



Achieving enterprise search satisfaction

The Search Network

October 2018
Martin White



Search Network members

Martin White, [Intranet Focus Ltd](#) (UK)

Martin is an information scientist and the author of Making Search Work and Enterprise Search. He has been involved with optimising search applications since the mid-1970s and has worked on search projects in both Europe and North America. Since 2002 he has been a Visiting Professor at the Information School, University of Sheffield and is currently working on developing new approaches to search evaluation. Follow Martin on Twitter [@IntranetFocus](#).

David Hobbs, [David Hobbs Consulting](#) (USA)

David helps organisations make higher impact digital changes, especially through early strategy to best frame these initiatives before they begin. He is the author of Website Migration Handbook and Website Product Management. His clients include the Center for Internet Security, the Library of Congress, the Mideast Broadcasting Company and the World Bank. Follow David on Twitter [@j davidhobbs](#).

Charlie Hull, [Flax](#) (UK)

Charlie is the co-founder of Flax, which builds open source search and Big Data solutions for clients worldwide. He writes and blogs about search topics, runs the London Lucene/Solr Meetup and regularly speaks at, and keynotes, other search events across the world. He co-authored Searching the Enterprise with Professor Udo Kruschwitz. Follow Charlie on Twitter [@FlaxSearch](#)

Miles Kehoe, [New Idea Engineering](#) (USA)

Miles is founder and president of New Idea Engineering (NIE) which helps organisations evaluate, select, implement, and manage enterprise search technologies. NIE works and partners with most major commercial and open source enterprise search and related technologies. He tweets as [@miles_kehoe](#), [@Ask Dr Search](#) and [@Search-Dev](#).

Helen Lippell (UK)

Helen is a taxonomy consultant. She works on taxonomy development projects, including taxonomy audits, ontology modelling, tagging initiatives, semantic publishing, metadata training and more. Her clients include the BBC, gov.uk, Financial Times, Time Out, RIBA and the Metropolitan Police. She writes and speaks regularly and is the programme chair of Taxonomy Boot Camp London. Follow Helen on Twitter [@octodude](#).

Agnes Molnar, [Search Explained](#) (Hungary)

Agnes is the managing consultant and CEO of Search Explained. She specialises in information architecture and enterprise search. She shares her expertise on the [Search Explained blog](#) and has written and co-authored several books on SharePoint and Enterprise Search. She speaks at conferences and other professional events around the world. Follow Agnes on Twitter [@molnaragnes](#).

Doug Turnbull, [OpenSource Connections](#) (USA)

Doug is CTO of OpenSource Connections and the author of Relevant Search. His goal is to empower the world's best search teams. He has assisted with search at organisations in a variety of domains. His clients include Wikipedia, Snagajob, Careerbuilder, and many search organisations. Follow Doug on Twitter [@softwaredoug](#)

Acknowledgements

I am very grateful to Charlie Hull, Miles Kehoe, Professor Udo Kruschwitz, Agnes Molar, Lee Romero and Tony Russell-Rose, and a number of former clients, for their insights into enterprise search management in the course of the development of this report. Cecilie Rask, Intranet Manager, Danish Police, provided the information presented in Section 19 on the development and adoption of its new search application. Val Skelton improved this report significantly with editorial advice as I struggled to find a structure and focus. Final production was undertaken by Simon Flegg at [Hub Graphics](#).

Introduction

Enterprise search can trace its origins back to 1951 and enterprise search applications have been on the market since the mid-1980s, so by any measure this is a mature technology and a mature industry. It is therefore surprising and concerning that surveys have indicated that employees are experiencing considerable difficulties in getting the best from search applications. When conducting interviews for an intranet project, managers very quickly turn the conversation towards poor quality search and in twenty years of listening to conference presentations about enterprise search enhancement I have never yet heard a search manager describe a dramatic increase in the number of search queries or a significant improvement in search satisfaction

An immense amount of research has been conducted on web search. The number of research papers is in the thousands. For comparison there are perhaps only a dozen or so research papers on enterprise search. A major reason for this is that it is very difficult to undertake research inside of organisations. The vacuum has been filled with a range of anecdotal pronouncements about what it takes to deliver high-quality enterprise search, with a strong focus on technology and on 'time saved'.

Computational ethnography is now being used to track how employees are seeking information, and what the role is for search in the process of seeking. This differentiation between seeking and searching is important. Employees have a range of ways of seeking information and it is important to understand why they decide to use a search application. For too long the focus has just been on the technology of search even though the impact of context on search has been recognised and researched for over two decades.

Since being appointed a Visiting Professor at the Information School, University of Sheffield in 2002, I have tried to bridge the gap between academic research in the field of information retrieval and the management of enterprise search applications. One of my objectives in writing this report is to show practitioners how much relevant research and insight is available to them from research papers. I hope that this report will encourage academics and practitioners to work together on innovative research projects which look in detail at the adoption and use of enterprise search applications. A substantial amount of research has been undertaken into enterprise resources planning and enterprise content management applications and on the adoption of enterprise social networks. Why enterprise search has not been given at least equal attention is a mystery.

This report summarises the outcomes of a personal selection of this academic research, together with the lessons learned from projects I have undertaken since 2001 and discussions with a range of search managers and search consultants, notably members of The Search Network. However, the report does not set out to be a comprehensive literature review. The wealth of detail in the academic research papers is such that in this report I can only provide a brief summary and a link to the published paper.

I hope that this report will stimulate discussion about how employees are using enterprise search applications. Organisations and search vendors may well have conducted internal research that contradicts my analysis. I would encourage the publication of this research and critiques of this report so that the global enterprise search community can benefit from all the available research.

Contents

- Introduction
- **1.** One day in September
- **2.** To begin at the beginning
- **3.** Metrics for search performance assessment
- **4.** The enterprise information environment
- **5.** Information quality
- **6.** Information seeking and information search
- **7.** The complexity of enterprise searching
- **8.** Computational ethnography
- **9.** Work tasks and search tasks
- **10.** Professional searching
- **11.** Time spent searching
- **12.** Query reformulation
- **13.** The relevance of relevance
- **14.** Document snippets
- **15.** The Scent of a SERP
- **16.** Stopping strategies
- **17.** The implications of information seeking failure
- **18.** The value of search logs
- **19.** Search literacy
- **20.** Personalized enterprise search
- **21.** Considering the evidence
- **22.** Measuring search satisfaction
- **23.** Next Steps

1. One day in September

The morning of Monday 10 September 2001 in Washington DC was simply glorious, with a clear blue sky. The Intranet Focus team of four consultants walked down 19th Street and into the lobby of the International Monetary Fund for the start of a three-month long intranet strategy project. We very quickly felt at home, helped by the warmth of the welcome from the IMF project team and the fact that one of the team was a retired long-service economist from the IMF who seemed to know everyone and everything about the Fund.

In addition, we had read [“Inside the IMF”](#). This was the title of a book authored by Richard Harper in 1998 whilst working for the Rank Xerox Research Centre in London. This 300 page book, subtitled “An Ethnography of Documents, Technology and Organisational Action”, provided us with an immense amount of insight into how information flowed around the Fund and how it was used to further the Fund’s objectives. This remains the only book ever published on the working practices of a major organization. During the course of the project we suggested to the Fund that a user survey would be useful to fill in some gaps in our research. We were told that surveys rarely achieved more than a few per cent response levels. Nevertheless, the Fund kindly agreed to our proposal and was then somewhat surprised to find that we achieved a response rate of over 50%.

One of the reasons for this was that for the first time Fund employees were able to comment on the dissatisfaction they experienced with the search application. Although functionally rich the results contained a large number of relatively low-quality documents and in addition the user interface left much to be desired. In principle the search application was not in the scope of the project, but this quickly changed. For me it was a classic learning experience and from that project forward I have always made sure that all intranet projects I work on have the search application in the scope.

You may be wondering how I remember the weather on 10 September. It is because the following morning the weather was just as exhilarating, but that day is now and forever known as ‘9/11’.

2. To begin at the beginning

The [history of enterprise search](#) dates back to a dissertation written by Philip Bagley when he was a Masters student at MIT in 1951. It was entitled ‘Electronic digital machines for high-speed information searching’. It was not until the 1980s that stand-alone enterprise search applications came on to the market, notably from IBM, Verity and Fulcrum. Then in the early 2000s intranet-specific search applications, such as Ultraseek, started to emerge. The very different challenges of intranet search to web search were recognised early on, notably in [Enterprise Search: Tough Stuff](#) by Rajat Mukherjee and Jianchang Mao, two search engineers working at Verity.

From 2007 to 2013 there was considerable consolidation in the enterprise search application market, notably the ill-fated acquisition of Autonomy by Hewlett Packard. Even so there are still around 30 major vendors offering commercial, open-source, appliance and cloud applications. In terms of installed base Microsoft dominates the market with the search application embedded in SharePoint and Office 365.

The core ranking model used by most of the vendors is [BM25](#) (for Best Matching), developed by Stephen Robertson and Karen Spärk Jones in the 1980s and 1990s. Although there are many variants of BM25, and a number of more complex ranking models, broadly speaking the enterprise search business is a mature one. Recently there has been substantial promotional effort from search software vendors about the benefits of artificial intelligence and machine learning as a means of improving relevance but there have been no published case studies that document the scale of improvement gained through adoption of these algorithmic techniques.

Given the maturity of enterprise search it is surprising that surveys from [AllIM](#), [Findwise](#) and [NetJMC](#) over the last five years all indicate that users are finding it very difficult to locate the information they need. It is quite noticeable that in presentations at conferences (or even in the corridors afterwards) speakers are very reluctant to disclose the volume of queries they are achieving. This might indicate that query levels are much lower than they anticipated. Reporting increases in these levels following the adoption of a new search application would surely be in the interests of both the search vendor and the corporate search team.

Sam Marshall (Clearbox Consulting) commented in a [blog post in 2018](#)

When I do employee focus groups for clients on their digital workplaces, it takes around three minutes before somebody complains how awful the search is. Even if that isn't what I asked about. Everyone grumbles and empathises, and the conclusion is invariably 'it should just work like Google'

I have had similar experiences in organisations in many different business sectors. Compared to the immense amount of research that has been published on how people search on the web very little research has been undertaken into how employees use enterprise search applications. By comparison there has been extensive research into the adoption and use of enterprise resource planning applications (ERP) and enterprise social networks (ESN). The situation is now changing and over the last few years (and especially in 2018) some very valuable research papers have been published which provide indications of why enterprise search applications are not delivering the anticipated levels of search satisfaction. The publication of this research has been the primary catalyst for this report.

3. Metrics for search performance assessment

In the case of enterprise search performance there is a wide range of information retrieval metrics that can be used, almost all of which evaluate the extent to which the search application is able to deliver relevant documents to the search user. This immediately raises two critical issues. The first is one of the semantics between [information retrieval](#) as an academic topic and search as the process of information discovery. The second is defining 'relevance'. The scale of the relevance debate is illustrated by the [book](#) published in 2016 by Tefko Saracevic entitled "The Notion of Relevance in Information Science: Everybody knows what relevance is. But, what is it really?" Some of the issues of defining and assessing relevance are considered in Section 13.

In IT enterprise architecture terms an enterprise search application is regarded as an 'information system'. Over the last four decades a significant amount of work has been undertaken to evolve success metrics for information systems. The most important of these contributions has come from [William Delone and Ephraim McLean](#) in a research paper they published in 1992 and which they reviewed again in 2012. In their view the

overall success of an information system was dependent on a range of individual elements, all of which were interlinked. These included system (technology) quality and information quality, and these were linked to user satisfaction.

An assessment of search satisfaction provides a benchmark at an organisation level about the attitude of users to search performance, in a way perhaps similar to an employee engagement survey. Although comparisons between the outcomes of engagement surveys in a range of organisations can be of interest the major benefit is in focusing on improving engagement (and so search satisfaction) within the organisation.

In my opinion the value of 'search satisfaction' as a metric is that users take a pragmatic view on the ability of a search application to meet their requirements, recognising that even Google is not perfection. No one expects employee engagement to be 100% but people (or employees) do want to feel confident that the organisation is aware of the challenges that have emerged in the survey and has a plan of action to gradually improve the score.

Over the last two decades a significant amount of research has been undertaken into how people search on the web, especially by Microsoft Research. One visible sign of this research effort is [Interactions with Search Systems](#) published in 2016 by Ryen White, leader of the Cortana Research team at Microsoft. The bibliography lists over 1600 research papers but none of these addresses the way in which employees interact with enterprise search applications. The reason for this is that there have been no papers published on the topic except for a few examples which take a very narrow view of the topic.

In July 2018 Dr Paul Cleverley and Professor Simon Burnett (Robert Gordon University, Aberdeen) published what is without doubt a landmark [research paper](#) on the factors that influence user satisfaction with enterprise search applications. There are three reasons for me applying the 'landmark' label to this paper. The first of these is the scale, with more than 1000 users in a large multinational company providing feedback over a period of two years. Nothing on this scale has ever been undertaken, though there is an [earlier paper](#) from the research team that was published in 2015. This considers the relationship between user satisfaction and task performance. Both papers include extensive bibliographies.

The second reason for regarding this as a landmark paper is that Dr. Paul Cleverley is a geophysicist who moved into information science roles in the oil and gas industry and in 2017 was awarded a [PhD for his work](#) on the use of filters and facets in enterprise search. So here is a discipline expert with a very solid understanding of research methodologies applying all his experience and expertise to understanding enterprise search behaviours. The final reason is that since six out of the ten largest companies in the world are in this sector there is at least a reasonable expectation that the outcomes will be similar in other large multi-national companies.

The authors identify three factors that predominately influenced satisfaction. They were *technology*, *information quality* and *information literacy and task utility*.

The categories and their descriptions are set out in the table below.

In total technology factors were the largest single group (38%). Information factors (36%) and literacy factors (26%) accounted for 62% of the reasons for dissatisfaction.

Table 2. Coding inference rules to move from comments to categories

Polarity	Category	Description
Satisfaction	Pre-disposed	
	- Internal	Comments 'comparing' performance >> than previous enterprise search
	- External	Comments 'comparing' performance == Internet 'Google' experience
	Technology quality	
	- Speed	Comments mentioning speed of performance deemed 'fast'
Dissatisfaction	- Ranking	Comments specifically mentioning ranking, items at or near 'top' of results
	- User Interface	Comments mentioning friendly user interface or 'ease of use' - easy to use.
	Task utility	Comments mentioning item was found/what was needed Task needs met
	Pre-disposed	
	- Internal	Comments 'comparing' performance << than previous enterprise search
	- External	Comments 'comparing' performance << Internet 'Google' experience
	Technology quality	
	- Speed	Comments mentioning speed of performance deemed 'slow'
	- Reliability search tool	Comments mentioning errors using search tool (permission lags, timeouts)
	- Reliability systems IT	Comments mentioning errors after clicking on results (e.g. broken app)
- Ranking	Where the item the user was seeking was not ranked optimally	
- Syntax handling	Where simple spelling mistakes were not corrected by the search parser	
Information quality		
- Accuracy (Versioning)	Where multiple versions of the same information/out of date - led to issues	
- Completeness (Item)	Where the information sought was missing/not created/insufficient	
- Completeness (Index)	Where the information sought was in a collection which was not indexed	
- Labelling/tagging	Where information was poorly labelled in the title/tags (or missing)	
- Semantics handling	Where Acronym/synonym/hypernym vocabulary issues led to issues	
Information literacy		
- Query term choice	Where the query terms used were sub-optimal for the need identified	
- Results scanning	Where the user did not notice the correct result on the search results page	
- Seeking channel	Where the search tool was not the optimal channel for given info. needs	
- Tool literacy	Where lack of re-formulation/use of filters likely led to not finding info.	
Task utility	Comments mentioning item was not found/Task needs not met	

There appeared to be significantly more areas for dissatisfaction than for satisfaction.

Journal of Information Science, 2014, pp. 1-25 © The Author(s), DOI: 10.1177/0165551510000000

My view of this paper is that search satisfaction is a function of search technology, information management and information literacy. Unless all three are addressed there is no possibility that a reasonable level of search satisfaction can be attained.

The issue of search literacy was raised by Anne April Lewis in her [Masters thesis](#) in 2010. The thesis describes research carried out at the Oak Ridge National Laboratory (ORNL) into the use of the organisation's intranet search application. 65,000 queries undertaken over a three-month period were analysed. The outcome was that users are uncertain of how to best formulate queries. They also tend to be unfamiliar with how a search engine works. Search engine administrators were not familiar with the user's perceptions of how information is indexed on the intranet. The average results returned per query for ORNL were too high (3,395) and in the view of the author this reflects a knowledge gap for both users and implementation designers. A [research study](#) carried out at the University of Bristol and published in 2016 also considered the linguistic differences between web and intranet searching behaviours and as with the ORNL study noted the increased complexity of enterprise search queries.

4. The enterprise information environment

Before looking in detail at the process of search it is important to understand some of the complexities of managing information in an organisation, especially larger diversified organisations working in many countries, cultures and languages.

There are three categories of information within any organisation.

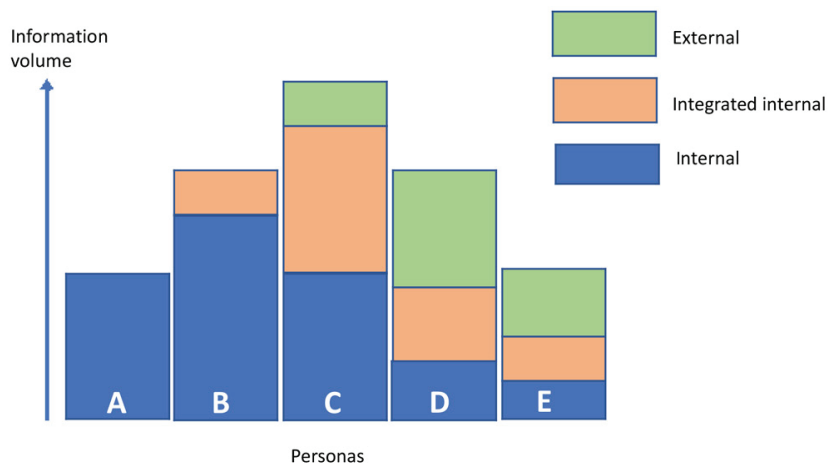
Internal information usually relates to organisational performance and internal management. Many employees in support roles working with systems that are based around defined workflows may well only use this category of information. Many other employees may need to find a specific item of information on an occasional basis. An example might be salary ranges for a grade of employment, which may change on an annual basis.

Integrated internal information will be used by managers to assess the performance of teams, departments and projects against objectives. Although the basis of this integration may well be data-based and be presented in dashboards there will always be a requirement for a text-based opinion of the trends and implications.

External information comes from sources external to the organisation, such as news, market surveys and competitive assessments. Patents and research reports/papers also come into this category.

The chart below presents five 'content personas', indicating a typical blend of the three categories of information and a sense of the relative volume of information that the personas have to manage effectively.

I developed this approach in 1979 when undertaking a consulting assignment for a strategic development group in a company in the B2C sector where external information on market trends and competitors was not being well integrated into development opportunity appraisals.



Persona A would represent employees in manufacturing, call centres etc. working with one or two enterprise applications for much of their working day. They would of course be interested in news about the organisation and need to find out about personnel policies (for example) but these would be a very small element of their information volume.

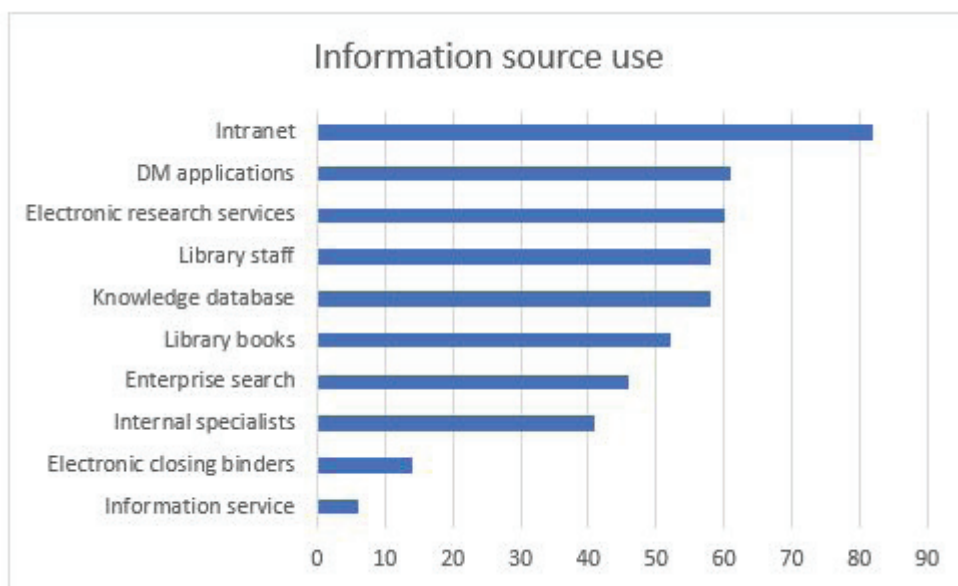
Persona B brings together output information from the processes being used by Persona A so that trends and opportunities are identified. These managers will be working with small number of applications that they know well and writing reports for colleagues and managers. Another Persona B might be technical support teams.

Persona C managers build on the outcomes of Persona B and put them in the context of the wider organisation and also external economic, market, technical and competitive information. Much of this work will involve working with a range of local and virtual teams, possibly with a range of language skills.

Persona D managers are often in technology, product and market development. They are probably going to be working backwards from external information to provide advice to senior managers on the strategic and operational directions the organisation should be taking. Although information volumes are lower than for Persona C the balance of internal, integrated internal and external information is probably fairly equal.

Persona E are senior managers up to Board level, who are dependent on the quality of the work carried out by Personnas C and D and is now synthesised into documents and presentations. Persona E will still be interested in internal information outcomes, often as dashboards.

The chart below comes from research carried out on a regular basis by a global law firm and illustrates the wide range of applications that are being used by both lawyers and support staff.



The appearance of library books in this list may seem strange but lawyers make extensive use of books in which the authors review the legal precedents that are the basis of common law. Indeed, in most business areas there are essential reference works that provide informed assessments of techniques and processes. A cursory check of Amazon UK indicated over 200,000 books on engineering alone. Each profession has a set of core information resources that are critical to the success of the organisation and understanding what these are, and for what purposes they are used, is very important in the process of information seeking. These books and similar resources are invisible to an enterprise search application.

The use of web search is not considered in enterprise search projects and yet the evidence is that there is a constant use of public web search engines and even the search application in the corporate web site. Some years ago, user testing at a global manufacturing business with around 45,000 employees revealed that more employees used the product information on the corporate web site (which was not indexed by the enterprise search application) than the intranet. The main reasons for this was that there was a team responsible for the level of detail and quality of product information on the web and that the search application on the web was better supported than the internal search application.

5. Information quality

Just because a document appears to be relevant does not mean to say that it is useful. Content quality is a major issue in search implementation because it is often very obvious that the most 'relevant' references listed on the SERPs (Search Engine Results Pages) are widely different in terms of quality and value to the employee.

Some [recent research](#) on the use of Enterprise Content Management applications (which invariably have good search functionality) strongly indicates that the problems lie not with the functionality of the ECM with regards to either adding content or finding it but with the quality of the content that is retrieved. The research suggests that there are two aspects to quality assessment.

The first is representational quality.

Conciseness – the information is widely dispersed throughout a document, making it time-consuming for a user to be sure they have found all the relevant information.

Presentation – the information is presented in a format that is not easy to use, for example a PowerPoint presentation which has been rendered as a pdf so that copying and pasting important content is very time-consuming.

Understandability – the information is written in a way that makes it difficult for a user who is not familiar with the topic to understand it. This is common in the medical sector where the use of clinical language and in particular abbreviations and acronyms can be a significant barrier to comprehension.

The second is contextual quality.

Completeness – a document refers to another document to which the user does not have immediate access to.

Relevance – it does not immediately enable the user to take a decision because (as an example) the information refers to the situation in the business several years earlier and may be longer applicable.

Timeliness – the information may not be current and so the user cannot rely on it being the latest or current version.

There are many other approaches to categorising information quality. A good starting point in gaining an understanding of information quality is [The Philosophy of Information Quality](#), edited by Luciano Floridi and Phyllis Illari and published in 2014.

There is of course no realistic hope of going back through perhaps millions of documents and removing all those that do not meet quality standards. It is not just a question of resources but also the lack of definition of quality standards, or even quality guidelines. Organisations need information management policies and standards to be in place and to have identified categories of content where quality is paramount in meeting business objectives and avoiding breaches of regulatory compliance.

The vast majority of research papers use English as the query language, but never consider the linguistic expertise of the search user. Many companies have a de facto language policy that states that English is the corporate language. It may be the default

corporate language but in any multi-national company many different languages will be used. The impact of language on business operations is now the subject of many [research studies](#).

The table below shows the percentage of documents by language in a major pharmaceutical company with the headquarters in EU. The total number of content items is close to 100 million.

Language	Content items as % of total	% speaking the language as their primary language
English	73	24
German	13	25
Spanish	4	11
Portuguese	3	4
Japanese	2	6
Italian	2	5
French	1	6
Chinese	1	4
Polish	1	3

The right-hand column does not total 100% as there are a significant number of employees in Arabic-speaking countries and also in India and the Nordic region, where English fluency is very high but is not a national language.

There are two implications of this language diversity. The first is that although the number of content items in (say) Portuguese is small as a percentage the importance of these content items to the Brazilian subsidiary of the company is very high. The second is that although the majority of the content items are in English only around 24% of the 120,000 strong workforce have English as their mother language.

As a result, the majority of the searches are carried out by employees working in their second language. A German national may have a fluent command of spoken English but may not have a good enough vocabulary to construct the optimum query. In 2018 Stephanie C Segura Rodas, an MBA student at the Management School, University of Sheffield wrote a dissertation on the potential impact of multiple ad hoc language policies on search performance. In the course of a very thorough literature review no previous research on this topic was found.

Another complication is that in some cases ([business and finance](#) for instance) many English words have been assimilated into other European languages but may not have quite the same definition as they would to an English financier.

This leads into the topic of metadata management. There is no dispute that the quality of the search experience depends heavily on the quality and consistency of metadata. The Danish Police were one of the the award winners in the 2018 [StepTwo Intranet and Digital Workplace Awards](#). The core priority for the new intranet was to provide the best possible search experience. To achieve this a metadata model was developed and applied rigorously across the force to ensure that users were able to find both global (country-wide) and local information presented in a common SERP. The search consultant was [Agnes Molar](#) who has a great deal of experience in optimising SharePoint and Office 365 search implementations. It is noteworthy that the team member with

specific responsibility for search is a serving police officer. The outcome is that the team has even heard feedback such as “it’s almost like Google” and users say that the search is now their preferred method of finding information.

6. Information seeking and information search

Most companies of any size and complexity rely on a range of enterprise applications. The term ‘enterprise’ is used to denote an application that is managed by IT for the benefit of those employees who will make use of it in their work. With a few exceptions (email, SharePoint etc) none of these applications will be used without training by all employees. Each cohort will only use these applications after training ahead of using them for a significant proportion of their working day. Many of these applications will have a search application built in.

When it comes to ‘enterprise search’ the model is completely different. The aim here seems to be the delivery of an application that can be used by any employee in the organisation without any training so that they can find exactly what they need from the entire information resources of the organisation, or at least those resources they have permission to access.

Over the years many eminent academics have developed models that try to codify the way in which information flows around organisations. Twelve of these models are described in detail in the book ‘[Looking for Information](#)’ by Donald Case and Lisa Given which in its most recent 4th edition runs to 500 pages and around 1600 citations of research papers. Professor Nigel Ford, from the Information School, University of Sheffield, has written a very readable [introduction to information behaviour](#).

In 1999 [Professor Tom Wilson](#), also working at the Information School in Sheffield, developed a very useful schematic for the positioning of information behaviour, seeking and searching.



This positions information search behaviour, the process of using a search application, as just one element of information-seeking behaviour, and that in turn reflects organisational information behaviours. The definitive analysis of information behaviours in organisations is '[The Inquiring Organisation](#)' by Professor Chun Wei Choo, published in 2016. The sub-title of the book is 'How Organisations Acquire Knowledge and Seek Information'.

Using a search application is just one way to seek out information. We realise that we do not seem to have all the information we need to make a decision or undertake a task. This information gap could be filled by (as examples)

- Reading through documents we have stored on our personal or team files
- Browsing through email folders
- Using an enterprise application (HR, ERP, e-Learning etc)
- Sending an email to one or more people we know
- Talking to a colleague or an acknowledged expert
- Posting a request on a social media channel
- Browsing through an intranet
- Checking through a department or team wiki
- Asking for assistance at the next team meeting
- Searching on the web
- Searching on a specific application (e.g. a corporate intranet)
- Searching across multiple applications

The act of searching must be put into this wider context so that we not only know how employees search but why they chose search as their option and what they then do with the information they find. It is important to note that only in the case of search would there be a requirement for training. All the other information seeking options are commonplace.

One of the earliest [studies of human – machine interaction](#) in responding to queries was carried out in 1967 by Robert S. Taylor, working as the Director, Center for Information Sciences, Lehigh University, and it remains valid today in the context of both enterprise search and digital assistants.

He proposed a taxonomy of eight classes of information use which might well apply to enterprise search situations.

Understanding the context – information is used to develop a context or make sense of a situation

Understanding the problem – information is used to develop a better comprehension of a problem

Instrumental use – information about how to undertake a task

Factual use – information to determine the accepted truth about a phenomenon or truth

Confirmational use – information used to verify information already discovered or known

Projective use – information used to predict what is likely to happen in the future

Motivational use – information to initiate or sustain personal involvement

Personal use – information to develop relationships or to enhance status, reputation or personal fulfillment.

In effect these are different use cases, and it is arguably not a taxonomy. The categories are not mutually exclusive, and information gained from one of these cases may be

used to address other needs. They do indicate that the reasons why employees use a search application are very diverse. An important implication of these use cases is that the assessment of what is relevant content will almost certainly be different in each case.

Looking through the literature there are many other ways of categorising information use in the context of seeking and search. The important point to note is that it is not just a ‘high precision’ search or an ‘exploratory’ search but just about every point on the line between these two extremes. Not only will users have different views of what are relevant documents they will have different reasons for looking for them.

One of the major changes in organisations over the last decade, and especially the last five years, has been the widespread adoption of collaborative working and collaboration applications. These applications enable participants to have access to a centralised (for the team) store of documents, not only those created by the team but also related material that has been posted to the ‘team site’ (used in a generic sense!). If members of a team realise that they need additional information one member of the team is likely to undertake to find it and share it with the team. This will inevitably lead to a reduction in the number of unique users and searches. A significant amount of research has been conducted into [collaborative information seeking](#) and the benefits (and challenges) of having two or more members of the team work together on a seeking task but none of the leading search vendors offer this capability

An implication of this move towards collaboration is that a high percentage of the information needed for decision making is probably being contributed collectively by members of the team, and so the need to undertake a search for this information is being significantly reduced.

The scenario presented by many search vendors is of a solitary employee seeking business-critical information and having to depend on a search application to meet their information needs. Given the range of potential options in an organisation this is not a realistic scenario.

7. The complexity of enterprise searching

The information seeking models referred to in Section 6 do not go into an explicit analysis of the search process itself. Progress is now being made on the development of a [Complex Searcher Model](#) which has been developed progressively by a number of research teams in an effort to try to understand the ways in which searching is undertaken.

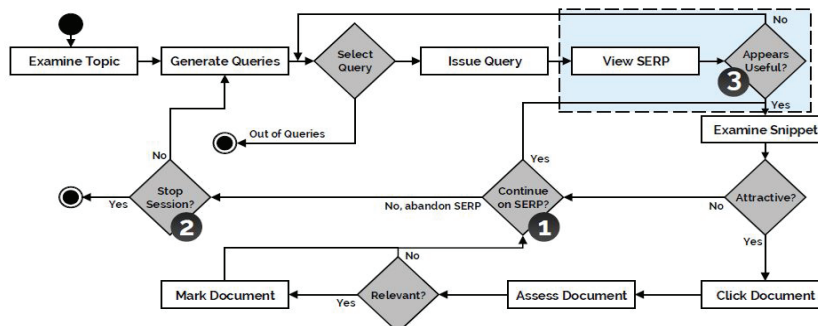


Fig. 1. The updated *Complex Searcher Model (CSM)* with the key decisions (shown in grey) and actions (shown in white). Stopping level decision points are numbered 1–3. The new SERP level stopping decision point is highlighted in the dashed box.

This diagram illustrates the complexity of the search process, especially when the SERP stopping strategy is included. Stopping strategies are discussed in more detail in Section 15 but from this point onwards it is useful to have this diagram in mind when considering how best to optimize the search process and improve search satisfaction. In any process with this number of potential pathways it is quite possible to focus on a specific point in a pathway and not make any overall impact on search success. It is also important to consider the extent to which search logs are able to provide a dependable metric on each of the three stopping points.

This is of course a generic model. More research needs to be undertaken in an enterprise search context to understand more about the factors that come into play at each stopping point.

8. Computational ethnography

Section 1 makes a reference to Richard Harper's ethnographic study of the International Monetary Fund without defining the meaning of 'ethnographic'. The term originates from a combination of the Greek words *ἔθνος* ("folk, people, nation") and *γράφω* grapho ("I write"). It describes a method initially used by social science researchers (cultural anthropologists in particular) to examine the meaning in the lives of a cultural group. Researchers conducting ethnographical studies, or ethnographers, strive to develop 'thick' descriptions of everyday life and practice. The challenges in adopting ethnographic research include the amount of time and effort needed to undertake the observations and interviews, the possibility of a bias being introduced by the research team and the analysis of the outcomes.

Over the last few years [computational ethnography](#) has emerged as a powerful means of overcoming, or at least reducing, the challenges of interview-based ethnography. Computational ethnography leverages computer or sensor-based applications to unobtrusively or nearly unobtrusively record end users' routines in undertaking work tasks. Because data is automatically captured it potentially provides higher objectivity, less intrusion, more inclusiveness (i.e., into spaces and time where/when direct observation by human observers is not possible), and better scalability for data collection, aggregation, and analysis. Among these applications are keyboard and screen-logging software developed to mitigate the loss of sensitive data. However computational ethnography research needs to be integrated with qualitative research designs such as interviews, diaries and observations. This is termed a mixed-methods research strategy and is an essential tool for enterprise search managers.

9. Work tasks and search tasks

There has been a significant amount of research into task-based search for well over a decade, with much of it coming from Professors [Pertti Vakkari](#) and [Kalervo Järvelin](#) working at the University of Tampere (Finland). Professor Elaine Toms has also undertaken [research into search tasks](#).

In the introduction to a [review article](#) published in 2003 Professor Vakkari wrote

The rationale for using information systems is to find information that helps us in our daily activities, be they tasks or interests. Systems are expected to support us in searching for and identifying useful information. Although the activities and tasks performed by humans generate information needs and searching, they

have attracted little attention in studies of information searching. Such studies have concentrated on search tasks rather than the activities that trigger them. It is obvious that our understanding of information searching is only partial, if we are not able to connect aspects of searching to the related task. The expected contribution of information to the task is reflected in relevance assessments of the information items found, and in the search tactics and use of the system in general.

There has been a [Task Search track at TREC \(Text Retrieval Conference\)](#) for a number of years. The [introduction](#) to the 2017 Task Track reveals a substantial research issue

Research in Information Retrieval has traditionally focused on serving the best results for a single query, ignoring the reasons (or the task) that might have motivated the user to submit that query. Often times search engines are used to complete complex tasks (information needs); achieving these tasks with current search engines requires users to issue multiple queries. For example, booking travel to a location such as London could require the user to submit various queries such as flights to London, hotels in London, points of interest around London, etc. Similarly, a person who is trying to organize a wedding would need to issue separate queries in order to locate stores to buy a wedding gown, arrange catering, book honeymoon, etc. In some cases users may not even be aware of all the subtasks they need to achieve to satisfy their information need, which makes search an even more difficult experience.

Ideally, a search engine should be able to understand the reason that caused the user to submit a query (i.e., the actual task that caused the query to be issued), and rather than just showing results relevant to the query submitted, the search engine should be able to guide the user to achieve their task by incorporating the information about the actual information need. The goal of this track is to devise evaluation methodologies for evaluating the quality of task-based information retrieval systems.

The focus so far has been on individual search tasks and in a web environment, and that is substantially different to the enterprise environment.

Many different approaches have been tried out to understand the context of enterprise search, including self-completion surveys or interviews at the end of a morning or afternoon. The problem with these approaches is that search users are relying on their memory of how long they spent seeking and/or searching. It is very difficult to obtain reliable and consistent results. Over the last few years academic research teams, primarily in Finland and the UK, have started to examine in detail how employees go about their daily work within the context of seeking and searching for information.

Professor Järvelin has published three very important [papers](#) co-authored with one of his PhD students Miamaria Saastamoinen. The approach they have taken marks a breakthrough in enterprise search research in that they logged the use made of digital resources in the work place, rather than rely solely on diaries or self-completed surveys. This is good example of the value of computational ethnography though it is interesting to note that this term does not appear in any of the three papers. A search for case studies in the use of computational ethnography will not find them!

To quote from Professor Järvelin

To design information search systems to properly serve WT requirements, it is necessary to study how the WTs as actions are connected to searching. Therefore, information (retrieval) systems development and evaluation should not take place in isolation but take the work context into account and find out for what purposes and how the systems are used. Failing to do this may result in developing suboptimal systems for expected but biased search needs.

This statement goes to the core of the likely causes of search dissatisfaction. Without understanding the relationship of work tasks and related search tasks in implementing enterprise search applications there is a substantial risk that the application will be fit to the specification (which, as an IT application focuses on functionality and technical performance) and yet not be fit for the purpose of providing an effective and satisfactory application for users.

As noted above almost all of the research has focused in on specific search tasks (ST). However, Professor Järvelin has noted one Work Task (WT) might will involve multiple Search Tasks.

His insights into task complexity are of importance

“Interestingly, if we assume that search (or seeking) tasks are part of work tasks and these subtasks vary in their complexity, we come to a situation where the effects of complexity work in both ways. The complexity of work task affects information seeking, and the complexity of seeking affects the complexity of the work task. In this article, we discuss the phenomenon only one-way, that is, how work task complexity affects information resource use.

However, the interaction of these phenomena should be kept in mind. Logic requires that if any subtasks are complex, they increase the complexity of the work task. Thus, if information seeking is considered complex, the task performer may feel the whole work task complex (or difficult) even before starting it, and if information seeking proves complex, it certainly affects her estimate of complexity afterwards.”

There are further complications. The search tasks within a work task may well be carried out using different search application and this makes it very difficult to decide on how long a search session takes and this depends on when (from a user, not a system perspective) it is actually completed.

In the research the main categories of resources were email/instant messaging, web search, other web sites, the PC, and finally other organisational resources. The logging software provided ‘dwell time’ information, defined as the time in seconds that the participant kept the resource open as an active window during the work task. The log files were supplemented by a limited amount of workplace shadowing to provide a context to the way in which the digital resources were being used. The study participants were also asked to record information about the complexity of the tasks they were undertaking, and these tasks were also broadly categorised into communication, support, editing and intellectual tasks. In effect this is a very practical application of computational ethnography though this term is not used in any of the three papers.

Within the constraints of this report it is not possible to summarise the range of data analysis that was undertaken. What does stand out is the amount of time spent on email and instant messaging, and the limited time spend on web search. The use of an

internal enterprise search application was not measured as an individual parameter, which is disappointing, but the evidence from this paper suggests that the time spent searching for information is more likely to be measured in minutes rather than hours. (See Section 11).

It is important to be careful in scaling up this outcome more widely. The research reported in the paper was carried out in a university, a public administration and a commercial company, so the sample was very small. The primary purpose of the research was to validate the methodology and not to derive a definitive statement about the time spent on searching.

10. Professional searching

For many people, especially those in law, medical research, patent research and recruitment, search is an essential application to discover information both internally and externally. Over the last couple of years, the concept of a '[professional searcher](#)' has emerged. In 2018 Tony Russell-Rose, Jon Chamberlain, Leif and Azzopardi published the results of a [comprehensive assessment](#) of how these four groups of professional searchers went about their work. In total 340 people took part in the study, making it both the largest study to be undertaken of individual search traits and the first comparative study across a number of disciplines.

The characteristics of professional searchers are

- Confidentiality: the search process and its outcomes are stored securely and not disclosed to potential competitors;
- Timeliness: the evidence gathered is up to date and not superseded by more recent work;
- Repeatability: the search strategy and its outcome may be reproduced by colleagues;
- Transparency: the process is open to scrutiny by others who can see why certain results have been returned
- Comprehensive: all relevant results are returned for a specific information need.

The survey investigated their search practices and preferences, the types of functionality they value, and their requirements for future information retrieval systems. The results reveal that these professions share many fundamental needs and face similar challenges. In particular there is a continuing preference to formulate queries as Boolean expressions, the need to manage, organise and re-use search strategies and results and an ambivalence toward the use of relevance ranking. The results stress the importance of recall and coverage for the healthcare and patent professionals, while precision and recency were more important to the legal and recruitment professionals. The results also highlight the need to ensure that search systems give confidence to the professional searcher and so trust, explainability and accountability remains a significant challenge when developing such systems.

One of the most interesting outcomes was that the use of Boolean logic was the number one criterion for formulating effective queries across all professions. It is particularly high for patent and healthcare information professionals, which reflects a need for transparent and repeatable search behavior, and an associated requirement to demonstrate due diligence and accountability for their search practices. Some search vendors are keen to present Boolean logic query creation as 'old search' but the reality at present is that algorithmic alternatives do not provide the level of control and creativity offered by Boolean logic.

The research presented in Sections 9 and 10 is in line with my own experience working on a range of enterprise search projects. The complexity of enterprise search use is always seriously under-estimated.

11. Time spent searching

It is quite common for search technology vendors to highlight the benefits of time being saved by the use of their applications, and this is usually because they are aware of the research carried out by Susan Feldman and Chris Sherman at International Data Corporation in 2000/2001. The research was published in [The High Cost of Not Finding Information: An IDC White Paper](#) in June 2001. In the paper the authors state

We use a general estimate that the typical knowledge worker spends about 2.5 hours per day, or roughly 30% of the workday, searching for information.

No basis is given for this estimate, but this percentage has been widely used by companies to make a business case for search as a means of increasing the productive time of employees. A forecast outcome of reducing the time by half is then worked out as a salary-based cost. Search vendors have also been quick to highlight the importance of 'blazing fast search' in reducing the time spent on searching.

It is important to read the full paper as the estimate was based on the assumption that approximately 35–50% of the information available within an enterprise was not centrally indexed and resides in databases and desktop or notebook computers. This is certainly not the current situation. Interestingly in 2012 IDC reduced the estimated time spent searching from 30% to 13% on the basis of its annual Information Worker Survey. This figure, by comparison less convincing, is rarely quoted by either search vendors or search managers.

Another factor to take into account in using these figures is whether the people surveyed by IDC understood the potential difference between information seeking and information searching. The search may have been very effective at identifying documents but the employee is then faced with looking through these documents to find the exact information they need. Although technology has moved on with thumbnail presentations of HTML-rendered pages once the initial fascination with the technology has passed users tend to find these presentations as unreadable and unhelpful.

In a [paper published in 2012](#) Pia Borlund et al considered what the time spent searching actually indicated, and it is of note that she did not cite any of the IDC research.

To quote from the introduction to the paper

In evaluation of information retrieval (IR) systems the time a user spends, mostly reading, is commonly interpreted as an implicit measure of interest. Similarly, the user's overall interaction time is interpreted in this way. In short, this means that when a user's use and interaction with IR systems and retrieved information objects is evaluated, it is assumed that longer time spent expresses more interest on behalf of the user compared to shorter time spent.

However, more time spent might also, for example, be a consequence of the retrieval of large sets of information objects that requires more handling time of the user, or the user's dealing with a complex information problem. Likewise, less time spent is interpreted as the user showing less interest (perhaps being

bored?), although it may be related to the retrieval of fewer information objects, or the result of a search problem that is easily grasped and quickly dealt with by the user.

This statement goes to the heart of the matter; what implications can be drawn from the speed with which a search is undertaken? The research methodology compared the outcomes of both a lab-IR study and a 'naturalistic' (i.e. on site) study to see if there were any significant differences between lab and on-site tests and is one of the first studies to undertake a rigorous assessment of the time spent on search using employees in their workplace.

The paper also presents the outcomes of an on-site study of engineers, who by the nature of their work are likely to be intensive and experienced users of search applications (A good example of professional searchers). This study was not about the absolute time spent on search but instead used work-task journals to assess the extent to which professional engineers needed to find information quickly or were prepared to spend time undertaking a more in-depth search.

The work-task journals set out seven categories of task

1. Ordinary administrative
2. Unordinary administrative
3. High-quality tasks
4. Just-to-get-done tasks
5. Regular teamwork
6. Unique tasks
7. Inventive teamwork

The effectiveness of information seeking and searching seemed to be the main goal of the activities in task 1-5, whereas aspects related to learning and considering new alternatives was related to the tasks 6 and 7. It seems that the tasks 6 and 7 are open for considering information as interesting, instead of easily found, necessary and relevant that are more readily related to tasks 1-5.

In 2009 a study was undertaken of a group of Microsoft employees working in a collaborative search situation. Although the results in the [paper](#) should not be extended too far in the direction of single-user enterprise search the outcomes are of interest in the context of the time spent searching.

Of the co-located collaborative searches that took place at work, most (52.4%) were informational (e.g., searches for work-related documentation such as network protocols or design principles) or navigational searches (38.1%) (e.g., navigating to the company's website). A few (9.5%) were transactional (e.g., applying for a parking permit). The majority were spontaneous (61.9%) and lasted only a few minutes. Searches that lasted for an hour (14.3%) were all informational in nature.

12. Query reformulation

Query reformulation is where a lack of information literacy can be a significant handicap in finding relevant information. Over ten years ago Helene Hembrooke, Laura Gran-ka, Geraldine Gay and Elizabeth Liddy categorised the ways in which both novice and expert search users [reformulated their initial query](#). In summary the strategies were

Elaboration Expanding and extending search statements using details and searchers' intrinsic ideas.

Redundancy An overall index of the extent to which search terms are used repeatedly on successive queries. Redundancy can be further divided to backtracking, topic terms, and plural making/taking.

Backtracking

The frequency with which a searcher reuses prior search terms over successive trials.

Topic Terms

The extent to which the given topic terms are used as main querying terms.

Plural Making/Taking

Reflects instances when a user repeatedly incorporates similar nouns into their search attempt, with the slight modification of making the word plural or singular.

Broadening The extent to which a user begins with a specific query and expands the scope of the search phrase over successive trials.

Refining The extent to which a subject begins broadly and narrows the search with increasing specificity.

Kitchen Sink The extent to which a searcher incorporates search terms related to the subject, but not specific to the query task.

Poke-N-Hope The extent to which a searcher retains the same basic structure throughout all search queries, changing only a single word within each trial.

As an information scientist I have spent most of my career working on query reformulation skills and yet still run into problems in trying to optimise my Google experience. This is because I do not keep up with the changes that Google make in their search algorithms. Specialist search trainers, such as [Karen Blakeman](#), are finding a very strong demand for web search training and yet internally organisations assume that since the vendor has promised an intuitive interface any problems that arise are down to the incompetence of the user.

A novel approach to query formulation has been developed by Tony Russell-Rose and Phil Gooch. Designated [2d Search](#), query formulation is managed by expressing concepts as objects on a two-dimensional canvas. Relationships between objects are articulated by manipulating them using drag and drop. Automated search term suggestions are provided using a combination of knowledge-based and statistical natural language processing techniques. This approach has the potential to eliminate many sources of inefficiency, make the query semantics more transparent, and offers further opportunities for query refinement and optimisation.

In effect query expansion is treated as a recommendation task rather than an information retrieval task, i.e. given one or more query terms already entered by the user, can we provide a list of further recommended terms. Reframing the task in this way is particularly significant, since the visual approach offers an opportunity for the user to engage meaningfully with candidate expansion terms and exercise more informed judgement regarding their value and contribution to the current search strategy.

As mentioned in Section 5 query reformulation requires good linguistic skills, including knowledge of synonyms and professional jargon. Most search applications provide ways to support query formulation, such as auto suggestion and auto completion. How effective this functionality is in enterprise search has not been evaluated. With the low levels of search use, the very wide range of search terms and often multiple languages there may not be enough data to drive effective automated query suggestion.

13. The relevance of relevance

It is time to consider the issue of relevance. Probably several thousand research papers and thousands of hours of enterprise search tuning have been dedicated to ensuring that the 'most relevant' information appears in the first 10 results, and ideally in the first 5. The level of interest in relevance tuning is illustrated by the success of the [Haystack](#) conferences and the recognition of the potential value of [relevance engineers](#) in search teams.

When it comes to enterprise search the concept of relevance and the value of relevance tuning needs to take in account a range of related issues. Enterprise search users balance three criteria when assessing search results. The first is the extent to which the document is nominally relevant to their query. I use the term 'nominally' because this is the extent to which the retrieval model has determined the relevance based (for example) on the BM25F algorithm's assessment of the text content of the document.

The second criteria is that of quality. This will be a very personal view, based on a knowledge of the organisation and the people within it. Trust is an element of this quality assessment and a user is more likely to select a document that has been written by someone they know and trust than a document from an unknown author, or an author who (according to the people search functionality) is no longer an employee. Currency is also important.

The third criteria is usability, in the context of (for example) whether the format of the document is one that is relevant to a need or is in a language that the user can understand. A PowerPoint file may well be highlighted as a highly relevant document (especially in a SharePoint search!) but the user is faced with being unsure of the back story to the presentation and whether (for example) the presentation was heavily criticised when presented to a team meeting.

Although improving the ranking of relevance documents is essential the research discussed in Section 3 indicates that 38% of the dissatisfaction factors were categorized under the Technology heading and of these 32% were categorized under 'ranking'. My perspective is that although improving ranking can improve search satisfaction it is just one of many other factors. Search teams need to be certain from user feedback on the extent to which poor relevance ranking is impacting search satisfaction when placed against other factors. It is about prioritizing the workload on what is almost always an under-resourced search team. It is also important to take into consideration the relevance engineering challenges. These may be easier to address in an open source application than in a commercial search application where AI and ML routines may not be well documented and may not be able to be validated on an A/B basis.

14. Document snippets

Over the years there has been an enormous amount of research into the optimisation of search interfaces, with a great deal of attention being paid to facet design and layout and the positioning of the search box. These provide dialogue support in the quest to arrive at a ranked list of results on what is often termed the Search Results Presentation page (SERP). Once the list of results has been presented the task of the user is to review them and decide which might be the most relevant for their purpose. This involves a very close scrutiny of the summary text for the result. A document that may be tens of pages long is represented by a document surrogate of perhaps 50 words together with some metadata. Requests For Proposals (RFP) search applications rarely cover functionality for snippet derivation, and technical documentation provided by a search vendor the issues of document summarisation are not considered. How else is the user going to gauge the potential relevance of a result?

The challenge of creating document summaries has been on the search technology agenda since the pioneering work of [H.P Luhn](#) in 1958. There are two core approaches to creating document summaries. [Extractive summarisation](#) locates and presents sections of sentences from the document. First an intermediate representation is created, the sentences in the document are then scored against the query terms and finally a summary is presented that may include sections of more than one sentence, usually with the query terms highlighted. The fundamental challenge is whether the combination of the sentence extracts and the highlighting of the query terms will give the search user a valid representation of the value of the document. This is especially the case when query terms are in tables and charts, which are very difficult to deconstruct through extractive summarisation.

The second approach is abstractive summarisation in which natural language processing is used to create (in effect) a new document which is a summary that in principle would be a best match for the query terms. This is technically very difficult but over the last few years advances in NLP and machine learning are giving a fresh impetus to providing solutions. There are many [extractive summarisation software applications](#) available as well as methods such as [ROUGE](#) which can be used to assess the quality of a summarisation technique.

Two recent research papers shed new light on important aspects of document snippets and SERPs in general. Leif Azzopardi, Paul Thomas and Nick Craswell have looked at the [utility of SERPs](#) and the study by David Maxwell, Leif Azzopardi and Yashar Moshegi considered [snippet length and informativeness](#). The paper on the utility of SERPs is focused on web search but the paper on snippet length has more general applicability. Both are rich in experimental data. For the purposes of this report I am going to present my own interpretation of the research outcomes.

In an enterprise search situation relevance is obviously important but equally so is the quality of each result. This cannot be directly assessed but it is likely that the user will be considering a range of clues from the result snippet.

These might include

- The quality (especially clarity) of the title
- The name of the author
- Their position in the organisation
- The department for which they were working
- The origination date of the document
- The file format if it might have an impact on usability
- The language

In an enterprise context the origination date is of considerable importance. In effect this is the date on the front page of the document. (I am using 'document' in a generic capacity). The problem with the very common Last Modified Date filter is that there is no ability to evaluate what was modified. If it was a substantial modification then in an enterprise situation it would have been republished in a revised version. It could be anything from a change to the issuing department title or to change a sales forecast by an order of magnitude.

The quality of the snippet information is totally dependent on the extent to which there is consistent governance of metadata. The presentation of file format, document type and last modified date in Microsoft applications is largely because this can all be inferred from the metadata associated with document preparation in SharePoint.

There is a very interesting perspective on this issue from Frances Johnson at Manchester Metropolitan University in a [paper](#) entitled From Accessibility to Assess-Ability: An Evaluation Heuristic Based on Cognitive Engagement in Search.

To quote from the paper

The aim of this investigation is to work towards developing a framework for the heuristic based evaluation of the interface designed to optimize users' core activities in search and discovery. This seeks to complement the use of existing and well known heuristics in interface design and metrics to gauge user satisfaction and engagement. The framework proposed is based on key questions the user might ask when searching – such as 'has the retrieval engine worked?', 'is the information retrieved relevant?', 'is the information credible?', 'has my query worked?', 'what have I learnt?' 'what else do I need to know?'; and overall, to evaluate the assess-ability of the information retrieved and presented at the interface. This critical approach which leads the user to actively search for information, with intention, assessing both the information retrieved and the interface supporting this activity is arguably a vital literacy necessary for instances of search where the user (with the information need) is bought back into decision making, and for a better experience of searching with current search technologies.

This paper is important as it reinforces the complexity of the search experience and strongly suggests that assuming that enterprise searchers only take into account the relevance rank of a result is far too simplistic. It seems likely that search users have developed a heuristic for evaluating SERPs but are not explicitly aware that they have done so. Of more importance is whether they have a single heuristic which may not be relevant for all types of search or have developed a range of heuristics.

15. The scent of a SERP

Peter Pirolli and Stuart Card developed the [information foraging model](#) for information seeking while working at Xerox PARC in the early 1990s. [Ed Chi](#), a fellow Xerox PARC employee, further developed the model in the late 1990s. The concept of an “information scent” refers to the way (for example) pigs can find truffles even though they are well hidden.

The connection between truffle hunting and Search Engine Results Pages (SERPs) is that the search user’s view of results pages is informed by a wide range of proximal clues, which together create an information scent in the mind of the searcher. For example, a glance at ten PowerPoint files listed on the first page of results could bring a search to an abrupt halt before it has even started.

A great deal of research has been published in papers and several books on search user interface design but none of this research considers the role of the SERP in enterprise search. What are the proximal clues that employees use to assess relevance. It is likely that one of the clues will be the person writing the document or giving the presentation. This may be a ‘respect’ judgement but could also be an ‘access’ issue. My experience in enterprise searching suggests that the document is used as a proxy to locate either a person who can then be contacted or perhaps a department. This will always be a subjective assessment, based on the degree of distance the author seems to be from the search user’s network. This is of course supposition on my part but is derived from a range of usability tests and comments made during interviews.

[David Maxwell](#), a PhD student in computer science at the University of Glasgow and [Leif Azzopardi](#), associate professor at University of Strathclyde, presented a [paper](#) at the 40th European Conference on Information Retrieval in March 2018 which takes search relevance in a new direction. In their paper the authors hypothesise, model and then validate the impact the information scent of a SERP has on stopping strategies and therefore, search performance. In summary (and there is a substantial amount of data and analysis in the paper), they believe the role the quality of SERP presentation has had on search effectiveness and satisfaction has been significantly underestimated.

The paper goes on to discuss the search ability of users. Again, in the “click count” world, all users are assumed to have equal search proficiency and an equal command of the languages being used on the SERP. The paper shows search proficiency influences opinions about the usefulness of the page based on information clues from SERP, and the authors set out some potential categories of user proficiency. Recognising that a SERP itself may be a stopping point is one of the key elements of the Complex Searcher Model set out in Section 7.

In the concluding section of the paper the authors comment

Furthermore, we found that including the SERP level decision pointed to more accurate modelling of actual stopping behaviour. This represents a major shift in modelling interaction and has ramifications for how IR systems are measured, which typically assumes people examine ranked lists. These results suggest that future work needs to be directed towards measures that consider abandonment and should also include how the sequence and quality of queries affects interactions taking place with ranked lists.

It is important to stress that these new insights into how users gain clues from SERPs

does not invalidate existing retrieval metrics but instead provide additional information to assist in enhancing the user experience and ultimately their level of satisfaction with the search application.

16. Stopping strategies

During the search process, searchers need to decide when they should abandon the current query (and perhaps issue a new query after examining the current results list), and when to curtail their search by stopping the search session altogether. Knowing when to stop is considered a fundamental aspect of human behaviour. Stop too early, and important information may be missed. Stop too late, and time and effort is wasted. Worse still, the examination of fruitless result lists will mean not having time to examine other lists which may potentially contain greater yields for the searcher. David Maxwell and Leif Azzopardi together with Kalervo Järvelin and Heikki Keskustalo (School of Information Sciences, University of Tampere, Finland) have published a fascinating [research paper](#) on this topic. The four authors are in the very top echelon of IR research so what they have to say should be taken very seriously.

Two of the earliest stopping rules proposed were devised by W.S.Cooper in 1973 (search goes back a long way!) who proposed:

- the frustration point rule, where a searcher stops after examining a certain number of non-relevant documents; and
- the satisfaction stopping rule, where searchers would stop only when a certain number of relevant documents were found.

The frustration point rule is especially interesting. The authors define it as counting the number of non-relevant documents seen in the ranked list at position k . If the total number of non-relevant documents exceeded a given threshold, the searcher would then stop. So if we have a personal rule that if we have got to the third page of results (say $k = 30$) and found few if any relevant results then we give up in frustration. Our time is too precious and we may, or may not, start again.

Numbering the SERPs enables users to jump to say page 5 and take a view on how steep the relevance decay curve seems to be. If page 5 is displaying a significant number of relevant results then the user may well revise the query to something more specific than work through 50 or more documents.

I will probably do the authors a great disservice by jumping to the end of their paper but the main outcomes are that the two most common stopping strategies are

- Fixed Depth. Under this stopping strategy, the simulated searcher will stop once they have observed a self-defined number of results snippets, regardless of their relevance to the given topic.
- Contiguous Non-Relevant. The searcher will stop once they have observed (say) five non-relevant snippets in a row (contiguously).

The authors caution that a great deal more work needs to be done to understand these behaviours. However the research indicates that we may need to rethink our approaches to evaluating search success and search failure, at least taking into account search users strategies which may be internalised and pragmatic rather than just a function of relevance.

17. The implications of information seeking failure

There are many differences between searching the public web and searching enterprise repositories. People posting content on web sites do so because they want others to find this content and make use of it. That is not a primary driver within the enterprise and so little attention is paid to adding in good quality metadata and perhaps providing additional context so that an employee working in a different department, location or subsidiary is able to judge the value of the content.

Of greater importance is that enterprise search has many modes of failure. These include

- The document does not exist
- The document may have existed and although it has been deleted the record remains in the index
- The document may be in an application which has not been indexed
- The document may not yet have been indexed
- The user does not have the security permission to access the document
- The document is written in a language that is not fully supported by the search application
- The document has been indexed but is not able to be downloaded because the document itself has been deleted but remains in the index
- The document has a low nominal relevance (based on the ranking engine) and the search has been halted before the document is presented in the search results (The term 'document' is used as a generic descriptor to cover an individual file in any format.)

The search user is unlikely to be able to determine the reason for the search failure but whatever it is there is no immediate fix. This puts the user in a very difficult position. In a search of the web almost any item of information being sought has multiple sources, but that is not the case internally. The only enterprise option is to find someone who may know the information or has a copy of a relevant document, and neither of these are easy to undertake. The search journey ends with a failure that could have organisational and personal implications.

Sam Marshall (Clearbox Consulting) has developed an elegant framework to support organisations in understanding the potential [reasons for search failure](#). The categories of issues that can lead to potential search dissatisfaction based on a wide range of client engagements undertaken by Clearbox Consulting are

- Content quality issues
- Technical issues
- IA (UI) issues
- User training issues
- Other issues

These match and so confirm the outcomes of the research conducted by Paul Cleverley and Simon Burnett referred to in Section 3.

18. The value of search logs

My experience in talking to search managers about information gained from search logs can be summarised as

- They are very unwilling to disclose publicly (in a presentation) any information about the volume of searches, queries or unique users
- Very little analysis is carried out on search logs. The primary reporting of log outcomes is usually up the line to a more senior manager as a means of continuing to justify the search investment

In my view the primary reason why log disclosure is so rare is that search managers are at a loss to understand why so few searches are being carried out and don't want to present the numbers in case the audience considers the search application to be a failure!

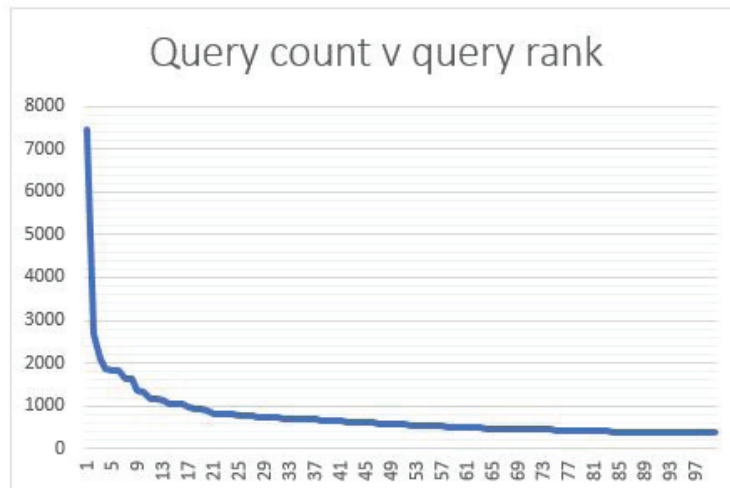
The table below shows data from three quite different multinational companies.

	High tech manufacturing (IF client)	Oil and gas (Cleverley and Burnett 2018)	Pharmaceutical (IF client)
Employees	87000	110,000	100,00
Total queries a month	25%	14%	5%
Number of unique users	73000	70000	62000
Queries per unique user	2.2	6.4	2.3

All three have a well-developed and supported enterprise search application. What is immediately apparent is the small number of queries across all unique users. This suggests that search is not widely used even though all three companies are in highly competitive global business sectors.

Another feature of enterprise search applications is the extremely long tail of low-posted queries, and the graph below, taking from the logs of the high technology manufacturing company illustrates this very well.

With only a few exceptions all these queries are about applications, of which the company has over 500 active enterprise-wide. The most highly queried term is the employee self-service application. Although it is linked on the corporate intranet most employees



just type in ESS and are led immediately to the application which (unlike the link on the intranet) leads them through single sign-on to their personal area of the application. This outcome is supported by the work referred to earlier by the [University of Bristol](#).

This might be an extreme case in terms of the rapid fall-off of query volumes but is indicative of the general situation.

From employee research the reason for the low volume of queries is that this particular company has a long tradition of team working. Each team builds the information resources it needs, and these are topped up from time to time by team members contributing reports and other resources they find, many of which come from being a member of other teams. The role of Professional Support Lawyers in law firms has been mentioned earlier in this report and similar roles can be found in many organisations.

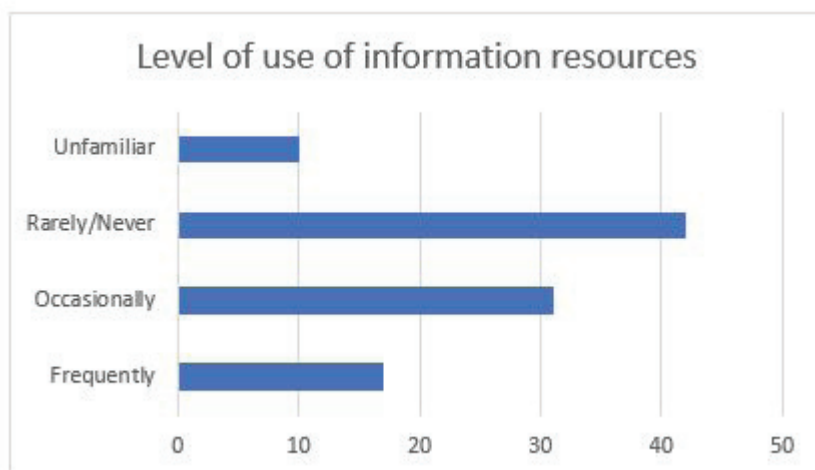
There is another implication in this use of a wide range of applications. Many employees may spend a significant amount of time in a specific application. Good examples would be an enterprise resource planning system or a project management system. As they work in these applications they may well wish to find information on a topic that could be within the application, or (such as a corporate policy document) may be in another application. Most enterprise search applications do not support users undertaking a search from within their 'operational' application. This is a use case that is rarely identified and even more rarely supported. It is of interest that this is a feature of the recently announced Google Enterprise Cloud Search.

In the case of the company referred to above there was very little periodicity in the query ranks. In the case of universities, I have seen significant differences across the academic year, and in the case of large engineering consultancies bids and project start-ups invariably distort the inflexion point and the shape of the long tail.

Lee Romero, Senior Manager, Knowledge Services for Deloitte Touche Tohmatsu Limited has posted a number of helpful [blog](#) posts on some of the issues around search query log analysis. Although these were published in 2009-2011 they remain valid today. The same is true of [Search Analytics for your Site](#), published by Lou Rosenfeld in 2011. The focus of this book is on web site statistics but most of the principles apply to any search application.

19. Search literacy

An implication of this log data is that the view that a wide number of employees are (or should) be using enterprise search on a frequent basis is not supported by evidence. An outcome of the research carried out in a global law firm (shown in Section 4) is that only 17% of employees in the law firm used information seeking and searching applications on a regular basis, and just over 30% used them on an occasional basis.



The fact that over 50% employees seem to make little use of these resources should not be taken to indicate that they work in an information desert. As with so many organisations information is shared and discussed in practice and matter teams where personal expertise is added to baseline information.

Even so as the chart indicates there are a core of regular users that are generating the use rather than every employee of what would certainly be a very information-dependant organisation.

An interesting outcome of the study by Paul Cleverley and Simon Burnett was the impact of search literacy on search satisfaction. Given that the indications are that individual usage of search is low and that exploratory search is important, search literacy is an important skill to acquire. If the number of searches is low and information being sought is wide ranging then almost every enterprise search is a new experience. The situation is exacerbated by the language diversity referred to above.

Complex user interfaces present considerable challenges to users. The typical approach is to have 'one UI to serve them all' and as the number of applications increases federated search UIs can quickly become a challenge even to frequent users. The example shown below is quite typical, with an emphasis on filters based on metadata. For example, 'Result type' offers a search user the ability to just view documents in a Rich Text Format. It is very difficult to imagine the utility of this filter.

Modified Date	Document Type	Result Type	Site
Any Modified Date	Any Document Typ...	Any Result Type	Any Site
Past 24 Hours (1)	Updates (12)	Portable Docum... (842)	BusinessDev (916)
Past Week (7)	Templates (11)	Microsoft Word (363)	Practices (402)
Past Month (32)	Internal Communi... (8)	Microsoft Powe... (297)	Services (250)
Past Six Months (181)	Human Resources (6)	Email Message (30)	Offices (44)
Past Year (340)	Legal Opinions (4)	Web Page (30)	People (19)
Earlier (1452)	Policies (4)	Microsoft Excel (21)	
	Business Forms (3)	Web Archive (21)	
	Client Training (2)	Rich Text (4)	
	Credentials (2)	ZIP Archive (2)	
	Guides (2)		

The complexity of search user interfaces for enterprise search has not been the subject of any specific research projects. However some indication of the complex can be gained from a paper by Ben Steichen and Luanne Freund in which they undertake user tests on a number of different ways of presenting [results in multiple languages](#).

The authors comment

Using the most common designs in both multilingual information retrieval and monolingual aggregate search, our crowd-based evaluation has shown that participants particularly like an interface that presents results from different languages in separate 'Panels' on the same screen, and that language separation is generally preferred over the common approach of interleaving results. Also, our results have shown that the user's proficiency and search content type play a role in user preferences, and that different interfaces elicit different user behaviours.

In the case of public web search a great deal of research and experimentation on user interface design has been carried out, but in the enterprise space the view is that search should be intuitive. Yiwei Wang and Chirag Shah have conducted [research](#) on the impact of various search factors on search failure.

Among the main internal issues were

- Unable to articulate information needs
- Unaware of the relevant information sources
- Poor search skills

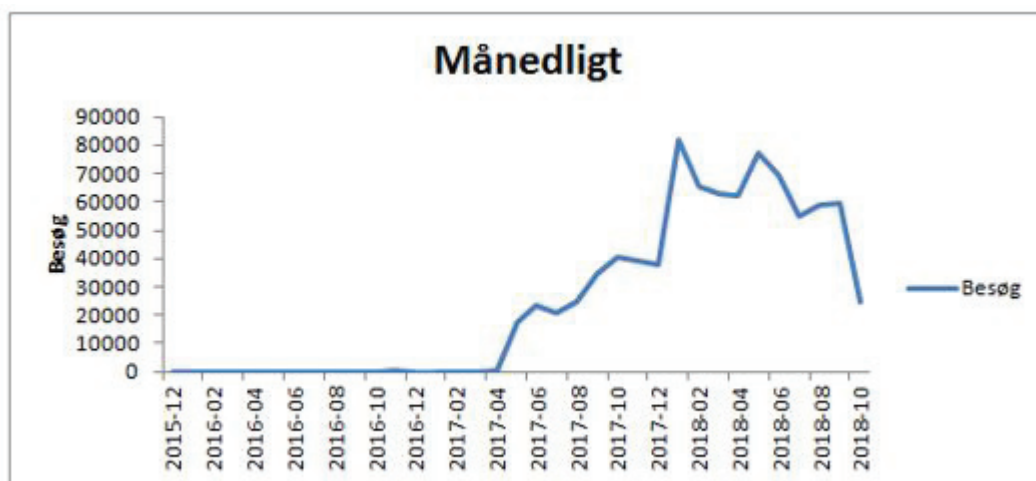
The impact of training on enterprise search literacy has not been considered in any detail. There is [one paper](#) that reports on the use of library research applications, which do have complex user interfaces. Using a quasi-experimental design over a year and a half, eight doctoral students (novice searchers) participated in a series of five sessions with an expert searcher who was an information professional. A novice-expert comparison examined the differences between novices and experts in information searching; and the effect of scaffolding sessions in which the expert information searcher helped novice information searchers was examined. Findings showed differences existed between the novice and the expert searchers in use of complex formulation of query statements, choice of keywords, and operators. Scaffolding sessions with the expert searcher resulted in self-reported and observable improvement in information searching among the novice searchers. The research also considered the extent to which various approaches to changing query formulation were used by novice and expert users.

The priority given to search by the Danish Police force was mentioned in Section 5 In addition to providing high-quality metadata attention has also been given to providing guidance to users. To quote from the [StepTwo report](#) that describes the work of the winning intranets

The team has produced a number of one or two minute videos to educate users on how to do more precise searching, for example using the metadata terms directly in the search query. So as not to confuse users, the videos don't mention metadata but instead talk about the document model which the user can relate to through their daily work.

The team also spend a few hours every week improving the governance around search by reading through default search reports. However, instead of publishing more promoted results, which the team try to minimise the use of, they use the reports and any trends they spot to meet with authors in the police districts and talk about best practices around tagging content. While search logs can be informative, the team believe the results cannot tell you about the search experience. Therefore the team also regularly speak to content owners and users to find out about any issues. If there are dissatisfied users, the team might use workplace observation to gain valuable insight and to allow for the development of the best solution.

When the new intranet and search application were launched in mid-2017 the team was concerned about the low level of use. They decided to place an A4 guidance poster alongside the coffee machines in the offices and as the chart below shows there was an immediate and significant increase in the volume of searches. The dips are due to popular vacation periods. The drop off in 2018-10 is because the chart was produced early in the month in order for the chart to be included in this report. The Danish Police has 14,000 employees generating around 60,000 searches a month, a very high level of use and a testament to the amount of effort that was put in to the implementation of the search application.



© 2018 Rigspolitiet, København

This is an excellent example of the benefits of even a very limited amount of user guidance and encouragement.

20. Personalised enterprise search

Many search vendors are promoting the idea of providing a personalised search experience so that the employee only sees a small number of results of very high relevance to them. Technically this involves indexing applications being used by the employee in at least near real-time and using a range of parameters to filter the search results before presentation. In addition, this information can be used to enhance expertise profiles. The approach is not new. Profiling was a feature of Autonomy IDOL in 2005. The approach is now being positioned as a way of using AI and machine learning to ‘understand’ the interests of the employee. Gartner refers to these applications as Cognitive Search and Forrester as Insight Engines.

Autonomy was offering a profiling functionality in Version 5 of the IDOL application in 2005. This functionality was then used in a bar at the top of a page of text that would display terms as they were being keyboarded so that the user could click on the term and convert it to a search query. Users found this to be distracting and have very limited utility and it was quickly discontinued.

Although AI could be of value in determining relevance from document tracking it can be undertaken in other ways. In 2011 Matthias Reichhold, Jörg Kerschbaumer and Günther Fliedl reported on how enterprise search results could be enhanced by using a [role-sensitive ranking algorithm](#).

In the information retrieval community, the term applied to personalised optimisation technology is Contextual Search. The 1st International Conference on Information Interaction in Context (IliX) was held in Copenhagen in 2006, with further events in 2008, 2010, 2012 and 2014. It was then merged with the ACM Human Computer Information Retrieval (HCIR) symposium to create the [CHIIR](#) conference which focuses on elements such as human involvement in search activities, and information seeking and use. The papers at these conferences not only set out how contextual search can be delivered technically but what the user experience needs to be to trust the profiling and take advantage of the technology. A [review](#) by Lynda Tamine and Mariam Daoud was published in 2018 of the experience gained over the last two decades, citing over 160 research papers.

In the context of this report the primary issue to consider is whether personalised search is going to enhance search satisfaction. In one sense enterprise search has always had an element of personalisation through security trimming but an individual employee has no way of knowing what they are not allowed to find through a search query.

We are now well used to the fact that Google, Bing and other search sites make use of a range of associated data on our searching habits, location and much more to target the results we receive from a search. Amazon was probably the first to do it at scale, though having selected a book I find I’m more interested in the people who purchased books seemingly totally different from my own selection. In the case of these public search sites there are two important issues to consider. The first is that if you don’t like the Google search bias you can run the search in Bing or DuckDuckGo or any number of other sites. The second is that you are probably not going to make a life-changing decision on the basis of a single search in a single search application.

Inside the enterprise things are totally different. The first is that there is (in general) only one search application across core business-critical repositories. The second is that if you make a decision based on a “personalised search experience” and that decision is

not the correct decision it could be career-changing in all the wrong ways. The issue that is never discussed is whether limiting the scope of the search by incorporating factors which are invisible to the employee (and that they cannot verify or change) is ethical. It is presenting a 'truth' without there being any way of verifying the truth. In the public space we have options, not just from using different search sites but interrogating individual web sites. Take weather information. In the winter of 2018 the BBC Weather Website stated that there was heavy snow in Philadelphia. On the freeway cameras there is no snow in sight and flights were taking off on time from Philadelphia airport.

Back inside your company office in London you might be looking for sales by customer in Saudi Arabia. The search application provides a personalised table but it is not clear if the numbers are in Riyals, £ Sterling, or \$US, because you work for a multinational organisation and you don't know where the sales are consolidated. What you also don't know is whether you are seeing the full picture because your security clearance or office location might mean that sales to Saudi government agencies are not included.

The approach might have some degree of value if employees just used (for example) SharePoint as a means of creating text documentation. In reality employees are using a wide range of applications. Indexing all of these involves the complexity of connectors and the licence and per-item costs of indexing.

Much is also made of running analytics across query logs to identify patterns. However, there is just not enough data on the vast majority of queries to provide a 'better quality' result than an investment in metadata and taxonomy management applications and good natural language processing applications. Named entity extraction in particular can make a significant difference to search effectiveness.

Are employees aware of the extent to which search has been personalised? Who is making the judgements on the filters, and is there a way in which an individual can verify their personal filter profile, and if needed challenge it within the compliance structure of the organisation? What if the technology does not work correctly – will the vendor accept a case for damages if the organisation makes an inadvertently uninformed decision? Is there an entry on the corporate risk register about the potential outcomes of inappropriate (to the occasion and decision) personalisation?

Another issue is personal data privacy. Does my profile contain information derived from indexing which might inadvertently prejudice my career? The review by Lynda Tamine and Mariam Daoud referred to above considers privacy issues in some detail. In effect personalisation is a form of profiling and any introduction of personalised search needs to take account of the ['Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679'](#) prepared by the Article 29 Data Protection Working Party of the European Commission. So far no search vendor has published the research they use to validate the success (in terms of search satisfaction) of personalised search on an enterprise-wide/multiple-application basis.

21. Considering the evidence

The outcomes from evidence presented in this report are set out below to provide an extended Executive Summary to the report.

1. Searching for information is just one way in which people can find the information they need to make decisions which are very likely to impact the organisation, colleagues and their personal career journey. It is important to understand the context of why they are seeking information, why they have chosen to use a search application, and what their expectations are of the outcome.

2. Enterprise search applications first came onto the market in the late 1980s. The BM25 ranking model was formulated in the early 1990s and forms the basis for most commercial and open source applications. Enterprise search capabilities are a core functionality of Microsoft SharePoint and Office365 applications and there are many commercial and open-source search vendors. By any measure enterprise search is a mature technology and business.

3. Despite this technical maturity the evidence from many surveys is that users find these challenging to use. Case studies from vendors provide virtually no information on quantitative metrics about the volume of queries and the number of unique users. These metrics are also rarely mentioned in conference presentations. If there were marked improvements over the legacy application it is curious that this information is not highlighted.

4. There are many case studies of intranet implementations, notably those published by Nielsen Norman and Step Two, and many conferences where the level of success of intranet implementations are presented in some detail. There has been a substantial amount of academic research into ERP and social network adoption and success factors, but virtually no research into the factors affecting enterprise search adoption. However over the last few years a number of research papers have been published on enterprise search, and on aspects of search use that, although not specifically about enterprise search, do provide insights into enterprise search behaviour.

5. The picture of a lone employee spending many hours a week searching for information is not based on reality. Enterprise applications are presenting employees with streams of internal data, and the intranet should be an effective source of internal documents. Indeed the problem is one of information overload, compounded by search applications which do not seem to be delivering the quality of results expected by users. The extent to which organisations work collaboratively has increased dramatically over the last few years and the process of working together and sharing information has inevitably changed the requirements for search. Interestingly no search vendor offers a collaborative search option.

6. Investing in more technology is not in itself going to result in an overall improvement in search satisfaction. Ensuring that there is consistent and appropriate metadata is essential. At present there is no independent evidence that AI and machine learning are going to eliminate the metadata requirements in enterprise search in the near term, let alone the short term, nor provide the level of work and personal context to provide users with personalised results they can trust.

7. Despite the wealth of click-derived data generated by search applications it has always been difficult to understand why people search and how much time is taken to complete a search session. Computational ethnography offers a methodology for gaining a better understanding of search processes at the level of an individual user. The initial application of this methodology has already shed significant light on these processes, especially the relationship between work tasks and search tasks.

8. Each work task is likely to give rise to many search tasks. Some of these may be almost subliminal in duration and others may extend over several related search sessions and take several hours or more to complete. Reducing the time to complete a search cannot be the basis for investment in technology.

9. Typically, employees will make use of many enterprise applications. They will take advantage of the highly curated content on the corporate web site and the World Wide Web and specialised subscription-based information services. The extent of this use will not be visible to the enterprise search team but could well influence the overall level of search satisfaction.

10. Providing a single user experience for all employees through a UI with a very wide range of filters and facets is not a realistic option. One of the important research outcomes is that search literacy is very low and presenting a single 'intuitive' search interface for employees is not realistic. Users of enterprise search applications should have a level of training and support that users of other enterprise applications are offered. Why should search be different in this respect?

11. The challenges of query management are significant, which is why professional searchers still make use of Boolean query strings. Query development requires a sound knowledge of the topic to be aware of synonyms and specialised vocabulary. The default language support will always be English but in many multi-national organisations a minority will have English as their first language. They will have to construct queries to be matched against largely English language content without perhaps an adequate command of the language.

12. The process of working through the results presented by the search application is much more complex than looking through the search results one-by-one. Ensuring that search results are relevant on inspection has to remain a priority, but it also has to be appreciated that people may scan one or more SERPs without clicking on a single link and then decide to modify their query. It would not be correct to see this as a failed or abandoned search, just a realisation that there might usefully be change in direction based on a rapid scan of complete SERPs. This is of course easier for experienced users.

13. Attention is now being focused on stopping strategies. It seems that users are developing their own personal heuristics on when to stop a search and either revise the query or adopt another information seeking approach. In an enterprise situation these heuristics are not necessarily based on relevance but might take into account who authored the document and which language it is written in.

14. Relevance measures and search metrics that have been developed for web and e-content search cannot be relied upon to provide a basis for enterprise search performance evaluation and enhancement. Changes made in response to these metrics should result in an improvement in search satisfaction. A measurement of 'search satisfaction' does not replace these metrics but complements them, providing indications of where more detailed assessments should be undertaken.

15. A major challenge for users is that there are multiple points of failure in a search application. Most enterprise applications support workflows, and it is usually very apparent when the application fails. With enterprise search the user has no option but to place 100% trust in the application. If the search application does fail this may not be apparent at the time that the search is being undertaken. It could well be that low levels of satisfaction with search are at least partially due to personal embarrassment and inconvenience when it becomes apparent in a team meeting that they are basing their contribution on a document which is not of the quality anticipated.

16. The overall picture that emerges of enterprise search practices from the evidence is one of immense complexity which cannot be solved by technology alone. The fundamental problem is that enterprise search is implemented on the basis that it can be used (in principle) by any employee without the need for training and support. The reality is that very few employees have the expertise and experience to construct effective queries and to assess the value of the results that are delivered. Each employee has their own domain knowledge and expectations, and has multiple information seeking options of which search is just one of many.

17. The assumption that seems to be made by search vendors and by IT managers is that anecdotal information about search use from a small number of interviews can be scaled up to all employees. No evidence is presented to justify this approach. The lack of claims about the high levels of search use and search satisfaction would indicate that there are fundamental challenges in enterprise search implementation which are not being addressed, and that the outcome is a low level of search satisfaction that could hold short term and long-term risks to the attainment of business objectives.

18. In my opinion the most significant cause of poor search performance is a lack of investment in search team skills to address the issues of information quality and information literacy highlighted in this report. A search team should have the skills and resources to understand user requirements and frustrations, and from business knowledge be able to draw inferences from search logs, no matter how many languages are being indexed. The team should also have the technical skills to be able to optimize relevancy ranking and system performance. There is no magic formula that (for example) calculates search team numbers as a function of employees. After all it could be that only 20% of employees depend on the search application. First understand the scale of the tasks the team will have to undertake and then staff up to ensure that search satisfaction steadily increases.

22. Measuring search satisfaction

Four ways of measuring search satisfaction are set out below. All have benefits and challenges that need to be recognised.

Search UI icon

The search UI has icons representing a range of satisfaction levels, or a drop-down list of degrees of satisfaction

Pros

- The score will be given at the close of a search session and so provide immediate feedback
- By definition it will be completed by people who are search users
- If the information is available it could be tied into other factors such as location and business department

Cons

- For short searches there is a danger that people may not be bothered to respond
- It will generate a substantial amount of data that needs processing and analysis
- The end of a search session is always difficult to determine.

Survey

A web-based survey is undertaken, either on an enterprise-wide basis or to specific departments and roles

Pros

- This survey can ask a range of questions about information seeking and information searching
- It is easy to target locations and departments, but usually roles are not defined well enough

Cons

- Organisations dislike surveys
- Employees dislike taking part unless there is a significant personal benefit
- It is difficult to persuade employees to give a fairly prompt feedback

Dipstick

As an example, place some bins outside of the staff restaurant and ask people to take a card pre-printed with a satisfaction level (1-5) and ask them to drop it into a bin

Pros

- Makes search very visible and could start some interesting conversations
- Very quick to implement and analyse

Cons

- May not be a suitable location, or locations in different locations are not consistent in footfall
- Probably can only be done once a year at most
- Very difficult to obtain user segmentation information

Meetings

Ask team leaders to add an item to a team meeting agenda about search satisfaction

Pros

- Very targeted survey methodology

Cons

- Team members may not be willing to go on the record with their comments
- Difficult to gain a prompt response as the item might be dropped from the agenda at short notice and then not reinstated
- Seen as an opportunity to tell stories about a specific search and not provide an overall assessment

With any of these options there should be a pilot trial, and this should include enough analysis of the results to estimate the scale of the analysis with a larger sample of respondents. There is no point in trying to get to a level of four significant figures. Just getting to two is quite good enough, especially for the initial survey.

23. Next steps

This report sets out the evidence that achieving high levels of enterprise search satisfaction requires attention to be paid to technology, information quality and information literacy factors. Just enhancing the technology is not in itself a solution. In addition, search has to be placed in the context of information seeking to gain an understanding of why people are using search. Tracking search queries through log information provides some insights into how people search and what content they judge to be relevant. We are still a long way from understanding why people search. The initial outcomes of research using computational ethnography is already indicating that enterprise seeking is much more complex than it might seem on the surface. A good example is the way in which SERPs are being used to provide an initial sense of the outcomes of a search before starting to work through individual results.

It would be easy to list out the ways in which search satisfaction could be increased, but in reality it comes down to investing in a search team that understands search technology and the organisation. This search team must be visible, providing training and mentoring and having the time to sit down with people and listen to their aspirations, their search success stories and the frustrations they have in both seeking and searching. A core skill in the team is that of relevance engineering. Although search is often referred to as an enterprise solution in reality there are going to be groups of users with specific requirements that will be met by the development of equally specific ranking models and user interfaces.

Search satisfaction is of course a subjective metric, but mapping variations in search satisfaction across the enterprise should provide a starting point for improvement. Search is a journey, so we must start out being as certain as possible that we are heading in the right direction.