emails per day; the 80% of the enterprise information is unstructured; 988 Exabytes is the worldwide data volume in 2010.
  - o The digital data growth is enormous, expected to be of 35 zettabytes in 10 years.
  - o 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). Large organisations often have 30 or more separate information processing systems (ERP, CRM, CMS, etc) and ES systems must be able to reveal the data in all of them.

- **Types of search**
  There are three types of search:
  - o Unique search box (that people have learned to use in the past ten years)
  - o Faceted search (e.g., for prices, for colours, etc.): it is semantic, it allows visualization, people familiar with it from shopping sites;
  - o Form based search: traditional search on database, but it is complicated (for the several fields to fill in and users need to know field values).

- **Facets**
  - o Facets are semantic dimensions
  - o They are visualisations of semantics that users can understand, coming from databases or from text and linking together structured and unstructured data..
  - o Search based applications are possible because of semantics

- **Semantics and databases**
  - o Search engines now handle rich semantics
  - o Database semantics is imported easily into search engine facets
  - o Semantics is extracted from unstructured text: natural language processing in an ES system reveal this semantics **Facets** can make search results look like Business Intelligence, provide interactive reporting and real time operational tools (enabling real time reporting on questions that could not be anticipated), build decisional tools on unstructured information.

---

[1] As defined by M. Bennet, New Idea Engineering, Inc- Vol 5, 4, 2008.

■ Search based applications

- o Search based applications **(SBA)** merge information from databases and other sources using search engine technology
- o Advantages are: usability, performance, agility for a large number of users and all data (structured and unstructured)
- o they work connecting the data sources, processing the data, creating virtual feeds, allowing users to design their pages
- o A SBA addresses critical business issues by enabling easy Search & Discovery into key data by key users;

Mr Grefenstette also stressed that a SBA should **not** be considered as **a replacement** for traditional applications, not a reproduction of all complex, historical business logic, but as a easy way of revealing data stored in these traditional applications

He proposed some **examples** of SBA, from the private or the public sector, such as 'Restminer', capable to unify structured and unstructured data sources or 'Urbanizer' for smart phone, offering semantic analytics for mood-based local search, or 'Sourcier', useful for Government duties such as map-based access and reporting for water quality data.

Mr Grefenstette condensed the current and future needs in the "domain specific semantics",recognising, as remaining challenges, intelligent query interpretation (i.e., intelligent analysis of query to extract relevant term classes, for instance, in the cases of Yellow pages or National train system web site).

He concluded his talk by pointing out that, given that search engines can handle the semantics of databases (though not the transactions), that facets are semantic dimensions and that semantics allows 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[2].

Enterprise data all over the place

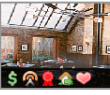[2] G. Grefenstette, L. Wilber, *Search-Based Applications*, Morgan & Claypool, 2011
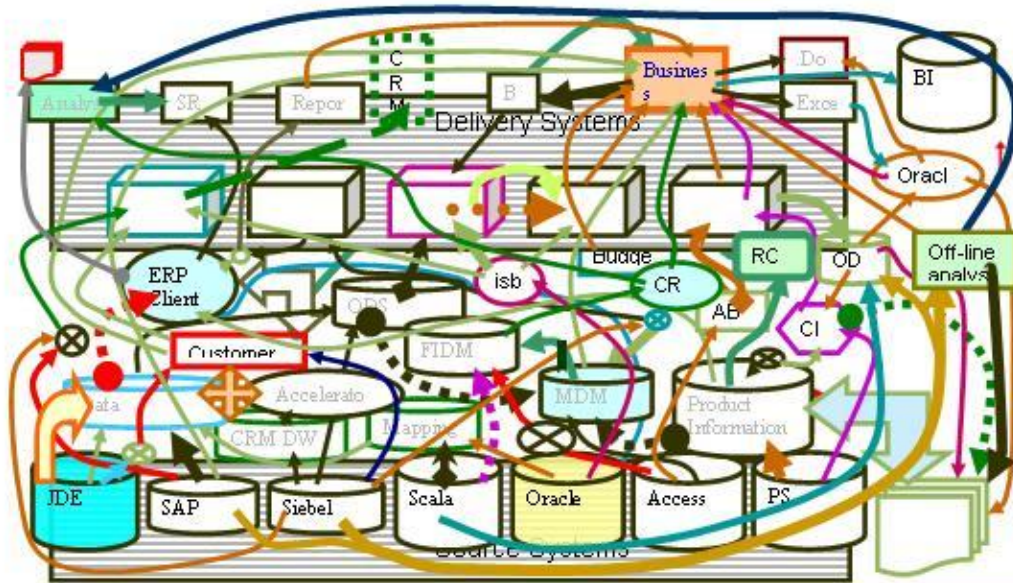
Source: Leveraging Search to Improve Contact Center Performance
Richard Snow, VP & Research Director, Customer & Contact Center Management
March 2009

Unify structured and unstructured data sources

In the Q&A session following his talk and before the three sessions of short presentations Mr Grefenstette has the occasion to address some issues raised by the participants' questions. He clarified, inter alia, the concept of business intelligence, that is, the use of data about the enterprise community; regarding the relevance of audio-video search in ES and Search Based Applications (SBA), he pointed out that every company now has audio/video data, the quality of which is growing, as well as that people are going to record almost all moments of their life. Therefore, these data should be captured in enterprise Search. Moreover, he clarified that SBA will be able to feed into transactional database system. Finally he partially confirmed the concerns regarding the possibility to process personal data through ES: considering, for instance, the use of Blackberry for company's tasks, all the valuable data (possibly also personal), that are available by the employee's desktop, will be indexed outside.

3. Semantic Enterprise Search and Content Intelligence
Matthieu Jonglez (Smartlogic)

Matthieu Jonglez's talk started with the provocative question: "what went wrong with enterprise search?" aimed at pointing out the main challenges that enterprise search has to face.

According to Matthieu Jonglez, **the challenges that ES has to face are higher than in web search, but investments in promoting content production or in education are not enough**.
He presented data showing that researchers are successful in finding what they seek only at the 50% of the case[3] and this did not change in the last ten years. Smartlogic research has shown that half of people cannot find information they are seeking using their own organization's enterprise search facility within an acceptable amount of time[4].

He pointed out that while Search Engine Optimization (SEO) helps on the Internet, this is not the case for Enterprise.
He considered then the **different degrees** of order with regard to the management of enterprise information:

- I degree: **Filing management**. Jonglez noticed that the 80% of enterprise information is unstructured, doubling every 19 months, and that there is an increasing burden of compliance for enterprises
- II degree: **Index management**. This degree of order relies on file plans and metadata schema, characterized by mono-hierarchical standardized taxonomies and manually applied classification that can ensure only a low level of consistency and quality
- III degree: this degree is the result of the **computerization** of the I and the II degrees, based on the adoption of a proper human interface to access to content services, through search engines. Jonglez commented that following to this computerization we assisted to an explosion of information, giving the example of the U.S. administration, where the amount of data passed from 4 to 80 Terabytes between the 1993 and the 2009. The III degree has been defined by Jonglez as *Content universe*, in which the relevance of the content and its management increased significantly and where people also create content.
- IV degree: **content intelligence**. As content developed enormously during the last years, also through social computing (user-created content), there is an increasing need of 'content intelligence', able to make sense from the content, to give it meaning, to organize people data (like they usually do with their items) and to enrich the user's experience.
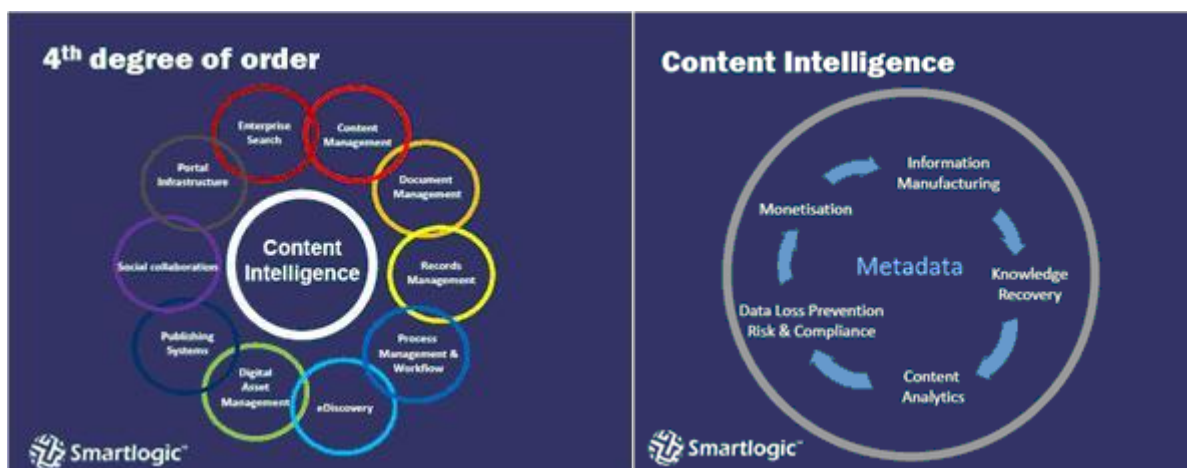
The slides above shows the relevance of metadata and of a content intelligence platform (comparable to the relevance of having labels for supermarket products) that, according to Jonglez, do help users to see what he/she

---

[3] Source: Quantifying Enterprise Search, IDC, May 2002
[4] Smartologic/MindMetre survey, April 2011

needs and what he/she likes, contextualizing and classifying content and therefore **increasing his/her experience**.

Jonglez concluded his talk stressing the need to integrate ES with semantic technology (content technological solutions).
He clarified that Content Intelligence is software technology that adds a layer of intelligence and context to the underlying information management systems, such as search, content management and workflow that an organization has already deployed.

In particular, content intelligence delivers:
- Exceptional user **experience**;
- **Automation** of business processes involving unstructured content;
- Automatic **classification** of content thus improving the quality of the information assets.

4. Open Source Techniques Push enterprise search & Search Driven Applications
Christoph Goller (Intrafind)

Christoph Goller's talk focused on **Open Source Techniques** as solutions to promote enterprise search and Search Driven Applications.

He started by offering an historical overview on the development of the open source, since the emergence of Lucene and Solr in late 90's as library for indexing and ranking, until their wide acceptance from 2005 by technology organizations and their following expansion as web service, converters and connectors. He gave some examples of media enterprises that use them as application, such as Apple, IBM, Twitter.
Goller continued his talk highlighting the **strengths and weaknesses** of Lucene and Solr:
🔴 Strengths:
- o   Best practice segmented index (like Google, Fast)
- o   flexible ranking (term/field/doc boosts, function queries, custom scoring…)
- o   Best overall query performance and complete query capabilities (Boolean Operations, Wildcards, Similarity Search;
- o   Multilingual Stemmers, Filters, Memory Mapped Indexes Near Real-Time Search, Cache Management, Replication, Faceting;
- o   Rapid Innovation, Extensible Architecture, allowing complete control (open source)
- o   **Core technology as good or better than any other and open source**
🔴 Weaknesses:

- o   No formal support, limited access to training & consulting
- o   Certain features for Enterprise Search missing:
  - ▪ Secure Search
  - ▪ Graphical user- and administration interfaces
  - ▪ Integrated search of structured and un-structured content
  - ▪ Connectors and stable Converters
  - ▪ Viewer Component (preview or view docs in their original layout)
- o   Not satisfactory user interface

Goller stressed then the current competitive situation, arguing that, while full-text search has become a commodity (especially thanks to the market strength and features of established enterprises), it became hard to justify the high prices (especially for large applications) as well as closed and proprietary technology.

Goller expressed the view that good Enterprise Search solutions have to offer more than just full-text search: therefore, companies, like Intrafind, can concentrate on solving the real problems in Enterprise Search.
He presented then the advantages of Intrafind Software AG that may mean: **semantic search, semantic linking** (combining structured and unstructured data), search-driven applications (linking to business Intelligence) and much more, as illustrated in the slide below.

After having illustrated the business model of Intrafind (that is an open source business which can built on more ten years of experience, of more then 700 customers and a wide partner network), Goller reminded the audience that Intrafind is a product company (of: iFinder, Topic Finder, Knowledge map, Tagging service), combining open source components with in-house development) and that one of the strength of Intrafind is to concentrate on relevancy and automated generation of metadata.

He presented then some examples of Intrafind products:

- the *Morphological Analyzer*, for applications in language search and based on a Lemmatizer (that maps words to their base form) and a Decomposer (decomposes words into their compounds); he contrasted it to stemming, usually based on a simple algorithm, therefore not having the morphological analysis advantages (e.g, combing high recall with high precision for search applications*).*

- *the Named Identity Recognition*, which permits the extraction of information from unstructured data and can be employed for several applications (e.g. search for 'experts', for companies or persons by names)

- the *Text classification*, that enables the automatic assignment of documents to specific topics based on their content and can be applied, for instance, to email classification, newsletter management system, etc.;

- the *knowledge map*, i.e. a graphic way of visualizing enterprise knowledge allowing to browse through the enterprise content.

5. An approach to overcoming obstacles to Enterprise Search Research
David Hawking (Funnelback)

David Hawking's talk started by questioning how the obstacles to the Enterprise Search, that has a strong potential, could be overcome. One of them is identified in the **lack of data**, as obstacle to the academic research in this field.

He mentioned TREC ("Text Retrieval Conference", trec.nist.gov), an initiative organized in the public sector in order to create data set for test collection and relevance judging resources, and argued that something similar should be thought in the ES domain, where, conversely, the research is difficult to carry out (difficulty in creating benchmarks for objective comparison of data) as the situation is characterized by heterogeneous repository, access rights, diversity of document types and diversity of use cases (e.g., legal search, professional search like patents, team building or everyday search). Hawking, thus, stressed the **very little academic research in ES,** despite the huge economic value of improving it, in particular in terms of competitiveness and efficiency, due to the lack of TREC-like resources.

Hawking continued his talk by discussing the importance of enhancing the collaboration between academia and industry, especially in this domain. One of the rare examples is that of the ARC Linkage Grant Proposal for research into Enterprise Search forwarded by the RMIT University in Melbourne, Australia, with Funnelback as industry partner. The objective of this research[5] is to provide open publication of the outcomes and to build a public test collection.

The basic idea behind this, Hawking explained, is that technologies allows us to do comparison of data set as we want, so, why do not use them? With the aim of charactering enterprise search in a wide range of organizations, Hawking is hopeful to find a way to build test collections which could accurately model a wide range of ES (outcomes will be announced soon).

Concluding, he suggested that the mentioned research makes to believe that it is possible to arrive to a collaboration also between US and EU on this domain.

6. Visual Search for Enterprise Documents
Mika Könnölä (Documill)

Mika Könnölä started his talk by presenting the activities carried out by Documill in the field of ES. He stressed that Documill operates in the ES domain as an indipendent software vendor (ISV), enabling browser-based access to MSOffice and PDF documents and powering large scale content processing solutions; Documill core competency is server-side enterprise document processing: " we take actions from the results of search".

Könnölä continued his talk, by illustrating the Documill offering. Documill has its own Engine (for document parsing, rendering and dynamic publishing solution) and, on top of this, is built the **Documill Visual Search Suite**, a complementary solution running on top of existing content indexing and search capabilities.

Searching documents is frustrating for enterprises, because enterprise documents are different from Internet pages (HTML):

- they are usually long;
- they include important visual objects, which are ignored by text based indexing;
- when the right document is found, getting the right page requires additional search.

Könnölä explained that the scope of Documill is to make enterprises' life better, offering a solution similar to Internet search: **visual search for documents**. A document is not downloaded before the person who is looking for it is sure that it is what he/she wants; the visual search solution brings this person directly to that page containing the document. Accuracy in the results is therefore one of the benefit. As illustrated in the image below, it also enables previewing (e.g. instantly *zoomable* page) and simultaneous view of multiple documents instead of just one.

---

[5] The research grounds on the data collected through an online questionnaire, available at tinyurl.com/6kndjzk.
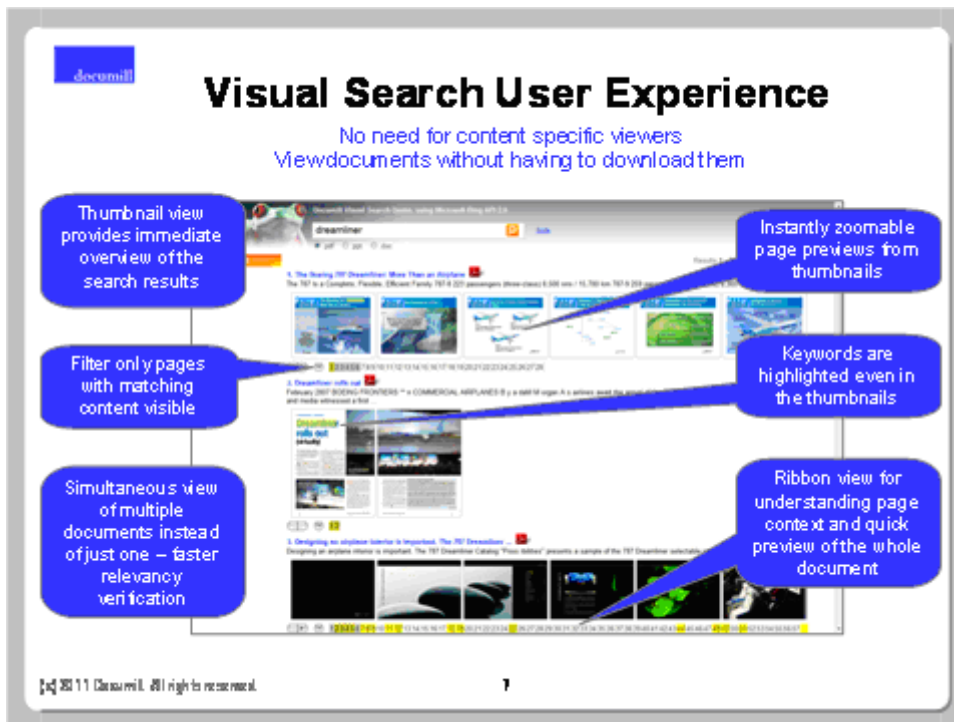
The visual search tools offered by Documill realize a user experience revolution in the sense that:

- Focus on most valuable corporate assets; data hiding in enterprise documents
- Significantly speed up document discovery experience
- Enable integrated high-fidelity content viewing

- When core search functionality is reliable, you can make it <u>easier for users</u> to:
  - Verify search result relevancy
  - Improve insight into documents; page-level, document length
  - Find documents based on recognizable visual characteristics
  - Offer fastest access into content
  - Enable visually enriched, enjoyable user experience

Moreover, given that, according to Könnölä, an enterprise is not only happy with finding documents but also with using these findings, Documill solution allows for the **re-using and sharing** of enterprise documents, ensuring not only fast way to read documents but also the possibility to use bookmarks, in order to manage the findings (including editability and collaborative actions) in real time. Könnölä pointed out the importance of **the Drag&Drop capability** which enables to:
- Choose relevant pages directly from the display of search results and document listings;
- Drag&drop pages to build collections;
- Share collections with other users ;
- Export collections as new documents

He considered then the summary of benefits deriving from using **Documill tools for Enterprise Search**, that are, among others, to speed up the document discovery process (up to 7x faster discovery); to save up to 3h/week (Improved productivity); to make less downloading (significant bandwidth savings)

Konnola concluded his talk by envisaging possible future scenarios for Documill applications:

- Enable collaborative search
- Enable convenient re-use of existing assets
- Provide complete, integrated solutions
- Offer search as a service (SaaS) easy to try and buy

Session 2: Services, applications and business models

7. Cloud-based Enterprise Search: Future or Reality
Kathrine Hammervold (Microsoft)

Katherine Hammervold started her speech by taking her personal information and career data as an example of tag cloud and illustrating the concept of *Cloud* in which we live and work today.

Her speech continued by revising the set of trends and emerging gaps she identified between the enterprise and the rest of technology world, that, so far, have not sufficiently considered.

She observed that the industry trends are influenced by social ('consumerization') and technological trends.

She illustrated these trends by providing the following data:

- There are around 330 million of Smartphones;
- Mobile workers are going to reach 1 billion in 2011;
- Revenue from IT Cloud services are expected to equal traditional in 2013.

As a general comment, she stressed that the social, consumer and mobile trends are driving the *cloud*.

People are already making use of cloud services in the daily life, they will expect the same from enterprise search; there is a quite big gap between the technological trends and what is going on within the enterprise. However, Hammervold pointed out that most Enterprise Search vendors are already offering some cloud services. She mentioned, as examples, the case of Searchbox, a solution which works in the cloud that is based on Lucene; or Google Apps (email, docs and communications) that also provides search where the data are; or, more, IBM, Oracle or FAST.

She expanded the discussion on the benefits from the cloud, by providing the Gartner's **definition of Cloud computing**, as "a style of computing where scalable and elastic IT-related capabilities are provided *as a service* to customers using Internet technologies". She pointed out that the cloud computing is already available, it is gaining momentum and it is going to be used more and more.

Hammervold illustrated then five attributes of Cloud the enterprise could benefit from:

- **Service-based**: consumer concerns are abstracted from provider concerns through service interfaces;
- **Scalable and elastic**: services are scalable on demand to add or remove resources as needed;
- **Shared**: services share a pool of resources to built economies of scale;
- **Metered by use**: services are tracked with usage metrics to enable multiple payment models
- **Internet technologies**: services are delivered through use of Internet identifiers, formats and protocols.

Hammervold presented the Cloud as inevitable, as shown by the scheme below:

The Cloud is Inevitable

She made the point that, on the one hand, we are already assisting to a transition towards the cloud, but, on the other, it does not mean that we will change our idea of using technology: for instance, it does not change the way *Microsoft* develop software. *Fast* has gone from delivering highly customizable search platform to delivering search in the cloud as part of Microsoft and SharePoint.

According to Hammervold, what the enterprise should always bear in mind is what the customers want when they go on the cloud.

This is illustrated by the **"cloud first" approach**, responding to the needs of customers:

- availability and support (it is easy to monitor the data)
- security and privacy (e.g., great concerns are related to the shared data that could be linked to employees' information; general concerns derive from hosting different enterprise data, for instance of *Coca-Cola* and *Pepsi*, on the same server)
- competitive price (scalable)
- ability to innovate, that also means partner integration, flexibility and App model (e.g. possibility to access to my bills when I'm travelling).

Hammervold passed then to consider the **opportunities** offered by the cloud for the enterprise search. Some of them are: having search as platform service; data-driven decision making; recommendations/suggestion; specialized search; mashed up applications; federated search, bridging cloud and on-premise data sources; adaptive search experiences; moreover, using complementary technologies.

She finally argued that as search will more and more follows the content and users location, we will assist to a sort of democratization of enterprise search, the main benefit of which should be to offer an **adaptive experience**.

In other words, what enterprise search needs is an integrated platform, services and tools to enable businesses to provide a continually evolving, effective user experience, to achieve business results and to grow their business online.

8. Google as a Role Model for an Enterprise Search Team
Harald Kirsch (Raytion)

he affirmed that, without having any Google insider knowledge but as a common Google user having ES implementation experience.

According to Kirsch, the four following assumptions regarding Enterprise Search are generally accepted, but **often do not hold**:

- availability of technical requirements (e.g.: copy of a vendor's feature list)
- availability of business requirements: at least part of a company will require search, but search does not work…
- source system owners will cooperate (Intranet, phone book, portal, company newspaper, database): on the contrary, this implies more work, additional system load and security issues;
- availability of management support (what is the business case?)

Kirsch continued his talk, by comparing Internet with Enterprise, in particular considering Google in its early stage: as illustrated in the picture below, it is possible to observe that all the assumptions are wrong:

He claimed that Google fulfilled its mission, as Internet without search engine is hard to imagine; what about Enterprise search's mission? In sum, neither enterprise is possible without search.

Kirsch referred then to a Steve Jacob's famous assertion, according to which the necessary attitude (to assume also in the ES domain) is to think that "it is not the consumers' job to know what they want".

He concluded his talk by summarizing what a search team should be able to do by itself:
- envision;
- start small, but useful;
- release early release often;
- define and measure success
- Incremental improvements;
- Talk about it

Finally, considering that the business case for Google search is in the side-effects (i.e. advertising), he questioned if it could be the same also for Enterprise search.

9. Visual Asset Search: Application Areas and Business Models
Ray Owens (LTU Technologies)

Ray Owens focused his presentation on discussing the meaning of **Visual Asset Search**, as intended at the LTU, the application areas in which Visual Asset Search is pertinent to the larger context of Enterprise Search, as well as the business models.

He first clarified that « Visual Assets » means photographs and videos stored in enterprise repositories and that LTU has partnerships with enterprise search companies which have called upon LTU to add VISUAL search to the more general enterprise search capabilities proposed.

Regarding the application areas, he pointed out that while there are many applications for visual search in general, within the context of enterprise search the **applications are,** at present, **limited, but growing**. He explained that LTU does not propose visual enterprise search on its own as it is always in partnership with an enterprise search solution provider, generally a joint proposal.

He then provided an overview of LTU technologies, which are technologies based on academic research (from INRIA, MIT, Oxford), the main activity of which is image recognition and image search. The historical market of LTU is government and law enforcement (e.g. French police and Interpol), but newer markets are in enterprises such as:

> - media intelligence (e.g., *Ebiquity*, for printed and web advertisement tracking);
> - e-commerce (*Fotolia*, photobank manangement – helping track millions of images contributed by a community of photographers)
> - mobile applications (*La redoute*, in essence, using the camera a data entry tool to access further information).

Owens reminded the audience that these examples do not represent enterprise search, but they enable certain specific operations and are not meant to be accessible by the general staff of the enterprise.

The LTU core technology includes two basic functions, 1) to **index** visual content and 2) to **search** for visual content.

Indexing is meant as creating a description of what is in the image. This descriptor of the visual content is sometimes called a signature, others a fingerprint of the image: a numeric descriptor used like a label for finding (and use) the image again.

Searching visual content allows searching if an image in the enterprise repository, eventually in a different size or file format and if there are similar images: it can be called "search by example". Just on the basis of an image as the search query, visual search allows the retrieval of all of the images within the repository: because they have common visual elements.

Owens passed, afterwards, to consider what happens if **applying visual search to enterprise search**. He stressed that at least a difficult fact of life about visual enterprise search should be faced: while many types of companies need enterprise search, as of today, only a very few types of companies need visual enterprise search.

Two examples were offered with regard to enterprise repositories:
- **Publishing** enterprises: images purchased or created for publications
- **Trademark** registration agencies: images of logos and trademarks registered

Owens argued that, in the case of VISUAL enterprise search combined with enterprise search even if all business units may benefit from enterprise search, only a limited number of business units within the enterprise may benefit from visual enterprise search. These would be:
- business units concerned with the choice of images or videos for publication purposes
- business units concerned with image processing (example: processing logo registration requests). The visual enterprise search would be, however, a component of ES.

Owens provided an example of application in a publishing enterprise (illustrated in the image below): a media group owns multiple magazines, each with its own collection of visual assets. Visual search would allow making search within the global visual assets of the group, with benefits for creators and editors. For instance, before purchasing a new visual asset, the editor may use enterprise visual search to know if:

> - the enterprise already owns the image, and where are located all its variants
> - the enterprise owns similar images, which may be complementary to the new image.

Another example of application is that in the field of Trademark Registration Agencies, where enterprise search is very useful for the textual data concerning patents and trademark.

Enterprise VISUAL search is useful in this domain in the operational activities concerning images: logos, trademarks, and in some cases, industrial designs.

Owens continued his presentation by pointing out the **varieties of Multimedia Indexing**: in multimedia indexing, there are many aspects which may be indexed. Beside the aforementioned aspects of image matching and image similarity, there are other visual aspects which may be indexed and applied in enterprise visual search, as illustrated in the scheme below:
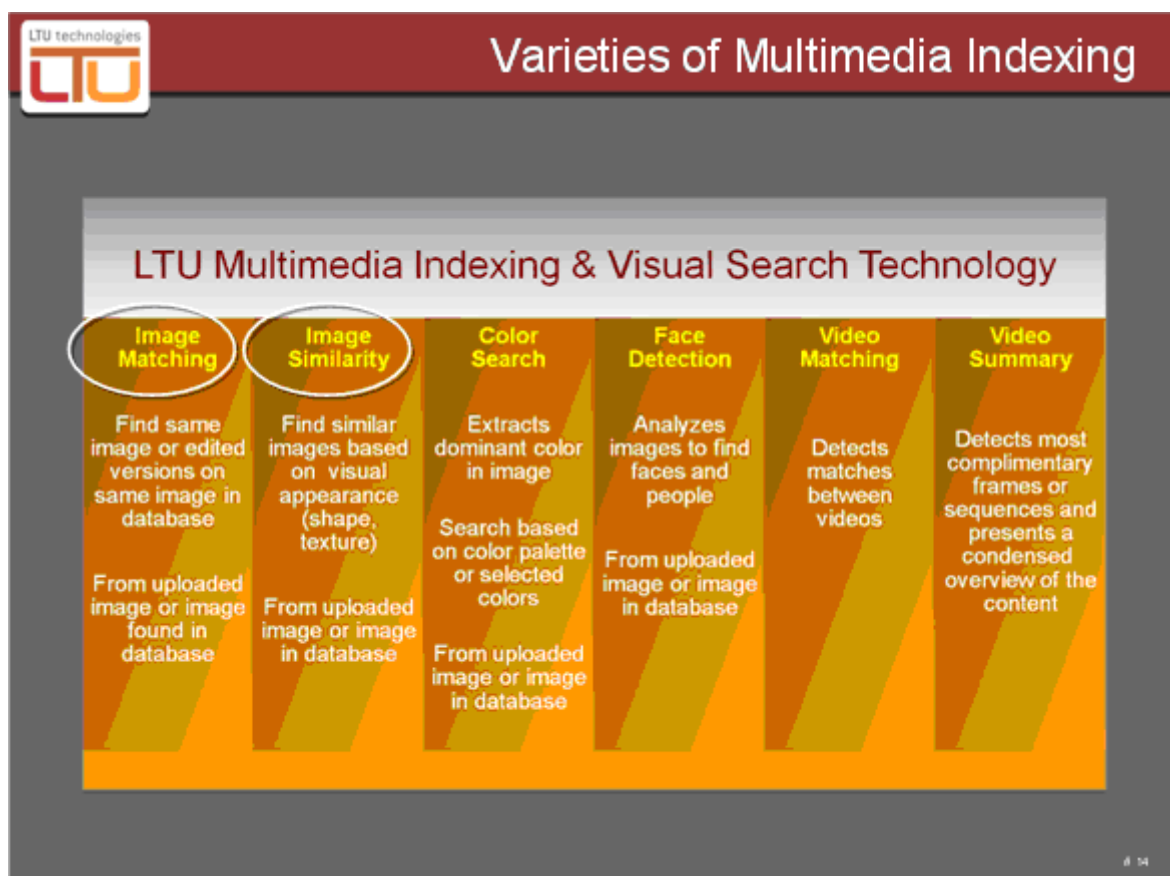
Owens concluded his talk by illustrating the **business model and architecture model** for including visual search within an ES solution. He argued that the business model is fairly simple, as it is a subsystem of the global solution: it would be the Enterprise Search solution provider who would integrate the Visual Search subsystem into the global solution. Therefore, LTU visual search would propose a software licence to the Enterprise search provider, while the latter would propose an integrated system to customer (including visual search subsystem).

Given that the visual search subsystem would be integrated with ES, its pricing may be aligned, according to Owens, to the same principles as the ES solution. For instance, if the price of the Enterprise Search solution is based upon the number of documents to be indexed, the pricing of the Visual Search subsystem is based upon the number of images and videos to be indexed.

10. Technologies for Mobile Application Development
Andrew Terry (Toumetis)

Andrew Terry's talk focused on Mobile search applications.

His first point was that global smartphone and tablet shipments combined exceeded PC shipments for the first time in February 2011. One year later they could be nearly double that of PCs[6].

Numbers

2011 Projected Global Device Shipments	Units (millions)	Annual Growth rate
PCs	384	3.1%
Tablets	63	260%
Smartphones	487	57%

Source: Gartner September 2011

He passed then to consider the situation in the enterprise search, where these aspects can be observed:

- the corporate blackberry is a smartphone now too;
- enterprise are increasingly supporting or specifying iPhones, iPads and other devices;
- "bring your own device" policies are increasing;

Using the motto "**don't shrink, rethink**", Terry stressed that what is relevant in mobile ES is not to make search becoming smaller, but different. This is possible because of some relevant features on which mobile applications can count on:

- Sensors:
  - Camera (still and video)
  - Location
  - Voice and sound input
  - Motion
  - Touch
  - and **text**
- Relationship:
  - Personal
  - With you all the time
  - Usually online
  - Simple to use
  - Not just "knowledge workers"

One has already a lot of possibilities with mobile solutions, but **Applications,** that offer rich and pleasing interfaces as well as access to all the sensors, can make a difference. A lot of innovation has been already observed in some early examples of mobile applications: e.g.: audio search (Shazam); bar codes (Amazon); location (Google); image (Google Goggles). However, Terry noticed, they imply more development work than a browser and multiple platforms.

He considered, then, the new objectives the mobile search should aim at, going from introducing new Voice interfaces (e.g. Siri), to allowing more innovative use of sensors, or to increasing accuracy and interactivity. The last aspect would mean, for instance, to find solutions to the problem of small screens (that make it difficult to scan results); to help user refine search and to provide him with smarter tools to extract key information from documents.

[6] Source: Gartner September 2011

In conclusion, Terry underlined that Mobile is becoming the mainstream and that there are lots of **opportunities** to take from it, such as:
- new ways to search;
- a wider set of users;
- potential for disruption.

SESSION 3: TECHNO-ECONOMIC ASPECTS OF ENTERPRISE SEARCH

11. Enterprise Search - A Gap analysis?
Martin White (Intranet Focus)

The focus of Martin White's speech was on the **market potential gap** of enterprise search. He provided some "big" economic data he collected, that, according to him, are the best available data on IT market, showing, that, though many enterprises in Europe with more than 1000 of employees have huge repository of data, very **few of them use enterprise search solutions**.
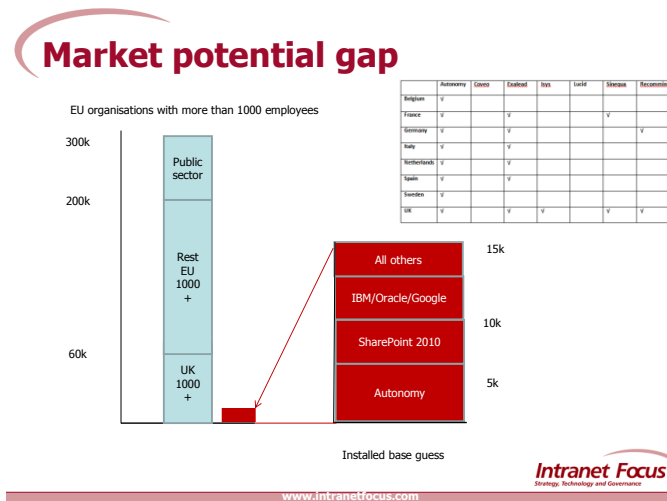
In particularly, according to the respondents, the amount of enterprise unstructured data will increase in the next three years. This implies that there is a market potential gap, as people are aware of the importance of unstructured data (more than 80%), but they do not resort to ES. In other words, **'big data' constitute 'big challenges'**.

White continued his talk by analysing and trying to 'understand' this gap. He stressed that more than 1/3 of respondents do not believe that the IT infrastructure and processes are adequate for managing unstructured data in the next three years. Moreover, the most employed technologies to search, analyze and deliver unstructured data are those related to content management system.

In conclusion, he pointed out that it is mainly a question of **skills gap**. In Europe, contrary to what happens in the U.S., there is not enough education and teaching programs on Enterprise Search, with the consequence that companies cannot find skilled and experienced staff to support enterprise search implementation (as indicated in the image below).

Skills gap

There are around 30 Specialist IR institutes in the EU

10 of these are in the UK

Only 4 have iSchool status

None teach a 3yr undergraduate course in IR

Building on a general undergraduate education, the IST program immerses students in information science; information architecture; human factors in information seeking and use; information organization, storage and retrieval; and applications of the Internet. Learn how information is retrieved, organized, manipulated, repackaged and applied in nearly every aspect of our increasingly digital culture and economy.

iSchool, University of Milwaukee

In 2011 Google will be hiring 1000 professionals in the EU

Companies cannot find skilled and experienced staff to support enterprise search implementation

Microsoft partners cannot find skilled and experienced staff to support enterprise search implementation

Search vendors, integrators and open-source developers cannot find skilled and experienced staff

Intranet Focus
Strategy, Technology and Governance

12. Search and Market Contraction - Lessons learned from the Catholic Church, Railroad Tycoons and Oil Magnates
Hendrik Speck (University of Applied Science Kaiserslautern)

Hendrik Speck's talk started by considering the technology and market maturity of Enterprise Search.

He mentioned four types of shareholders, as vendor classes:
- *type 1: large corporations* (e.g. Google, Microsoft, Oracle, IBM, Autonomy-facebook)
- *type 2: Proprietary solution providers* (Exaleed, Fabsoft, Sinquea)
- type 3: *Open Source Solutions* (Apache, Nutch, Solr)
- type 4: *Missing/Invaders* (Social Media, like Facebook; Vertically syndicated companies, like Apple; cloud, like Amazon; Enterprise Resource Planning, as SAP).

He argued that ES should only work if not alone, but as part of a large portfolio, as integrated part of wider search: in fact, there is no visibility of ES vendors in Europe without web search. In other words, Integration is preferable than consolidation and contraction of markets.
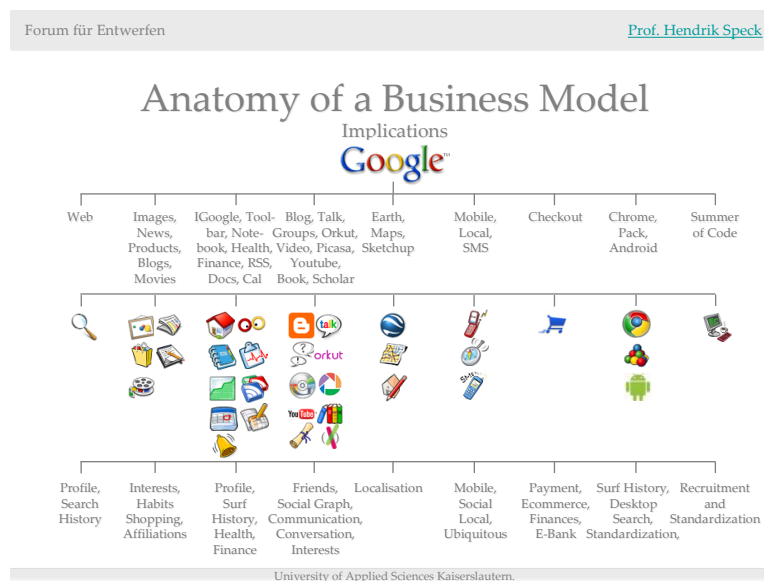
Speck stressed the importance of the lessons learnt from the social media, the future of which can be sketched as follows:
- Integration
  - Social media Newsroom/dashboard
  - Data engine/ Integrated communications client
- **Social customer relationship management**
  - Social media/ customer journey/ CRM/ support
- Organization meets social media
  - Research and development
  - Knowledge management
  - Customers, prospects, critics

Enterprise search has a lot to win from **integration with web search**, in particularly in terms of market visibility and brand recognition, but also in terms of Universal Interface, customer/corporation leverage, and business case.

Speck claimed that many media companies are positioning themselves in vertical perspective, in a "**vertical syndication**", the same that can be observed in the Catholicism. He then compared the Openness of Religion (e.g. Protestant vs Catholic) with the **openness of Business** (e.g., *iPad vs HP Slate vs. JooJoo vs??*). A powerful illustration of this openness would be that of the 'Holy War' between Mac vs. Dos, as indicated by Umberto Eco[7]

---

[7] Source: Umberto Eco. "The Holy War: Mac vs. DOS." *La bustina di Minerva. Espresso.* September 30, 1994.

and quoted by Speck. Accordingly, Macintosh would be *catholic* or better *counter-reformist* (it is cheerful, friendly, conciliatory and it tells the faithful how they must proceed) and MS-Dos would be *protestant*, or even *Calvinistic* (i.e., it allows free interpretation of 'scriptures'/program, demands difficult personal decisions to the lonely 'faithful'/user). The passage to Windows would represent an Anglican-style schism, but there would be always the possibility to return to DOS and take bizarre decisions.

Speck referred then to similar 'schisms' occurred in the occasion of the break up of the Standard Oil and of AT&T (into ATT and Babies bells)

He argued, however, that there are some decisions made in the technology domain that should remain the same, like it happened for the German Railroad Tycoons.

The following recommendations for the enterprise search concluded Speck's talk:
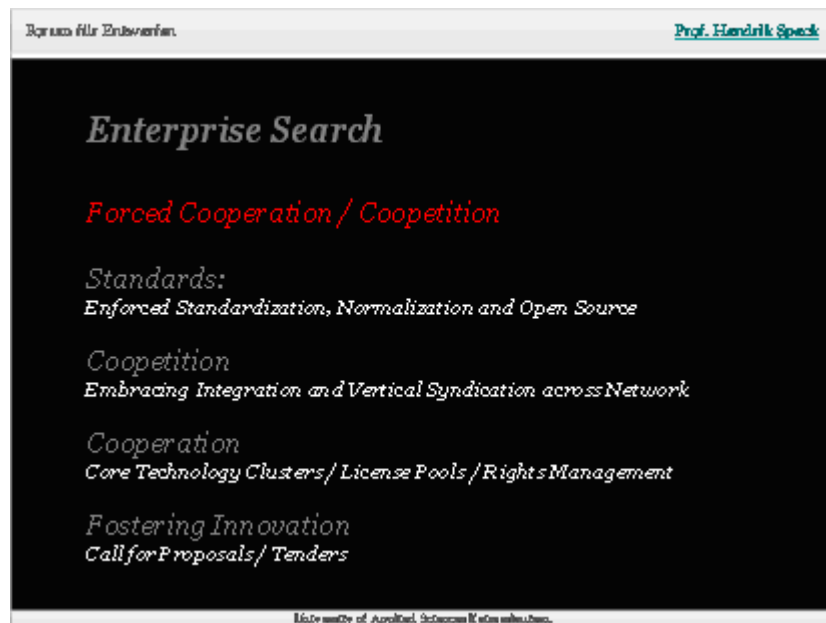
13. The future of Enterprise Search in Europe based on Altered Demographics *Henrik Strindberg (Findwise)*

Henrik Strindberg's speech started with an overview on the activities carried out by Findwise, a company based in the Scandinavian countries, involved in many successful Enterprise search projects together with big technology partners (IMB, Google, Microsoft) and the focus of which is on **Findability solutions.**

He passed then to show some demographic data concerning the old age dependency (OADR) in Europe in the period between 2004 and 2005, according to which the number of persons supporting each citizen aged 65 and above will decrease from four to two, i.e., we will have a higher number of older people. Strindberg also stressed that the '40 years-olds' became the new 30 (i.e. in 2005 the remaining life for 40 years-old people was of around 44 years, like it was in 1952 for the 30 years old).

He pointed out that, given the fact that today we assist to a growth of data available as information doubles every third month (not corresponding to a parallel growth of quality of data) these are the main **internal information challenges**:

- Difficult to find and reuse existing information
- Lack of necessary decision support
- Too much time spent searching for and re-doing documents
- Information is stored in numerous isolated sources
- Separate and costly search applications is used in different systems
- Existing search offers poor functionality, usability and result relevancy

Strindberg sustained that a more holistic approach is necessary. He referred then to an internal case study, in order to better understand to cope with these challenges through a findability solution. Ericsson identified **six key performance indicators** to measure the effects of the findability solution:

- Speed – find information faster
- Collaboration – converse, reuse and collaborate more
- Discovery – discover new information
- Accuracy – find more accurate information
- User coverage – find relevant information about co-workers regardless of geographical location or work task
- Content coverage – find information regardless of source and classification

Measurements of performance indicators show:
- Quicker access to information and knowledge – increased productivity
- Higher profit margins from improved support to business processes
- One platform – reduced costs

He sustained then, that, as shown in the graph, the findability solution allows for an improvement in percentage as compared to the Intranet search tool user prior to the launch of the new solution
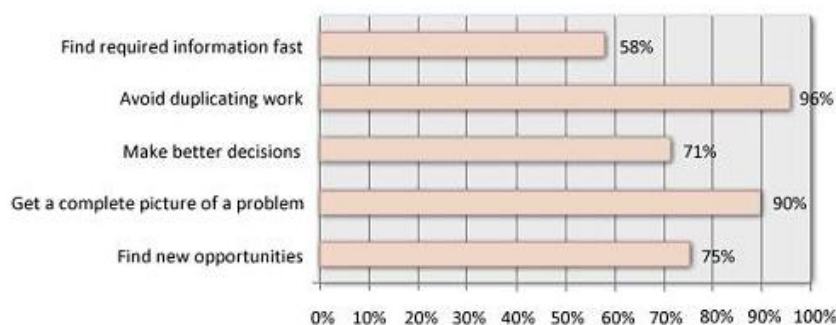
Strindberg concluded his talk by providing some reflections and recommendations.

According to him it is important for the future to adopt **a broader perspective**, such as that offered through Findability solution (by Findwise). He argued that search technology is not the silver bullet and that it should be accompanied by business processes. In other words, the use of search technology should support and leverage the existing business processes to assist in fulfilling and reaching their needs and goals.

Furthermore, he foresee a change in pension policies will happen in Europe (taking into account longer active working lifE, longer active lives and current economic situation) changing the personas of the users.Thus even more important to focus on:

User - The solution must be designed and tailored to fit the needs and capabilities of the users.

Information - The quality and structure of existing and newly produced information is an important success factor of the solution.

Organisation - The organisation must establish a process to govern the solution and maintain Findability for future needs.

14. Visioning Enterprise Search in Future Enterprises
## *Man-Sze Li (IC Focus)*

Man-Sze Li focused her talk on the future developments of Enterprise Search, trying to understand where the enterprise of the future needs to go.

She first affirmed that the topics of the previous talks were very related to the future of Internet too.

She reminded the audience of the Internet situation in 2009, in which was possible to find content (sharing files), Information (search engines), business services (Saas), communication (free email). She argued that, looking at the Internet development, it is clear that the IT companies have their B2B strategies and that it works.

What enterprises need to do is to **collaborate more**, to share information (at least to a certain extent).

According to Li, an Internet of future would imply that all the world is *Internetable* and serviceable
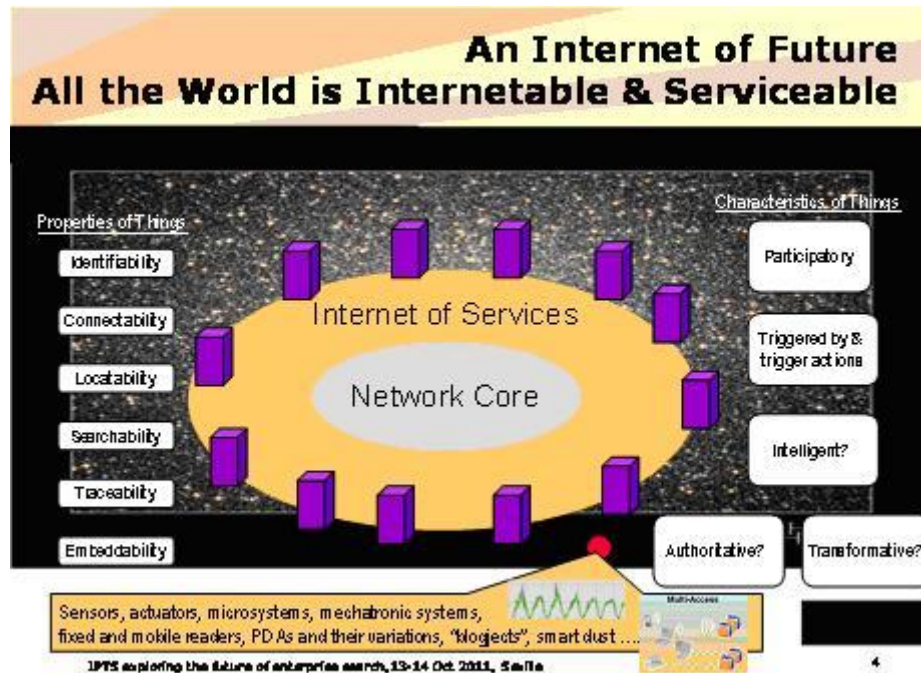
In terms of enterprise search, this is where it is possible to see the potential: in the Internet of all services and of Things. The enterprise should enhance their capability to process and use data from external sources. Li argued that a successful enterprise will be the **sensing enterprise**, that is a complex and smart entity capable of sensing and reacting to a wide ranging set of (business) stimuli. The characteristics of this enterprise would be the following:

- − "global" context awareness
- − decentralise intelligence
- − dynamic configurability
- − multi-identity oriented
- − "fusion" between the virtual & the real
- − Self-*, self-re-invention?

An important feature will be in particular the **decentralized intelligence**.

Li considered then the **implications for the enterprise search** that could be summarized as follows:
- • Enterprise inside out/outside in – from shifting boundaries to potentially borderless enterprises
- • Access / Retrieval and Push / Pull versus "Being there" and "Sensing it"
- • **Seamless transformation of (raw) data to (tailored) information to (experienced) knowledge**
- • New approaches to enterprise assets and new notions of assets
- • Enterprise systems as dynamically configurable collections of smart components, informed by smart things

Search as an always on, ubiquitous, inter-connected, multi-layered and multi-dimensional "background" capability.

6. Key findings of the Enterprise Search workshop 2011

The main findings of the 2011 enterprise search Delphi-type study organized by IPTS with Intranet Focus Ltd (July-October 2011) and of the expert workshop in Seville (October 2011) on "Exploring the future of enterprise search", can be summarized as follows:

- **Immaturity of the market.** Though there is a general consensus on the strong potential of ES the related market is not mature enough.
- **Market potential gap**. Though many enterprises in Europe with more than 1000 of employees have huge repository of data, very few of them use enterprise search solutions.
- **Skills gap**. Contrary to what happens in US, there are not enough trained people in Europe in the field of ES, as specific academic programs on ES are missing or very limited.
- **Few investments**. Though relevant European research projects in ES are not missing, there is not enough support to fill in the transfer gap between research and industry.
- **Lack of awareness**. More information about what a ES solution is (how it differs from web search, and what are the benefits for potential clients) is needed.
- **No missing technologies**. Though some unsolved technological challenges have been identified, technological development is not a limiting factor for the take up of ES.
- **Mobile**. ES is likely to be positively affected by the development towards a mobile world.
- **Integrated platform and Search Based Applications (SBA)**. One of the features of the successful enterprise search company will be the capability to provide semantic linking (combining structured and unstructured data) and semantic search (allowing intelligent analysis of query).

A JRC Scientific & Technical Report containing the main results of the aforementioned study and of the workshop on the current and future trends in the Enterprise Search domain (as of 2011) is forthcoming and will be available online, shortly. It will discuss the prospects of Enterprise search as well as the main challenges and opportunities.

7. Appendix A: Delphi-type Enterprise Search Questionnaire 2011

An online questionnaire was designed and developed by IPTS and Intranet Focus [7] in July 2011 and sent to 30 enterprise search professionals. The complete questionnaire is included in this Appendix.

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

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.

	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 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 than these technologies would be widely offered by most search vendors, but there was no consensus at all on a forecast date.

	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=	
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.

The Chorus+Project Consortium groups the following Organizations:

Partner Name	Short name	Country
JCP-Consult	JCP	FR
The French National Institute for Research in Computer Science and Control	INRIA	FR
Centre for Research and Technology Hellas - Informatics and Telematics Institute	CERTH-ITI	GR
University of Trento	UNITN	IT
Vienna University of Technology	TUWIEN	AT
University of Applied Sciences Western Switzerland	HES-SO	CH
Engineering Ingegneria Informatica SPA	ENG	IT
TECHNICOLOR	TECHNICOLOR	FR
JRC Institute for Prospective Technological Studies	JRC	EU

Author(s)	Date
Shara Monteleone (JRC) Stavri Nikolov (JRC)	9/11/2011
Ramon Companó (JRC) Shara Monteleone (JRC) Henning Mueller (HES-SO)	14/11/2011
Ramon Compano(JRC) Shara Monteleone (JRC) Ioannis Maghiros (JRC)	28/11/2011