Dubbed the ‘doyen of European Software’ by the *Financial Times*, Mike Lynch is founder and chief executive officer of Autonomy, a major provider of meaning-based technologies and a leader in unstructured data management and enterprise retrieval.

Launched in 1996, the company’s global operations have grown organically and through a series of targeted acquisitions including: etalk (enterprise-class contact centre performance solutions), Virage (rich media management technology), Neurodynamics (recognition systems) and, most recently (December 2005), enterprise search competitor Verity.

Autonomy’s customers range from government agencies to energy, pharmaceutical, media, e-commerce, technology, legal and financial services, and include blue-chip customers such as British Aerospace, Royal Sun Alliance, Shell, BBC, GlaxoSmithKline, Reuters, KPMG, NASA, US Department of Defense, Ford Motor Company and Merrill Lynch.

The company has won several nominations including the World Economic Forum’s Technology Pioneer award.

Mike Lynch holds a number of advisory board and non-executive positions in companies including Apax Partners, The Carlyle Group, the Quincentenary Campaign and Avocet Capital. He speaks regularly at major educational, political and industrial events including the World Economic Forum in Davos, and he holds numerous awards including the Electrical Engineers' Medal for Outstanding Achievement and the Confederation of British Industry's Entrepreneur of the Year (1999).

Dr. Lynch is a member of the Campaign Board of Christ's College, Cambridge of which he is an alumnus.

---

**Guru Interview:**

**Mike Lynch, OBE**

*Mike Lynch talks to Sarah Powell about meaning-based computing, a technology developed by his company, Autonomy.*

Interview by Sarah Powell
Your company’s pattern-recognition technology, ‘meaning-based computing’ (MBC), is designed to help organizations extract meaning from ‘unstructured’ data. Can you elaborate?

Mike Lynch:

First we really need to talk about the different types of computer information in the world. We’re all familiar with the sort of data we find inside a database, which is structured or organized so that the computer can process it. However, we actually spend most of our time dealing with ‘human-friendly’, unstructured information. By this I mean e-mails, TV programmes, videos, phone calls, newspaper articles and so on. Unstructured information makes up an estimated 85 per cent of all the information inside an enterprise.

The difficulty with unstructured information to date has been that computers can’t really understand it, which means human beings have to deal with it, retrieving and collating the information, and doing whatever job is then required. When attempting to retrieve this sort of information we have had to conduct keyword searches, i.e. choosing a word or words for the computer to search for. The computer then tells us which documents have these words in them. The important issue here is that the computer has no understanding of the meaning. All it is doing is finding all the documents with a particular word or combination of words in them.

With meaning-based computing the computer can actually derive a meaning from the information, i.e. can actually understand it and recognize the relationships between different data. What that means is that it can then carry out much more powerful retrievals. For example, if searching for information about a dog, the computer would know that a hound is another word for a dog – it recognizes that the meaning is similar.

In addition to this much enhanced search capability, instead of simply retrieving information to enable a human to perform a task, the computer can actually perform tasks itself, e.g. automatic diary entries, bookings. Meaning-based computing has moved far beyond keyword-based technology.

In a contextual search, how can the software recognize that the word hound, for example, means the same as the word dog? Do keyword synonyms have to be introduced into the programme?

Mike Lynch:

No, that’s really not a feasible method. People initially tried to do this, but it makes things far too complicated because not only do you have the work of putting in all this information, but it all becomes conditional – ‘she’s a star’ doesn’t mean ‘she’s a cosmic gas ball’! So, no, the clever thing about these meaning-based technologies is that they learn by reading and looking at the context, just as a human being does. For example, imagine you’re reading a newspaper article in the UK and it suddenly starts talking about ‘Gazza’ – you may not know that Gazza equals Paul Gascoigne, the British footballer, but you will see from the context in which the name Gazza is used that this is who it is referring to. Meaning-based computing works in the same way. Getting back to that dog search we talked about, through reading and looking at the context the system will deduce that the word hound relates to the word dog.

Autonomy’s customers are extremely varied in their operations. Can you offer examples of how your clients use your technology?

Mike Lynch:

Meaning-based computing responds to a generic need to process unstructured information. One of the things that surprises people is that Autonomy has customers in almost every single area of industry. Whether you are a concrete manufacturer, aeroplane designer or investment bank, you still need to handle the unstructured information that is being constantly generated.

Meaning-based computing is very useful in many different industries and is used in many different ways. For example, in an investment bank meaning-based computing can enable the computer to ‘read’ emails to ensure there is no risk of breach of securities laws, which could result in vast fines. In government, where a considerable amount of intelligence is being processed, MBC can prioritize what to look at, discerning the relationships between different pieces of data. In a company call centre, where the system listens in to conversations between reps and customers, meaning-based computing can automatically flag up to the rep the correct answer to a question.

There is an immensely wide range of potential applications and these are mirrored in our clients’ usage. A British police force employs meaning-based computing in its major crime investigations and at British Aerospace it enables analysis and sharing of the information that engineers are working on. The US Department of Homeland Security adopted the software across its 21 agencies to monitor suspected terrorist groups. The BBC uses it to locate picture and information clips for its news services while KPMG Forensic builds on its capabilities in commercial fraud investigations and disputes. AstraZeneca for its part was attracted by its
potential in the identification of R&D opportunities.

**Given its capacity for contextual search, this technology sounds as if it would be invaluable for researchers?**

**Mike Lynch:**

It is, and one of the nice things about meaning-based computing is that it is invaluable when you don’t know exactly what you are looking for. Because the computer can comprehend meaning, and understands which ideas resemble one another even though they are not exactly the same, a contextual search enables it to highlight links and relationships which no human being would be able to do given the sheer volume of information that would have to be read to make these associations. This capability means MBC is widely used in universities for making research material available, in the production of teaching aids, making videos searchable, and so on. Again, much of this is unstructured information and that’s what MBC really centres on.

**Meaning-based computing is based in part on your Cambridge doctoral research, submitted in 1990, six years prior to the launch of Autonomy. What were you working on at the university that led to the emergence of MBC?**

**Mike Lynch:**

Meaning-based computing comes down to the academic problem of pattern recognition which is the ability to take a mass of information and understand the pattern or meaning behind it. This is applied in many areas such as speech recognition, image recognition and understanding documents. At that time I was doing a lot of work in trying to get computers to understand hard meaning using pattern recognition. This work was based on the research into mathematical probability of an eighteenth-century English country vicar and mathematician, Thomas Bayes. Over the past 15 or so years there has been a resurgence of interest in the significance of his equations. MBC also builds on the work of Claude Shannon, ‘the Father of Information Theory’, whose ‘Principles of Information’ enable identification of the patterns that naturally occur in text. These two sets of research lie at the heart of MBC. Of course the work that I was doing while a Research Fellow at the University of Cambridge was theoretical, being at the mathematical end of things. It was when I left the university that we started to work on practical applications.

**What are the technological challenges of your pattern-matching techniques, e.g. do some language patterns pose difficulties for MBC’s technology, and is this an argument for retention of some degree of manual processing?**

**Mike Lynch:**

There can be difficulties when it is unclear what the subject matter is. Meaning-based computing also has difficulties with sarcasm. The problem derives from the fact that sarcasm and irony make perfect sense – but can mean quite the opposite of what the words might suggest. On the other hand, the programme is surprisingly good at understanding slang.

“One of the things that surprises people is that Autonomy has customers in almost every single area of industry. Whether you are a concrete manufacturer, aeroplane designer or investment bank, you still need to handle the unstructured information that is being constantly generated.”

In answer to your question about manual processing: yes, the ideal is a combination of human and machine input. If, for example, in a large company a decision needs to be made as to whom to forward an email from a customer, a machine-generated operation is likely to result in a lower quality performance. The solution is for the computer to recognize that it is unsure what to do, and to alert an operator.

**Can the technology be applied to other languages, including those with different scripts, e.g. Arabic, Russian, etc., and if so is there a requirement for technicians in these languages to adapt the software?**

**Mike Lynch:**

Meaning-based computing works in all languages and has already been used in Japanese, Arabic and Russian (and even, for sheer fun, in ‘Klingon’ – the Star Trek language!). There is no need for any knowledge of languages at all – it is all a question of getting a mass of material so that the computer can learn the meaning through reading.

You have spoken out strongly against President Jacques Chirac’s proposal to create a European
multimedia search engine, Quaero. What is your objection to this project?

Mike Lynch:

The Quaero project is a result of misguided and unnecessary nationalism, which will undoubtedly result in the suffocation of national entrepreneurs and start-up companies currently operating in the search market. The Franco-German consortium of large lumbering industrial corporations is aiming to research an area of technology that has already been well advanced by smaller, more nimble companies that are capable of achieving complex multimedia recognition and search and retrieval functionality. Image and sound recognition software already exists, and by the time Quaero develops its search engine the market will already have progressed to the next generation of search and retrieval applications.

Investing over £206 million in an area of the technology market that is already developed is not only a huge waste of money but will also create frustration amongst those start-ups, including various French semantic search companies, that are operating in the same market and are in search of government funding. Creating conditions under which innovators and entrepreneurs can flourish should be the focus for the French government, rather than wasting money and effort on trying to outstrip their own entrepreneurs. Historically such projects in non-capital-intensive areas like software have failed; they have only been successful in capital-intensive areas like airliners.

Autonomy was the first UK dotcom company to announce a profit (in 2000) and you were hailed as the UK’s first dotcom dollar billionaire. How was your research and development funded in the early days and, given your experience, what do you see as the major challenges facing UK academics wishing to create a spin-out company from their universities?

Mike Lynch:

The main problem for academics wishing to spin out companies is getting hold of what’s called ‘angel finance’. Getting the first £50,000 is very difficult in the UK – although much easier in the USA – because it is too small an amount to interest venture capitalists in the initial stage. In our case we managed to secure just £2,000 in funding for our university research – a loan from a rock band promoter I happened to meet in a wine bar!

At the time of our spin-out, the university was unsure of its strategy vis-à-vis such ventures, which meant it couldn’t really help us. From a business point of view we were constrained by lack of funding and were unable to hire the staff to establish a formal business structure. However, the company survived. Shortly after we moved into profit we sold part of it to venture capitalists – who did extremely well out of their investment, seeing their £2 million turn into £1 billion.

You make 52 per cent of your sales in the USA and are headquartered in both Cambridge and San Francisco. How important is your Cambridge base and how important to the university is the city’s thriving business and technology hub?

Mike Lynch:

It would be hard to argue that Cambridge is not the premier academic/business hub in Europe. Apart from the number of high tech companies here, for the business community the presence of a top university and the links between the two communities are very important, and vice versa. The presence of a leading university and a vibrant business hub attracts the best people both to the university and to the business community, creating an attractive environment that also draws in excellent students. It’s a symbiotic situation. And of course the fact that English is now the standard global language is a major asset.

My work as a Research Fellow at Cambridge was generic and purely theoretical. I developed meaning-based computing after I left, and Autonomy owns the intellectual property. Today the University of Cambridge is far more geared up to support spin-outs and I believe it is important for universities to take a share in IP created within their faculties.

Intellectual property is an excellent generator of income for universities and even a modest holding in a university set-up can generate a substantial return.

Republished from Emerald Now, August 2006.