

As a professor of linguistics, David Crystal has spent over 12 years developing software that can truly understand online content. Sean Hargrave reports

Word for word

As one of the UK's most respected linguists, Professor David Crystal has spent over 12 years developing software that can understand web content the way people can. But his latest focus is a refrigerator.

Crystal's Textonomy technology doesn't just recognise words, it understands the several meanings a word can have and decides which one applies according to context. Having come out of academia, the core technology is being continually updated to suit commercial contexts. So 'refrigerator' was a recent addition to the 250,000-strong list of words that it uses.

To take Crystal's favourite example, if a web page is talking about 'depression' it could be being used in its economic, psychiatric or meteorological sense. By looking at the word in isolation, software would have no idea which meaning applies, but Crystal believes years of semantics work means his system can.

"Many of the words we used every day have several meanings and we can judge which one applies by the context of the conversation," he says. "Likewise, if a page is talking about cold fronts, our software knows 'depression' is being used to talk about the weather; if there were medical terms, we'd know that part of the page was about psychological depression."

Other systems that look for strings of letters that make up a word are a quick fix, says Crystal, because they don't understand the meaning of that word. Without this ability, he claims that ad serving software can easily make mistakes and end up wasting budget and risking brands being embarrassed by their names getting associated with inappropriate content.

His favourite example here is of a CNN report on a stabbing that he once read online. "The ads down the side were all for knives," he says. This is why the ad network Ad Pepper bought his company, Crystal Semantics, for £2m in 2006, because it believes it can provide a more reliable way of matching content to ads. "If our tag had been used then it would have served up crime-related ads," says Crystal.

Ad Pepper has incorporated his technology into its iSense vetting technology, which can stop ads being served against unsuitable content. It's currently being trialled by the NSPCC (NMA 21.02.08). But Crystal says it could be used to get around the problem of publishers on an ad network packing pages with words and phrases they know will attract the biggest volumes of advertising, regardless of whether that's what the page is about.

While advertising is now the public face of his 12 years' work on web semantics, it wasn't his prime motivator for becoming involved in linguistics online. Advertising was chosen as

“IF A PAGE IS TALKING ABOUT COLD FRONTS, OUR SOFTWARE KNOWS ‘DEPRESSION’ IS BEING USED TO TALK ABOUT THE WEATHER”

the main route for Crystal Semantics because, six years ago, he could see that it was most likely to lead to a return.

After leaving his post as professor of linguistics at Reading University just over 20 years ago, to move back to his home town of Holyhead to pursue a career as an academic author, Crystal was asked to edit a concise dictionary for the University of Cambridge. When this was completed in the early 1990s, the university realised it would be useful to have the terms cross-referenced so people on campus could use a computer to search for a term and then filter the correct meaning before being given a screen full of information on the subject.

When the Cambridge project was completed,

it sparked the idea from Crystal that this could be used to help people find information on the embryonic internet. So he spent the latter half of the 1990s working at first with the university and then with AND, a Rotterdam-based company (which bought out the university's interest in the project), to make a database of terms searchable so that people could search for a word online, select the meaning of the word and then be delivered relevant information.

By then, though, a search engine called Google was making waves on the web. AND went bankrupt in 2000, so Crystal and a business partner bought the technology and continued to develop it from 2001 to 2005, until they were bought the next year by Ad Pepper.

When he set up Crystal Semantics, Crystal foresaw advertising as just one application. He still claims it could also be used to spot a paedophile or terrorist at work in a chat room or social media site, as well as provide better results for shoppers searching for keywords.

"Our software can understand what's being said, so I was, and still am, convinced it could catch paedophiles, because they tend to ask the children they're grooming predictable questions. It can also help link a search term on a site to its catalogue of goods."

He now splits his week evenly between writing and overseeing iSense. On the writing front, he has just released *Think On My Words*, a book about Shakespeare, and will follow this up in June with *Txtng: The Gr8 Db8*, in which he will argue that using text-speak doesn't reduce young people's vocabulary.

But what Crystal is about to find out over the next month or two is whether his labour of love for the past decade really is capable of being commercialised. This may not become clear until the beta triallists report back after Easter. But even so, Crystal believes still that, with the explosion of social media, the technology could have an important role to play in moderating sites and protecting users, as well as ensuring they see the most relevant ads. ■

CV Name Professor David Crystal (OBE) **Title** R&D director, Ad Pepper **Age** 67 **Education** St Mary's College, Crosby, Liverpool; read Linguistics at UCL **Career** 1963-65: Lecturer in linguistics, University of Bangor; 1965-84: Lecturer in then professor of linguistics, Reading University; 1984-2001: Freelance author and linguistics researcher; 2001: Founder, Crystal Semantics (bought by Ad Pepper in 2006); 2006: academic author and R&D director of Ad Pepper