

The future of search: It's how, not where, you look

By Alan Cane

Published: March 28 2007 10:13

Current situation

The time staff waste searching for “stuff” – the information necessary to do their jobs more effectively – has become legendary. Accenture, the consultancy, polled more than 1,000 executives in the US and UK and found that managers were on average spending up to two hours – a quarter of their working day – searching for stuff.

When they found it, moreover, at least 50 per cent was useless: irrelevant, out-of-date or just wrong.

Concerned that its intranet was becoming overburdened, BAE Systems, the aerospace group, carried out its own survey and discovered that four out of five employees on the network were wasting an average of 30 minutes a day retrieving information while 60 per cent were spending an hour or more duplicating the work of others.

The solution was a system from Autonomy, a UK company which, with 16,000 customers worldwide, leads the market for what is known as “enterprise search”, a family of technologies that make it possible to extract information quickly from both structured and unstructured sources. With the Autonomy system in place, BAE estimates that time spent in finding information is down by more than 90 per cent.

Another example: lawyers with the US firm Morrison & Foerster found they were drowning in information scattered through their systems: client histories were stored in accounting and customer relationship management systems, documents were stored in a document management system, communications in e-mail servers and so on.

The firm drew up a specification for an ideal solution, which it called AnswerBase, and commissioned a system from Recommind, a legal search vendor. Searches which had previously taken hours could be completed in seconds using AnswerBase; those taking days were reduced to minutes.

As Craig Carpenter, Recommind's head of marketing and business development, puts it, the days when enterprise search was a non-essential novelty are past; now the future lies with search technologies which will home in on concepts rather than keywords.

Enterprise Search - heterogeneous formats, privacy and security

Enterprise search is a comparatively recent phenomenon, forced on companies by the internet, e-mail, company intranets and the 20bn gigabytes of new data now being created by businesses each year.

Google currently leads the world in conventional internet search but as Mike Lynch, Autonomy chief executive, emphasises, enterprise search is different: “Unlike the internet, enterprise information is in different formats. A large company might support 300 different information formats scattered through 5,000 separate repositories.

“An enterprise search engine has to be able to understand all those formats and talk to all those repositories. And most staff are not allowed to see all the information a company has stored away. In a large group, for example, an individual might be allowed to see only one in every 10,000 documents. Each repository has its own set of complex rules governing who is allowed to see what and it is changing all the time.”

So Autonomy uses “spiders” and “ants” – intelligent software – to roam the intranet, indexing all the material available for a search: in that sense, even unstructured data has a structure of sorts. Ants are self-learning and capable of appreciating that particular pieces of information are frequently requested or that some categories of information change rapidly. Mr Lynch says attempts to create search tools without overall indexing – known as “federated search” – are unworkable: “They glow red hot and melt.”

Finding the relevant information

Tamara Alairys, global leader for search at Accenture, points out that using Google to search a word like "Turkey" will return thousands of hits but it will not distinguish between the country and the bird: "The challenge for people searching their intranets has been to get better search relevancy and to retrieve data that can help them make a better decision."

She argues that search technologies have improved "by leaps and bounds" in the past two years: "Early capabilities were limited: a user could only perform basic keyword searches and sort the results using parameters such as the date of creation. Much more is possible today. Structured and unstructured data can be searched. And natural language processing enables the search engine to understand the intent behind a user's query and give a meaningful response."

The cost of failing to retrieve relevant data can be high. Zia Zaman, in charge of strategic market development for Fast, a search company based in Oslo, Norway, recalls a pharmaceuticals company that entered into a strategic relationship with a drug delivery group: "The two companies invested years and millions of dollars in trying to figure out how they could work together but in the end they had to pull the plug on the deal. Then the pharmaceuticals company found a document in its own files which detailed how the drug delivery mechanism could never work. They had been making decisions in a fog."

Changes in the legal environment in the US is driving interest in enterprise search. The latest revision of the Federal Rules of Civil Procedure, the code for civil legal action, published in December year, gives companies involved in a lawsuit 99 days to produce relevant information stored electronically compared with three years or so previously.

Mike Lynch comments: "This will be impossible for a big drug company or manufacturer unless they already have a system in place. Companies which have some experience of lawsuits have already realised how important this is. Others are just waking up to it. Later this year I would expect to see the first prosecutions resulting from a failure to comply with the requirements." It is expected that other countries will follow the US lead in principle.

Improvements in the last two years

Over the past 18 to 24 months there have been significant improvements in search technology, and the number of vendors of enterprise search systems has grown. IBM, Microsoft and Google have offerings aimed at business. The top end of the market is dominated by Autonomy, Convera, Fast and Open Text while specialist players include Endeca, InQuira, Siderean Software and Vivisimo.

The result, as Jerome Pesenti of Vivisimo writes, is that the search market is fragmented and confusing but that should not stop companies experimenting: customers don't know what to ask, what features are needed and which vendors to look at, he notes, going on to argue that search should be seen as a long-term application, deployed quickly and improved in phases based on end-user feedback: "There is no limit as to how good and useful a search can be," he claims, "but modest goals, early rewards and especially, valuable user feedback, can be obtained through quick deployment."

The BBC has difficult information retrieval needs. It is awash with information: core business systems as well as financial information about programmes, approvals processes, e-mails, audio and video.

Keith Little, BBC chief information officer, says: "We have lots of information that is unsearchable – valuable information that nobody can access. We have systems with search facilities but these are silos and then there are e-mails and other repositories of unstructured data that go right across the organisation."

The BBC uses several search tools – Autonomy, Microsoft Sharepoint and OpenText Livelink among them. Mr Little says: "At the top level, our search strategy is to create a framework for plugging in, in a service-oriented manner, legacy and future systems.

"We need the ability to put those together to meet the search requirements from the business and we then have to think about how we provide access to our real audience – the people who pay our licence fees." "Infax", a simple programme search tool, was made available to the public last year.

Future developments

What lies ahead for enterprise search? Ms Alairys of Accenture sees four developments. First, advanced analytics and monitoring which will make it possible to tap information in real time and provide rapid responses. Second, sentiment analysis which uses textual analysis to gauge the tone of a document – whether results show a company in a positive or negative light, for example. Third, multimedia search across textual, video and audio sources. And fourth, guided information discovery – exploring information without a specific query.

So in future, even if you don't know what you want or where to find it, enterprise search will guide you to the right answer.