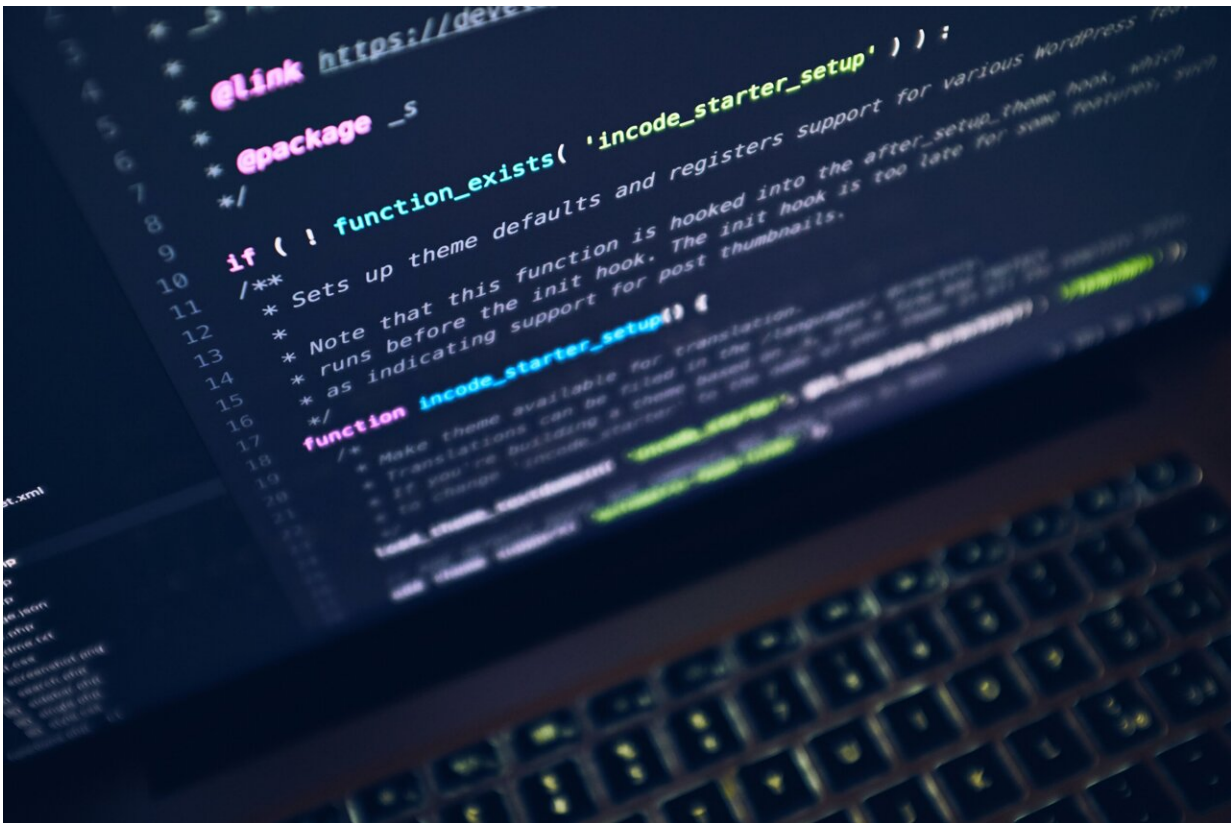# Search technology that can gauge opinion and predict the future

August 16 2012

Credit: Unsplash/CC0 Public Domain

Inspired by a system for categorising books proposed by an Indian librarian more than 50 years ago, a team of European researchers have developed a new kind of internet search that takes into account factors

such as opinion, bias, context, time and location. The new technology, which could soon be in use commercially, can display trends in public opinion about a topic, company or person over time - and it can even be used to predict the future.

'Do a search for the word "climate" on Google or another search engine and what you will get back is basically a list of results featuring that word: there's no categorisation, no specific order, no context. Current search engines do not take into account the dimensions of diversity: factors such as when the information was published, if there is a bias toward one opinion or another inherent in the content and structure, who published it and when,' explains Fausto Giunchiglia, a professor of computer science at the University of Trento in Italy.

But can search technology be made to identify and embrace diversity? Can a search engine tell you, for example, how public opinion about climate change has changed over the last decade? Or how hot the weather will be a century from now, by aggregating current and past estimates from different sources?

It seems that it can, thanks to a pioneering combination of modern science and a decades-old classification method, brought together by European researchers in the LivingKnowledge project. Supported by EUR 4.8 million in funding from the European Commission, the LivingKnowledge team, coordinated by Prof. Giunchiglia, adopted a multidisciplinary approach to developing new search technology, drawing on fields as diverse as computer science, social science, semiotics and library science.

Indeed, the so-called father of library science, Sirkali Ramamrita Ranganathan, an Indian librarian, served as a source of inspiration for the researchers. In the 1920s and 1930s, Ranganathan developed the first major analytico-synthetic, or faceted, classification system. Using this

approach, objects - books, in the case of Ranganathan; web and database content, in the case of the LivingKnowlege team - are assigned multiple characteristics and attributes (facets), enabling the classification to be ordered in multiple ways, rather than in a single, predetermined, taxonomic order. Using the system, an article about the effects on agriculture of climate change written in Norway in 1990 might be classified as 'Geography; Climate; [Climate change](); Agriculture; Research; Norway; 1990.'

In order to understand the classification system better and implement it in search engine technology, the LivingKnowledge researchers turned to the Indian Statistical Institute, a project partner, which uses faceted classification on a daily basis.

'Using their knowledge we were able to turn Ranganathan's pseudo-algorithm into a computer algorithm and the computer scientists were able to use it to mine data from the web, extract its meaning and context, assign facets to it, and use these to structure the information based on the dimensions of diversity,' Prof. Giunchiglia says.

Researchers at the University of Pavia in Italy, another partner, drew on their expertise in extracting meaning from web content - not just from text and multimedia content, but also from the way the information is structured and laid out - in order to infer bias and opinions, adding another facet to the data.

'We are able to identify the bias of authors on a certain subject and whether their opinions are positive or negative,' the LivingKnowledge coordinator says. 'Facts are facts, but any information about an event, or on any subject, is often surrounded by opinions and bias.'

From libraries of the 1930s to space travel in 2034...

The technology was implemented in a testbed, now available as open source software, and used for trials based around two intriguing application scenarios.

Working with Austrian social research institute SORA, the team used the LivingKnowledge system to identify social trends and monitor public opinion in both quantitative and qualitative terms. Used for media content analysis, the system could help a company understand the impact of a new advertising campaign, showing how it has affected brand recognition over time and which social groups have been most receptive. Alternatively, a government might use the system to gauge public opinion about a new policy, or a politician could use it to respond in the most publicly acceptable way to a rival candidate's claims.

With Barcelona Media, a non-profit research foundation supported by Yahoo!, and with the Netherlands-based Internet Memory Foundation, the LivingKnowledge team looked not only at current and past trends, but extrapolated them and drew on forecasts extracted from existing data to try to predict the future. Their Future Predictor application is able to make searches based on questions such as 'What will oil prices be in 2050?' or 'How much will global temperatures rise over the next 100 years?' and find relevant information and forecasts from today's web. For example, a search for the year 2034 turns up 'space travel' as the most relevant topic indexed in today's news.

'More immediately, this application scenario provides functionality for detecting trends even before these trends become apparent in daily events - based on integrated search and navigation capabilities for finding diverse, multi-dimensional information depending on content, bias and time,' Prof. Giunchiglia explains.

Several of the project partners have plans to implement the technology commercially, and the project coordinator intends to set up a non-profit

foundation to build on the LivingKnowledge results at a time when demand for this sort of technology is only likely to increase.

As Prof. Giunchiglia points out, Google fundamentally changed the world by providing everyone with access to much of the world's information, but it did it for people: currently only humans can understand the meaning of all that data, so much so that information overload is a common problem. As we move into a 'big data' age in which information about everything and anything is available at the touch of a button, the meaning of that information needs to be understandable not just by humans but also by machines, so quantity must come combined with quality. The LivingKnowledge approach addresses that problem.

'When we started the project, no one was talking about big data. Now everyone is and there is increasing interest in this sort of technology,' Prof. Giunchiglia says. 'The future will be all about big data - we can't say whether it will be good or bad, but it will certainly be different.'

Armed with the project's Future Predictor, Prof. Giunchiglia is well equipped to make that prediction.

More information: livingknowledge-project.eu/

Provided by CORDIS

provided for information purposes only.