Track the news

What is RSS?

Rich Site Summary or Really Simple Syndication is an easy way to track the latest news

Search

Click here for your how-to guide to searching FT.com

Back to results

FT REPORT - DIGITAL BUSINESS 2007: Intelligent systems grapple with phonetics

By Alan Cane, Financial Times
Published: Nov 07, 2007

I wish my computer could wreck a nice beach." A curious ambition, admittedly, and one that invokes images of sun-kissed sand and cool sea despoiled by heaps of discarded 1960s mainframes.

Fortunately for Bondi, Daytona and other famous resorts, the actual sentence was: "I wish my computer could recognise speech."

Mike Lynch, chief executive of Autonomy, the Cambridge company with a leading position in the market for the storage and retrieval of unstructured data, uses this example to illustrate the difficulties computer systems face in searching audio recordings for specific content.

Searching audio, however, is simple compared to searching video, where even the best technology falters. Even top digital forensic specialists such as the international group Kroll Ontrack has only just added audio search capabilities to its electronic discovery systems and has no video searching tools at present. This is a company with software capable of searching 30m pages of text in hours: searching voice mail and other audio files is much more difficult.

Martin Baldock, the company's operations manager says it is adding audio search tools from an as-yet unnamed supplier to its e-discovery system: "It takes the audio file and breaks it down into phonemes. When these phonemes are indexed, the file becomes very searchable. It's not 100 per cent accurate but from the initial testing we have done, we are confident of getting 70 per cent accuracy."

And that can save a lot of time and money. Kroll Ontrack is typically called in when searches are needed to support litigation: "We have recently been looking at trading floors in energy companies where strange things have been happening with trades," Mr Baldock says. "Historically, we would have sat down a paralegal who would have listened to hours and hours of recording. These systems can reduce that dramatically."

Every year the volume of stored digital audio and video rises and with it the need to be able to home in on content with reasonable accuracy.

Who needs this capability? Everyone from teenagers hunting down a YouTube video, to investigators looking for dodgy deals, to performing rights

administrators and call centre management.

According to Mr Lynch of Autonomy, audio search technology is in its third iteration. The first involved speech recognition engines of the kind now commercially available for domestic PC users which take an audio file and turn it into text.

But this has never been too successful: "A clean audio source for an unknown speaker using a large vocabulary in normal, connected, speech might give 80 per cent to 90 per cent accuracy," says Mr Lynch, whose PhD topic focused on speech recognition. Add in a poor phone line and background noise and the accuracy falls away precipitously. Second-generation systems tried to pick out phonemes - the smallest sound units - but, as the "wrecked beach" example shows, tended to fall over when matching phonemes to words.

The most recent systems use "language model based speech recognition" which combines guesses at what a block of phonemes might mean - with probability of accuracy appended - together with an attempt at context. These are intelligent systems which are self-learning and keep up with new expressions and jargon. Several months of trading calls could be searched in a couple of days using these systems, says Mr Lynch.

While there are systems that can recognise elements of video images - software used to detect pornography on the internet looks for an over-abundance of skin tones, for example, while searching for a picture of a horse in a field might bring up horses but also cows and pigs in fields, too - most of today's video searching systems depend on hybrid technologies. These search the soundtrack to a piece of video using a speech recognition system, the logic being that "Mr Clinton" and "Labour Party Congress" is likely to be linked to that part of a video showing the former US president addressing the party conference in Manchester in 2006.

According to Clive Longbottom of the consultancy Quocirca: "Graphics such as photographs can be searched through the use of colour, shape and texture (CST) technologies.

"Video is more difficult but is probably best dealt with as a hybrid system - the sound can be converted to text while the video itself can be stripped down to a representative set of stills that can be indexed through CST, along with some basic, manually applied metadata (searchable text added to the video images) providing extra back-up."

Bill Scott leads an IBM group working with media companies including the BBC to develop archiving and search techniques. The company has developed a way of searching video using a proprietary speech-to-text engine developed in its laboratories in Hawthorne, New York State, coupled with analysis of the resulting text.

"That uses another of our capabilities called semantic supercomputing which comes out of our Almaden Research Centre in San Jose. This is combined with analysis of the video stream itself. We see this being used by media groups especially for cataloguing archive contents," Mr Scott said.

Audio and video recording, however, carries with it legal responsibilities. As Paul Birch, head of forensic technology services at accountants BDO Stoy Hayward points out, the UK Data Protection Act requires that people know that they may be recorded, while the Surveillance Act forbids the recording of individuals covertly: "If the people carrying out the recording do not have an understanding of legal rights, then their material becomes inadmissable as evidence and it could leave them open to prosecution themselves."

He argues that somebody with a responsible position in a company should have charge of these activities.

And Mark Davenport of the law firm Eversheds thinks computerised search may be over-hyped. He says all documents - audio and video files - disclosed in legal proceedings should be reviewed first by a company's legal advisers: "This may well lead to a costly review process. However, voice transcription technology which transcribes audio content to readable and searchable forms does exist," with the implication of saved time and expense.

But he goes on: "The unavoidable reality is that information central to a client's case should always be manually reviewed for relevance. This is especially true while this new technology is in its infancy."

So those long-suffering paralegals may be condemned to a few more years in front of recorders with headphones on before industrial strength audio and video search comes to their rescue.