

[Home](#) | [Tech](#) | [Science in Society](#) | [News](#) | [Back to article](#)

Noise could mask web searchers' IDs

07 March 2009 by **Paul Marks**

Magazine issue **2698**. [Subscribe](#) and get 4 free issues.

ADDING noise to search-engine records could help keep surfers' identities private. A team from Microsoft Research in Mountain View, California, says the technique is a major step towards "provable privacy".

Records of internet searches made on websites such as Google and AOL are hugely useful to software engineers trying to improve search technology. Such data also give social scientists a valuable window on our largely uninhibited digital search behaviours. The problem is that such information can easily identify individuals who have carried out the searches, breaching their privacy.

Until now, search engines' attempts to anonymise this data have proved somewhat inept. In one case, [AOL replaced its customers' names with random numbers](#) before making the data public. But some queries proved so specific - such as people searching their own names and social security numbers - that reporters from *The New York Times* were able to use them to track down one individual.

People search for their own names and social security numbers, which renders them identifiable

Now a team of researchers at Microsoft's search technology lab – Krishnaram Kenthapadi, Nina Mishra, Alex Ntoulas and Aleksandra Korolova – have developed a safer way to release search query logs. The idea is to publish data associated only with the most popular queries, so that rarely performed searches cannot be used to identify people with special, or even downright peculiar, interests.

Another way to "crack" anonymous data sets is to see how they compare to known data, a bit like identifying countries on an unlabelled map by comparing their shapes to a labelled map. To prevent this, the team inserts noise into the data by adding digits to the figures it contains. "Adding the noise gives provable privacy," says Korolova. And the amount of added noise defines the level of privacy that can be guaranteed. The work will be presented at the [World Wide Web conference](#) in Madrid, Spain, in April.

Is the resulting data set still useful to researchers, though? "It depends what you use it for," says Korolova. The team has tested the technique by checking how search patterns reveal a ranking of people's phobias. They found a similar set of rankings regardless of whether they used the privacy technique or not. "We think we have a balance between a guarantee of privacy and the utility of the data sets," says Korolova.

Ian Brown, a researcher at the Oxford Internet Institute in the UK, is impressed: "This shows privacy can be preserved in many types of systems without having to trade it against functionality."

ADVERTISEMENT

The advertisement features a stack of three NewScientist magazine covers. The top cover has the headline "Home alone" and a sub-headline "The best job in science". The middle cover has the headline "Old killer, new hope" and a sub-headline "The speed of light is special". The bottom cover has the headline "Why Einstein was wrong about relativity" and a sub-headline "The speed of light is special". A blue circular button with the word "GO" in white is positioned above the stack of covers. The NewScientist logo is visible at the top left of the advertisement.

New Scientist

Not just a website!

Subscribe to New Scientist and get:

51 issues of New Scientist magazine delivered to your door

unlimited online access to articles from over 500 back issues

[Subscribe Now and Save](#)

If you would like to **re-use any content** from New Scientist, either in print or online, please [contact the syndication](#) department first for permission. New Scientist does not own rights to photos, but there are a [variety of licensing options](#) available for use of articles and graphics we own the copyright to.

[Back to article](#)



ADVERTISEMENT

