



Web spy software hacks into secretive online forums

■ 13 April 2010 by [Shehryar Mufti](#)

THE dark corners of cyberspace are being illuminated by indexing software that can reach into secretive websites that are normally inaccessible to search engines. This could allow search engines to cover online forums lurking within the "dark web", and provide insights into what is being said by groups who would rather keep their conversations secret.

Conventional [search engines](#) use programs called spiders or web crawlers that scuttle around the internet and index what they find. However, many websites are protected by security restrictions that fend off such software.

software also disguises its indexing activity by making it look like the traffic generated by users browsing the forum. What's more, it can attempt to sign up for membership on forums that require registration, though it has to seek help from Chen's team if unusual information is asked for. To help it index text in languages other than English it uses [Google Translate](#), Google's online translation engine.

Unlike a regular web crawler, Chen's software looks only at sites he has specified. It has compiled data on 29 restricted forums, containing about 13 million messages in total.

The software disguises its indexing activity to look like traffic generated by users browsing the forum

Screening out all traffic from IP addresses belonging to well-known search engines is one way to do this.

The dark web can provide a haven for extremist groups to exchange ideas, says [Hsinchun Chen](#), director of the artificial intelligence laboratory at the University of Arizona in Tucson. So Chen and his team devised software to access and index protected online forums (*Journal of the American Society for Information Science and Technology*, DOI: [10.1002/asi.21323](#)).

One of the tricks deployed by Chen's software is to regularly change the apparent IP address of the computer on which it is running. The

On one forum, it took just 39 minutes to index 29,016 posts made over a six-week period.

Chen's team is now analysing the conversations on these forums to build an overview of the links between participants. He suggests this may be useful in identifying prominent members.

The impressive thing about Chen's forum crawler is the way it combines human guidance and automated web searches to catalogue dark web forums, says Denis Roy, a spokesman for Yahoo. "The name of the game," he says, is to "find the right blend of the least possible number of humans and machines" to perform this indexing of restricted websites efficiently.