A History of Enterprise Search 1948-2020

Martin White

Intranet Focus Ltd.

May 2020

About the author

Martin White has gained international reputation over the last twenty years for his understanding of how to manage enterprise and intranet search applications for the benefit of the organisation. Many of his clients have been multi-national organisations with complex information management and information discovery challenges, He specialises in finding solutions for organisations working in multiple languages.

Martin has written four of the five books on enterprise search management, including Making Search Work in 2008 and the second edition of Enterprise Search in 2015. His report on Achieving Enterprise Search Satisfaction was published in 2018 and he is currently writing a book about the Enterprise Search Experience. He is the author of a monthly search column in CMSWire and is Editor of Informer, the newsletter of the British Computer Society Information Retrieval Specialist Group. Martin has been a Visiting Professor at the Information School, University of Sheffield, since 2002, specialising in information management and interactive information retrieval (IIR). He is a Fellow of the Royal Society of Chemistry, a Fellow of the British Computer Society and a member of the Association for Computer Machinery.

Introduction

In this report I have set out what is, inevitably, a personal and selective view of the evolution of enterprise search. One of the challenges in doing so is that there are a range of views about a definition of 'enterprise search'. For the purpose of this history I am using the definition of search software that is specifically designed for networked use within organisations. Some of the elements of this software were developed for other applications, notably for web search, but have been modified and enhanced for the specific requirements for scalability and extensibility of searching for a wide range of content items on internal servers.

The report is divided by decade and aims to introduce readers to the key initiatives, innovations and innovators that have shaped enterprise search.

Some readers may be astonished that this chronology of influential people and events starts over 70 years ago when post-WW2 scientists began to explore way to manage the ever-increasing and complex flows of scientific information.

The history features academics, researchers, entrepreneurs, and teams. Small and large businesses have made an impact and there are stories of mergers, buy-outs, IPOs and commercial failures and successes. There have been economic downturns and the dot.com bubble. Advances in computing capability and technology have greatly influenced the sector, of course, but so have changes in the way organisations and people work. Changes in the understanding of information needs, wants and behaviour of organisations and the individuals within them continue to influence the development of enterprise search tools and platforms.

Enterprise search applications now on the market have evolved over many decades. Search software vendors have been launched, acquired and disappeared but the fundamental technologies are probably the most robust and optimized of enterprise applications. Unlike other enterprise applications they do not mirror information workflows but react in around 500 milliseconds to queries from employees, each of whom has different expertise and different information requirements from those of any other employee. Each expects the search application to, in effect, read their mind and deliver a set of results which meets all their requirements. This is the story behind the achievements.

Intranet Focus Ltd

12 Allcard Close, Horsham, UK RH12 5AJ

http://intranetfocus.com/

1948 – 1959 In the beginning

The start date for this history is 1948, the year of the [Royal Society Scientific Information Conference](). This conference identified the challenges that lay ahead in managing the flow of scientific information, challenges that arguably we have not solved. The earliest research into how computers might help was undertaken by [Philip Bagley]() as part of a Masters project at MIT. His thesis was entitled [Electronic Digital Machines for High-Speed Information Searching](). He set out the basic principles of 'information searching' and wrote a program for the [Whirlwind computer]() at MIT. This Masters thesis should not be confused with his later [PhD thesis]() in which he was almost certainly the first to use the term 'metadata.

By June 1952 there was enough interest in the subject at a number of research centres across the USA to hold a Symposium for Machine Techniques for Information Selection at MIT. One of the speakers at the Symposium was [Hans-Peter Luhn](), at that time working on punched-card retrieval systems for IBM. Luhn would turn out to be hugely influential in information retrieval.

Another very influential person was to be [Eugene Garfield](), who in 1955 published a paper in Science about the value of citation analysis. From this approach Garfield launched his Institute for Scientific Information to commercialise citation analysis. His insight also became one of the innovations incorporated into Google at the outset on the 1990s, but that is another story. Of more immediate interest is a paper by Allen Kent and his colleagues at the Battelle Memorial Institute, Ohio. [In this paper]() the concepts of 'recall' and 'pertinency' are proposed as metrics for a search application; 'relevance' later replaced pertinency.

There were two further important conferences in the 1950s.The first was the [International Study Conference on Classification for Information Retrieval](), held in Dorking, UK in 1957. This was the first opportunity for UK and US research teams to exchange ideas and research on information retrieval. The USA may have had a technology lead, but the UK was held in high regard for research and implementation of classification and index frameworks.

A year later an [International Conference on Scientific Information]() was held in Washington D.C. to take note of developments since the 1948 Royal Society conference and much of the discussion was about information retrieval. The papers make for some fascinating reading. By 1958 Dow Chemicals was evaluating how computer-based systems could be used to manage in-house documentation.

The chemistry community have long had some special information retrieval challenges (such as searching chemical structures) and have always been in the vanguard of search development. It was at an American Chemical Society meeting in Miami in 1957 that Luhn gave a paper on [A Statistical Approach to Mechanized Encoding and Searching of Literary Information]() in which (in effect) he set out the constituent elements of a search application.

The following year Luhn published a [paper on his work at IBM](#) in which (according to the abstract)

 "Excerpts of technical papers and magazine articles that serve the purposes of conventional abstracts have been created entirely by automatic means. In the exploratory research described, the complete text of an article in machine-readable form is scanned by an IBM 704 data-processing machine and analyzed in accordance with a standard program. Statistical information derived from word frequency and distribution is used by the machine to compute a relative measure of significance, first for individual words and then for sentences. Sentences scoring highest in significance are extracted and printed out to become the "auto-abstract."

This was indeed a visionary approach. Luhn also proposed that the frequency of word occurrence in an article furnished a useful measurement of word significance. This is the origin of the now familiar term frequency – inverse document frequency model although it was not until 1972 that Karen Spärck-Jones developed a rigorous statistical basis for TF.IDF.

In 1959 Maron and Kuhns wrote a seminal paper entitled [On Relevance, Probabilistic Indexing and Information Retrieval](#) in which they defined 'relevance' and the use of 'probabilistic indexing' to allow a computing machine, given a request for information, to make a statistical inference and derive a number (which they called the "relevance number") for each document. They suggested that this could be a measure of the probability that the document will satisfy the given request. The result of a search would then be an ordered list of those documents which satisfy the request ranked according to their probable relevance.

The importance of the paper is that Maron and Kuhns then evaluated their proposal through a manual (rather than computer-based) trial, so setting out not only the fundamental principle of determining the probability that a document was relevant but the importance of system evaluation. Fifty years later Maron [published a short account](#) of the background to this paper in which he provides a fascinating insight in how he and Kuhns developed this principle.

Although Maron and Kuhns had shown that a probabilistic approach was superior to a Boolean approach virtually all of what might be seen as the first generation of commercial search applications used Boolean logic because the challenge of calculating a 'relevance number' had yet to be solved. It is of note that Maron was at the Rand Corporation which had set up [System Development Corporation](#) as a subsidiary. RAND spun off the group in 1957 as a non-profit organization that provided expertise for the United States military in the design, integration, and testing of large, complex, computer-controlled systems. SDC became a for-profit corporation in 1969, and began to offer its services to all organizations rather than only to the American military. played an important role in search development. Another important development in 1959 was the establishment of the [Augmentation Research Center](#) at Stanford Research Institute under the direction of [Doug Engelbart](#).

So by the end of the 1950s almost all the core elements were in place, including understanding the required modularity of the search process, the benefits of a probabilistic view of document retrieval (rather than using Boolean operators), the concepts of precision, recall and relevance, and the value of testing and evaluation. What was needed now was computing power. In the 1960s California would forces with Massachusetts in the quest to scale up search and make it widely available within and outside of the organisation.

1960-1969 The pioneers

Condensing the progress made in the 1960s is not easy and so this is a very selective perspective. As far as algorithm developments were concerned Bourne and Ford published a paper on stemming in 1961, Damerau reported on approaches to solve mis-spellings and in 1965 Rocchio and Salton considered how best to optimise the performance of retrieval systems. This was one of the first outcomes of the SMART project, initially at Harvard and then at Cornell, that will figure significantly in the history of the 1970s. Many of the developments of the period were reported in a new Information Retrieval section of ACM Communications from March 1964. A year earlier Information Storage and Retrieval was launched as a peer-reviewed journal, changing its name to Information Processing and Management in 1975.

Another initiative that started in the 1960s and lasted into the 1970s was ground-breaking work by Cyril Cleverdon, the librarian of the Cranfield Institute of Technology, UK on the comparative efficiency of indexing systems. It was funded by the US National Science Foundation. I had the good fortune to meet Cyril early in my career and his encouragement of my career choice was along the lines of "You will never be out of a job". How right he was.

In the 1960s advances in computer technology resulted in some very technical progress in search development in terms of both research and the availability of commercial services. IBM releases the 7090 range in late 1959 and the much more powerful 360 range in 1965. In parallel the technology to provide remote shared access to large computer centres was developed, with J.C.R.Licklider as the early innovator, leading directly to the Internet. At this point in the history of search a strictly chronological approach is not of value, and instead it is important to be aware of a number of major projects, several of which lead to commercial online services becoming available from 1965 onwards.

In terms of the impact on the underlying algorithms of search the work at System Development Corporation in the early part of the decade is of particular importance. Synthex was led by Robert Simmons with the objective of developing a system that could read and understand text, answer questions and compose an answer in readable English. The name was chosen as a tribute to the Memex concept of Vannevar Bush from 1945. There was a related ProtoSynthex project. One outcome of these projects was TEXTIR, an online search system developed for the Los Angeles Police Department in 1964 that could accept questions in natural language. Further development enabled it to incorporate synonyms into a search formulation and offer search term weighting. In parallel Hal Borko was

developing BOLD with a focus on the automatic classification of the text in documents. Yet another project was COLEX, the aim of which was to advance the development of using time-sharing services to provide online access to bibliographic databases.

These projects gave SDC the ability to launch the ORBIT online search service in 1967, a commercial service for information professionals and researchers which enabled them to search through large databases of abstracts of research literature. The project was led by Carlos Cuadra. Just a few months earlier the Information Sciences Group at the Lockheed Palo Alto Research Laboratories, led by Roger Summit, had launched the DIALOG online search service. The focus of this group was more towards scaling up online services and user interface development and one of its innovations was the display of set numbers at each stage of a query, a forerunner of facet hit numbers in current search applications. However probably the first public demonstration of computer-based information retrieval was at the 1964 World Fair with the LIBRARY/USA demonstration.

Other major centres of information retrieval science and application development in the 1960s included the work at Harvard and then Cornell University led by Gerard Salton, though this did not come to fruition until the early 1970s. Probably the most innovative was the work of Donald Hillman at Lehigh University on searching the full text of documents (the LEADER project) but mention should also be made of the SPIRES project at Stanford University (which remains one of the pre-eminent centres of information retrieval to this day) and TIP at MIT's Lincoln Laboratories. IBM was also very much involved in retrieval research on a global basis and research in to the use of computer applications for law research had been initiated. These and many other projects are described in detail by Bourne and Hahn in The History of Online Services 1963-1976 and in addition there is an excellent paper by Hahn based on the research for their book.

The importance of these online services to enterprise search is that they addressed the issues of scaling up the concepts developed in the 1950s and started to payse attention to user satisfaction, the user interface and user support. Probably the first user assessment of an online service was carried out in 1969 by Timbie and Coombs. It was not until the early 1970s that these services were available in Europe and indeed globally, a problem primarily of low network capacity and very high network access costs. The launch of these services also set a standard for the search experience for a generation of information professionals and researchers that was not challenged until the arrival of Alta Vista and then Google 30 years later. The online services showed that research services could be delivered on demand at the desktop. The next decade was primarily about improving search result relevance.

1970-1979 Enterprise search emerges

In the 1970s products emerged which are clearly the antecedents of what we would regard as enterprise search applications. From here on in the focus on academic research in this history will be significantly less, not because less research is being carried out but because it is well documented in a range of

books. In particular each chapter of [Introduction to Information Retrieval](#) by Manning, Raghavan and Schutze has an annotated bibliography and can be downloaded as a pdf. However, there are three academics that deserve mention. The first of these is [Gerard Salton](#). He developed the SMART software application as a 'test bed' at Harvard University and took it with him to Cornell University where he stayed for the rest of his career. Salton developed the cosine vector space model (VSM) to compare the relevance of a group of search results. The evolution of this model took place over a number of years and [David Durbin](#) has tried to unravel the way in which it developed, providing a good bibliography.

Karen Spärk-Jones worked in a number of departments at Cambridge University from the time of her PhD in 1964. A [profile](#) of her work whilst at Cambridge links to papers describing her research, all of which has had a major impact on information retrieval. Her [overview of information retrieval research](#) is essential reading. The third person is Stephen Robertson, a research colleague of Karen Spark Jones who went on to work at the Microsoft Research Laboratories in Cambridge. His work has extended from the mid-1970s until quite recently, the scope of which is indicated by his [list of research papers](#). Stephen is especially noted for his development of the [BM25 ranking model](#), which built on the work of Karen Spark Jones on the term frequency.inverse document frequency model.

If you want to choose a date to mark the beginning of enterprise search then 1970 is that date, It marked the launch by IBM of [STAIRS](#) (Storage and Information Retrieval System), an evolution of the AQUARIUS software that IBM developed to cope with the documentation for the defence of an anti-trust suit in the USA that started in 1969. STAIRS was specifically designed for multi-user time-share applications (the typical enterprise scenario) and remained on the IBM product list until the early 1990s. Jumping out of any sort of chronology in 1985 STAIRS was subject to a very thorough evaluation which raised doubts about the effectiveness of full text indexing, A [review article](#) by David Blair, published in 1995, is a must-read for anyone with an interest in enterprise search and evaluation as it looks back at the 1985 evaluation with the benefit of substantial hindsight, and benefits from the fact that although Blair was one of the authors of the original review it comes across as an independent and unbiased assessment.

In the Conclusions section Blair states

"We have shown that the system did not work well in the environment in which it was tested and that there are theoretical reasons why full-text retrieval systems applied to large databases are unlikely to perform well in any retrieval environment."

By the mid-1970s mini-computers were being adopted very widely, and many organisations and companies saw this as an opportunity to develop text/document retrieval software products for these mini-computers. These included BASIS (Battelle Institute) and INQUIRE (Infodata). So far this history has been dominated by developments in the USA but the mini-computer market stimulated software development in the UK, including [ASSASSIN](#) (ICI), [STATUS](#) (Atomic Weapons Research Establishment), [CAIRS](#) (Leatherhead Food Research Association) and [DECO](#) (Unilever). I had a role on the development team of DECO from 1979 – 1981. These and other applications all emerged towards the end of the 1970s.

An interesting [comparative review](#) of them was published in 1984. These applications all evolved from specific organisational requirements which were then productised for use more widely, demonstrating that you did not need to be a large academic institution or software company to develop retrieval software. These systems were accessed through networked terminals; the IBM PC was not launched until 1981. The scale of the development of these products can best be assessed from A Technical Index of Interactive Information Systems, published as [Technical Note 819](#) from the National Bureau of Standards in 1974. This report provides brief details of almost 50 software products.

The first Association for Computing Machinery (ACM) conference on information retrieval took place in 1971. The 1st Annual International SIGIR Conference on Information Storage and Retrieval took place in 1978. In 1979 the Institute of Information Scientists organised a two-day conference held at the Royal Society, London, entitled Computer Packages for Information Storage and Retrieval. The event attracted over 200 delegates.

As a footnote to this section on the 1970s it is important to highlight that the first assessment of potential role of [artificial intelligence in information retrieval](#) was published in 1976. Just over a decade later Verity, the prototype for all enterprise search applications, emerged from a company specialising in AI development.

1980 – 1989 Rapid evolution

In the early 1980s there was a great deal of interest in the UK around the use of text retrieval software running on mini-computers. The STATUS User Group in particular was very active. These vendors were not especially were interested in the commercial success of their products as the development had been justified on the need to meet internal information searching requirements within the organisation.

In the UK the Institute of Information Scientists played a very important role in stimulating interest in the capabilities of these applications through a series of Text Retrieval conferences between 1980 and 1990. The proceedings of these conferences make fascinating reading though sadly none are available in a digital format and only the 1998 and 1999 conference proceedings were published. However most of the conference documents are held by the British Library.

As far as the technical development of enterprise search was concerned probably the most important advance was the release of the [Snowball](#) English language stemmer developed by [Dr. Martin Porter](#). To be pedantic it was first released in 1979 but was not widely promoted until 1980. Martin Porter tells the story from a 2001 perspective on his [website](#) where his original stemming code and many more algorithms for various languages are available as open source. According to the [Wikipedia entry](#) the name Snowball was chosen as a tribute to the SNOBOL programming language, with which it shares the concept of string patterns delivering signals that are used to control the flow of the program.

Martin Porter also developed the Muscat (MUSeum CATalogue) search application while at Cambridge University. Released in 1984 it sought to take advantage of the work of Stephen Robertson and others on a probabilistic approach to information retrieval. Muscat Ltd. became a successful company with clients that included Fujitsu, the Japanese IT company. Muscat was eventually rewritten and released as the open source Xapian library which survived the eventual acquisition of Muscat Ltd. by a short-lived dotcom era company and is still available. There is a good summary of Muscat on the Flax website.

By the mid-1980s the IBM STAIRS full-text search application was setting the standard for enterprise search. In 1985 a wide-ranging research study was carried out by Blair and Maron of the retrieval performance of STAIRS, which at that time was being promoted as a litigation support tool. The results were far from impressive. The study is too detailed to try to summarise here, and it is better to read the paper by Blair published as a retrospect in 1995 than the original paper. This study remains the most comprehensive of its type, with nothing approaching it having been published in the last thirty years. It had commercial implications for the legal sector as this was the time when there started to be a number of major anti-trust cases brought by the US Department of Justice where reliable access to millions of corporate documents was of great importance. It should also be borne in mind that the IBM PC had been launched in 1981 and the 1980s were the time when documents started to be created on personal computers rather than being transcribed onto word processors.

I would suggest that the first commercial enterprise search application was developed by Fulcrum Technologies, established in Ottawa in 1983. This was a client-server application, rather than mainframe and offered the first API for writing information retrieval applications. It was most visible for the rest of the decade as a provider of search software for CD-ROM applications. From 1983 to 1988 Fulcrum pretty much had the search market to itself but failed to make much headway. The arrival of Verity, born in the entrepreneurial climate of California, marked a gradual decline of Fulcrum as a business. A succession of owners over the 1990s led eventually to being purchased by Hummingbird in 1997, which itself was then acquired by OpenText in 2006.

In 1985 Advanced Decision Systems was set up in San Jose, California with the objective of developing expert system and artificial intelligence applications. In 1986 David Glazer and Philip Nelson developed an innovative search application called Topic which was beta tested with success by the US Strategic Air Command. Topic made use of a probabilistic search ranking engine which offered significantly better management of ranking than the Boolean operators that had been used prior to the release of Topic, though STAIRS also used this model. This early success led to the spin-out of what was to become Verity from ADS, led by Michael Pliner with a technical team led by David Glazer and Philip Nelson. There can be no doubt that Verity was the proto-typical enterprise search application as unlike IBM STAIRS it was platform agnostic. At launch a multi-user licence cost $39,500, quite a substantial licence fee in the late 1990s.

Two other search software companies started out towards the end of the 1980s. David Thede set up dtSearch in 1988, initially offering a desktop search application. dtSearch remains one of the very few search software vendors to have been in the same ownership from start-up to the present day. Also in 1988 but across the other side of the world in Australia Ian Davies was developing the Isys software suite. This ended up being acquired by Lexmark in 2012. Several others were on the drawing board but did not emerge until the early 1990s.

The decade also marked the birth of a project at CERN in Switzerland to create what would become the World Wide Web. Tim Berners-Lee submitted his report Information Management – A Proposal in March 1989. It is important to appreciate that the initial purpose of the project was to be able to search through CERN documentation and thereby an enterprise search project was the start of the global web search business. W3C has compiled a very useful chronology of the subsequent development of the World Wide Web over the period from 1989 to 1995.

1990 – 1999 Innovation in retrieval technology

Before looking at the enterprise search business itself there were important developments in the understanding of how people searched, and in novel technical advances in search. Marcia Bates started to make us think about search behaviour in her 1989 paper on berry picking as a metaphor for the process of discovery. Peter Pirolli's work on information foraging was published in 1999. Although this is right at the very end of the decade being covered it is indicative of the research that was being undertaken looking at information systems from a user behaviour perspective, with Jakob Nielsen waiting in the wings at Sun Microsystems from 1994 to 1998. From an enterprise search perspective the work that was undertaken at the University of Huddersfield by Stephen Pollitt on faceted navigation was ground-breaking. The concept was taken up and developed further by Marti Hearst with her Flamenco project.

From a technical perspective the challenges of indexing and searching the World Wide Web were now starting to be addressed, taking search in some very different directions. Alta Vista was not the first WWW search engine but the team working on it gained an immense about of knowledge about web crawling and indexing at scale.Two members of the team founded Exalead in 2000. Google followed in 1998 and of course the arrival of enterprise web applications such as intranets opened up a potentially very large market for enterprise-level search. Sadly the IBM HITS algorithm (later integrated into the IBM Clever project) didn't have a chance against the Google PR machine. During the late 1980s and then into the 1990s advances in natural language processing were rapid as machine learning approaches and developments in machine translation opened up new opportunities for search. Latent Semantic Analysis first emerged in 1988 and Probabilistic Latent Semantic Analysis in 1999, the latter forming the basis of the Recommind e-Discovery application, now owned by OpenText. Lucene, written by Doug Cutting, also appeared in 1999.

The stage was set for the emergence of a significant number of search vendors. Verity was gaining momentum but finding it difficult to achieve profitability. In 1993 Retrievalware emerged but started a trend for search software companies to have multiple owners. How it ended up in FAST Search and Transfer via Excalibur is, to say the least, complicated.

The Infoseek/Ultraseek/Inktomi/Verity/Autonomy saga, which started in 1993, was yet another complicated journey. Interestingly Ultraseek was branded as Ultraseek Enterprise Search and by the time it was acquired by Autonomy had around 15,000 customers. Verity achieved an IPO in 1995, achieving **funding of $40m, double the amount anticipated**. This probably encouraged (at least indirectly) the arrival of Autonomy (1996), FAST Search and Transfer (1997) and Endeca (1999).

The development of the enterprise search business in the early 1990s is not well documented. Many of the entrepreneurs who had a vision for search have been interviewed by Stephen Arnold in his invaluable Wizards Index column. In the paragraph above most of the links are to Wikipedia entries, which inevitably vary in quality and depth but hopefully are at least a starting point for research. The distinguished journalist and philanthropist Esther Dyson tracked the development of internet companies during this period.

2000 – 2009 Verity, Autonomy and FAST

The decade from 2000-2009 was marked by the high visibility of Verity, Autonomy and FAST Search and Transfer and the beginning of consolidation in the search business. Verity grew rapidly over the period from 2000-2005 and started to achieve a respectable level of profitability, Revenues in 2003 were just over $100 million. These increased to $150 million by 2005 with the company sitting on around $250m in cash and investments. Inktomi (or rather the Ultraseek product) was acquired in 2003 and Cardiff Software in 2005. By late 2005 there were 160 employees and the company claimed that 15,000 companies and other organisations had licensed its software.

Over the same period of 2003-2005 FAST Search and Transfer revenues increased from $42 million to over $100 million, but the company had over 450 employees and the 2005 Annual Report is a tale of woe about a whole range of investments and other transactions. The FAST IPO had taken place in 2001. The company then sold off its web search interests in 2003, including AllTheWeb which has now reappeared as a component of Vespa. The acquisition of Retrievalware followed in 2007 but there were already concerns about the way in which the company was presenting its accounts.

In 2000 Autonomy raised $124 million of investment funds when it floated on NASDAQ and then in 2003 started the process of acquiring a substantial stable of companies, starting with the video software company Virage in 2003. Then in 2005 it acquired Verity for $500 million, a significant multiplier on $7 million net income. By 2006 Autonomy was reporting revenues of $250 million but probably half of this amount was generated by Verity. Over the next three years Autonomy also acquired Blinx, Zantaz,

Merido and probably most remarkably Interwoven, a WCMS vendor. In 2008 Autonomy became a member of the FTSE100, and by 2009 the company had revenues of £740 million and over 1600 employees.

The acquisition of FAST Search and Transfer by Microsoft in 2008 came as a surprise, as did the purchase price of $1.2 billion. It seemed to suggest that Microsoft was going to be an enterprise search provider, based around the very powerful FAST ESP search platform. However within months of the acquisition concerns were being raised about the extent to which the booked revenues of FAST Search and Transfer were being recognised, a situation that also arose in 2011 with the HP acquisition of Autonomy. One day the full story of both acquisitions may emerge. In the event Microsoft stripped out elements of FAST ESP and incorporated them into the FAST Search Server for SharePoint 2010. Such was the reputation of FAST that many organisations were under the impression that they had actually acquired the ESP product bundled into SharePoint.

Although Verity, FAST and Autonomy were the most visible enterprise search applications others were also being developed quite successfully, including Endeca, Exalead Vivisimo, ISYS Search and a number of others, but their independent existence continued for a few more years.

2010 – 2019 Consolidation

From 2010 to 2013 there was a rapid consolidation in the enterprise search business. Between 2010 and 2012 Exalead was acquired by Dassault (2010), Autonomy by Hewlett Packard (2011), Endeca by Oracle (2011), Vivisimo by IBM (2012) and ISYS Search by Lexmark (2012). Some of these vanished without trace, some notionally exist (Exalead) and of course Autonomy has now re-emerged following its acquisition by Micro Focus. Others emerged to fill the gaps. Funnelback was initially developed by CSIRO in Australia in 2001 but did not really move into the limelight until the establishment of a UK office in 2009 following its acquisition by Squiz. Lucid Imagination was set up in 2009 and was then renamed LucidWorks in 2012. BAInsight dates back to 2003 as a supplier of add-on modules to SharePoint but over the last few years has repositioned itself as more of a systems integration company. Mindbreeze, an Austrian company offering a search appliance, was founded in 2005 but as with the other companies mentioned above has flourished over the last few years. Mention should also be made to Google, which launched the GSA appliance in 2002. In 2016 it announced it was leaving the market and licences terminated in 2018.

There are currently around [50 vendors](#) of enterprise search software applications.

Looking back at the Gartner Magic Quadrant for Information Access Technology in 2005 there were four companies in the Leader/Visionary Quadrant, and they were FAST, Autonomy, Verity and Endeca. The majority of the companies surveyed in 2005 were towards the lower end of the Ability to Execute axis, and that has always been a challenge for the enterprise search business. Many companies with very good technology could not generate sales and cash flow to finance the marketing and sales effort

needed to get to a critical mass. Over the last decade the market has been dominated by Microsoft SharePoint in terms of an installed base of search functionality (perhaps close to 300,000 installations?) though Google built up a substantial installed base of appliance servers before leaving the stage. The Enterprise Search Summit was launched in New York in 2008 and the exhibition space was full with around 40 vendors. Those were the days.

The Enterprise Search Europe event was launched in 2011 but 2015 marked its closure as there were just not enough sponsors to keep the delegate fee at a sensible level. Thanks to Findwise we do now know much more about the way in which enterprise search is being implemented and used through the Enterprise Findability Surveys that started in 2011 and continued to 2016.
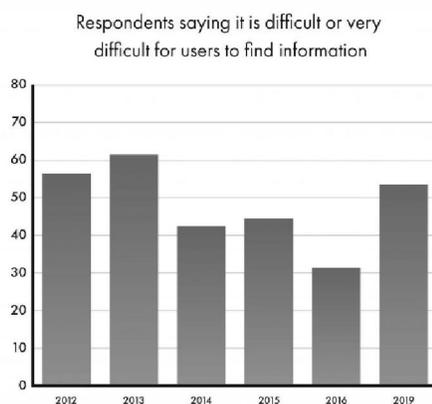
After a break of two years the survey was run again in 2019 and the chart above is reproduced from the summary report of the 2019 survey.

Over the last few years the concepts of 'cognitive search' and 'insight engines' have been proposed by two IT industry analysis firms, Gartner and Forrester. The basis of both is that search results can be customised down to the level of an individual employee based on what they are working on within the context of their colleagues. The aim of these applications is to deliver the most relevant information at position 1 on a search results page with the searcher just entering anything from a single word to a section of text they are working on. The technology involved is a combination of AI and machine learning allied to developments in natural language processing.

As yet there is no independent research that shows whether these approaches are scalable and extensible for enterprise-wide use in situations where employees are working on multiple tasks and projects simultaneously and in a range of languages.

2020 – Future positive?

Surprisingly there has been virtually no structural analysis of the enterprise search business. An initial attempt to do so was made by Deep Analysis in March 2020, just as the coronavirus pandemic struck. It

seems likely that the very rapid deployment of distributed working may surface issues over access to information, especially older information that could be of value in solving very new and novel problems. Organisations will have to support both 'Work From Home' and partial occupancy of offices and this will almost certainly have both near term and long-lasting impact on the demand for enterprise information. Enterprise search may have an almost 70 year history of development but could now become one of the most important of enterprise applications. The history continues.

Research Sources

The evolution of enterprise search is quite complicated and poorly documented. In this report I have set out a few of the milestone events and developments. They are a personal selection of history highlights and I make no attempt at being 'comprehensive'. It is 'A history' and not 'The history'.

The functionality that is now encapsulated in enterprise search software has been in constant development since the early 1950s, with especially rapid evolution in the 1970s and 1980s with the availability of large-scale commercial online search services such as Lockheed Dialog, SDC Orbit, BRS and Mead Data Central (Lexis). I started my career in search in 1975 and have had the good fortune to have met many of the early pioneers, notably Roger Summit, Charles Bourne, Carlos Cuadra, Jerry Rubin, Noel Isotta, David Raitt and Cyril Cleverdon. Whilst working in Cupertino in the early 1980s I also had the opportunity to meet research staff from Stanford Research Institute who had worked with Doug Engelbart. Other personal milestones include working on the development of one of the early UK enterprise software applications (DECO) in 1981-1982 and in 1983-84 inadvertently playing a role in the establishment of OpenText a decade later when I was involved with the conversion of the Oxford English Dictionary into a machine-readable format for editing and production.

Any history of enterprise search is intrinsically linked to the history of information retrieval, a term first used by Calvin Mooers in 1950. There have been many articles published on the history of information retrieval but by far the most readable is the chronology of information retrieval research written by Mark Sanderson and W Bruce Croft. I've always been intrigued that the URL id is 1066 and have often wondered if that was an accident or by design!

In 2019 Donna Harman published Information Retrieval: The Early Years, combining a very comprehensive bibliography of almost 300 research papers with her own experience of having been at the forefront of IR research. However there are no references to enterprise search, which is not surprising given the scope of her document.

Anyone writing a history of enterprise search is enormously indebted to Charles Bourne and Trudi Bellardo Hahn for their book A History of Online Information Services 1963-1976. Their book also provides a substantial amount of detail about enterprise search applications, though this term was not used at the time. Arguably the first ever enterprise/internal search service was set up in 1965 at the Cox Coronary Heart Institute in Kettering, Ohio by G Douglas Talbott. I would cite this as enterprise search because the application indexed content that the Institute was publishing in a quarterly internal publication.

Another excellent source is a literature review entitled Cooperation, Convertibility, and Compatibility Among Information Systems: A Literature Review published in 1966 by the US Department of Commerce that considered the issues arising from a multiplicity of information systems even at that early stage of development. This review provides a very good outline of the development of computer-based

information services dating back to the early 1950s as well as reflections on scientific communication in the widest sense from the founding of the Royal Society in London.

Stephen Robertson contributed a survey [on Computer Retrieval as Seen Through the Pages of Journal of Documentation](#) to B.C. Vickery, Ed., Fifty years of information progress: a Journal of Documentation review. London: Aslib (1994). It contains a bibliography of 146 items. Brian Vickery's career spanned much of the period covered in this history and his [personal account](#) of his work provides valuable insights into events in both the USA and the UK with regards to the role of computers in information retrieval.

Jeremy Norman's [History of Information](#) web encyclopedia, which is helpfully can be browsed through chronologically. A series of interviews with the pioneers of the pre-internet online search services was published in the [Searcher](#) magazine and these are an invaluable source of primary information on these services.

Also of **immense** historic value are the series of interviews carried out by Stephen Arnold between 2008 and 2013 and published in his [Wizards Index](#). Most of the founders of enterprise search applications tell the inside stories of how they created, launched and developed these applications.

All the links were checked on 14 May 2020 but sadly that is no guarantee that by the time you read this document they will still be active.

I would of course appreciate comments on factual inaccuracies, omissions and additional resources