NewScientist   Tech

Home   News   In-Depth Articles   Blog   Opinion   Video   Galleries   Topic Guides   Last Word   E-Newsletter   Subscribe   Look for Science Jobs

SPACE   TECH   ENVIRONMENT   HEALTH   LIFE   PHYSICS&MATH   SCIENCE IN SOCIETY

Software distinguishes between online namesakes

14:35 16 November 2006 by **Tom Simonite**

Scouring the web for someone with a common name could become easier with software that automatically distinguishes between individuals by analysing the details of search results.

The software tool, developed by researchers at the University of Tokyo in Japan, picks apart the results of a search engine query, identifying unique identities within these results. For example, it can tell the difference between Michael Jackson the pop singer and a travelling beer expert of the same name, who also appears on the first page of results produced by Google.

The program analyses the first 100 results returned by Google in response to a name search. It then examines common words in each summary to see if any results may relate to different people with the same name.

The program clusters together results thought to relate to different people. For example, it creates one cluster of results for pop signer Michael Jackson, defined by key words such as "music" and "trial", and another for beer expert Michael Jackson, defined by the words "beer" and "travel". In testing, the software was between 70% and 95% accurate at telling apart people with the same name.

Linguistic rules

The software also analyses full-page results to identify words or phrases that may be relevant to each unique person. It considers the frequency with which words and phrases appear and their grammatical relationship with the name in question.

"The keywords extracted by the algorithm can be used to suggest better queries to the user" to help them refine their search, says Danushka Bollegala, who developed the system with colleagues Yutaka Matsuo and Mitsuru Ishizuka.

For example, if a person searches for "Jim Clark" using the system, it will tell them how many different Jim Clarks there are and which keywords to use if they want to search a particular one.

But Bollegala believes the system could ultimately prove useful in other areas of computer science, such as language processing. This is because it can provide a better understanding of a statement by distinguishing between ambiguous statements. "We are working to extend the method to disambiguate other types of named entities, such as products, organisations and geographical locations," says Bollegala.

Semantic web

Victoria Uren, at the UK's Open University, agrees that such techniques could have widespread benefits. "Solutions to this problem are badly needed for the semantic web to work," she says. The "semantic web" involves enabling computers to process the meaning of online documents.

Andrew MacFarlane, who researches information retrieval at City University, UK, agrees that the system could be especially useful for web searching, but doubts whether any commercial search engine would be interested in implementing such a tool. "Since they get their revenue from advertising they will focus on improvements that return more clicks on adverts," he says.

The work was presented at the 17th European Conference on Artificial Intelligence this summer in Riva del Garda, Italy.

🖶 PRINT   ✉ SEND   ◄ SHARE ▪💬👍◆▾

More   Latest news

Trash trackers: The secret life of garbage 📷

17:21 18 September 2009
People fill their trash cans one day, and it's all gone the next – but no one can be sure where it goes. A pioneering experiment will find out

Better world: Pimp your house

15:27 18 September 2009
Our homes account for around a third of our carbon emissions, so every eco-warrior should try to make their pad a low-emission one
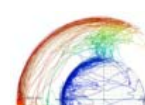
Army heli-Weeble hops to avoid rubble trouble 📷

13:08 18 September 2009
Toy-inspired remote-controlled aircraft aims to conquer rough terrain and reach places that other drones can't – through a series of hops

Visions of data

12:15 18 September 2009
See new ways of visualising data, showcased at a recent conference, from penguin tracking to beating forged signatures
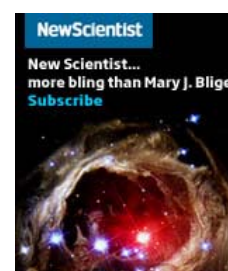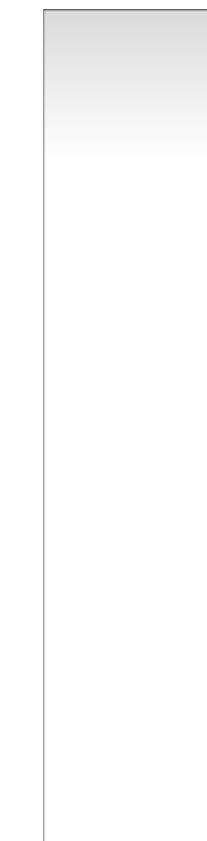
see all related stories

Most read   Most commented

This week's issue

there are a variety of licensing options available for use of articles and graphics we own the copyright to.

Have your say

Comment title

Your name

Email

Website

Comment

cancel   preview   **submit**

All comments should respect the New Scientist House Rules. If you think a particular comment breaks these rules then please use the "Report" link in that comment to report it to us.

If you are having a technical problem posting a comment, please contact technical support.

search New Scientist   Go »

Too much radiation for astronauts to make it to Mars

13 things that do not make sense

Could we create quantum creatures in the lab?

Brain science to help teachers get into kids' heads

Science fiction: The stories of now

TWITTER

New Scientist is on Twitter

Get the latest from New Scientist: sign up to our Twitter feed

Related Jobs

Software Engineer II - Cancer Group, Connectivity Map

Biochemistry and Molecular Biology Research Associate

Senior Research Executive - Healthcare / Pharmaceutical

Partners

We are partnered with Approved Index. Visit the site to get free quotes from website designers and a range of web, IT and marketing services in the UK.

Login for full access   **Login**

About us	User Help	Subscriptions	Links
New Scientist	Contact Us	Subscribe	Site Map
Advertise	FAQ / Help	Renew	Browse all articles
Recruitment	Disclaimer	Gift subscription	Magazine archive
Advertising	Ts & Cs	My account	NewScientistJobs
Syndication	Cookies	Back issues	The LastWord
Who's who	Privacy Policy	Customer Service	E-Newsletter
RBI Jobs			RSS Feeds

© Copyright Reed Business Information Ltd.