



XRay: Tool to track personal data use online

New Software Developed By American Researchers Aims To Combat 'Data Frenzy' On Web

Steve Lohr

© Tetra Images/Corbis



DIGITAL EYE: The XRay team hopes to have a more robust, general tool ready within a year. The most likely users, team members say, are technically adept staff members at privacy groups, offices of state attorneys general, and journalists

der an open-source licence, which allows programmers to use and modify the code as they see fit for any non-commercial purpose.

XRay is essentially a reverse-engineering machine that models the correlations made by web services. The group's three initial efforts have tried to determine the kinds of ads shown to Gmail users based on the text in their email messages; the product recommendations Amazon shows users based on their wish lists and other data; and the video recommendations made by YouTube determined by the videos users have previously viewed.

The researchers set up accounts and fed them "inputs" like email messages, searches and products viewed. They then observed the "outputs" like ads and product or viewing recommendations. And then they modelled the correlations between the inputs and outputs, so XRay could observe and predict the results of contextual and behavioral targeting by web services. XRay's research is by turns predictable, amusing and unsettling. Take the Gmail message and ad correlation study. In email

messages that suggest the subject of depression, using words including depression, depressed and sad, the ads delivered were more offbeat: "Shamanic healing over the phone" and "Text coach — get the girl you want and desire". Geambasu pointing to the shamanic healing ad, which is correlated with depression, asks if that is a correlation that is used widely. For example, if you click on an ad for shamanic healing in some other context, are you assumed likely to suffer from depression? "That leaked targetting information," Geambasu said, "can potentially be used for all sorts of purposes. It can be used for a very hidden kind of discrimination." That is precisely the concern raised a few months ago by the White House report on big data, which called for limits on how companies use personal data collected online.

With further development, the XRay team hopes to have a more robust, general tool ready within a year. The most likely users, members say, are technically adept staff members at privacy groups, state attorneys general offices, and journalists. NYT NEWS SERVICE

SHORT CUTS

Squids inspire

MY FAMILY AND OTHER ANIMALS