

# User Privacy in Web Search

Josep Domingo-Ferrer

Universitat Rovira i Virgili  
UNESCO Chair in Data Privacy  
Department of Computer Engineering and Mathematics  
Av. Països Catalans 26, E-43007 Tarragona, Catalonia  
`josep.domingo@urv.cat`

Web search engines gather a lot of information on the preferences and interests of users. They actually gather enough information to create detailed user profiles which might enable re-identification of the individuals to which those profiles correspond, *e.g.* thanks to the so-called vanity queries or to linkage of several queries known to have been submitted by the same user. In this way, a broadly used search engine like Google becomes a “big brother” in the purest Orwellian style.

In this talk, a survey will be offered of the solutions which have been proposed to preserve anonymity in web search and to fight profile creation. We will start with Private Information Retrieval (PIR) and we will highlight its lack of practicality. We will then look at some relaxations of PIR, based on standalone defense by the user or on a defense based on a peer-to-peer community in which one user submits queries by other users and viceversa.

Finally, we will sketch a new theory, called coprivacy or co-operative privacy, whose goal is to find out under which conditions the best rational option for a peer-to-peer user is to help other peers in preserving their privacy.

## References

1. Aguilar-Melchor, C., Deswarte, Y.: Trustable relays for anonymous communication. *Transactions on Data Privacy* 2(2), 101–130 (2009)
2. AOL Search Data Scandal (August 2006), [http://en.wikipedia.org/wiki/AOL\\_search\\_data\\_scandal](http://en.wikipedia.org/wiki/AOL_search_data_scandal)
3. Beimel, A., Ishai, Y., Malkin, T.: Reducing the servers’ computation in private information retrieval: Pir with preprocessing. *Journal of Cryptology* 17, 125–151 (2004)
4. Castellà-Roca, J., Viejo, A., Herrera-Joancomartí, J.: Preserving user’s privacy in web search engines. *Computer Communications* 32(13-14), 1541–1551 (2009)
5. Chor, B., Goldreich, O., Kushilevitz, E., Sudan, M.: Private information retrieval. In: *IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 41–50 (1995)
6. Chor, B., Gilboa, N., Naor, M.: Private information retrieval by keywords. Technical Report TR CS0917, Department of Computer Science, Technion (1997)
7. Chor, B., Goldreich, O., Kushilevitz, E., Sudan, M.: Private information retrieval. *Journal of the ACM* 45, 965–981 (1998)

8. Domingo-Ferrer, J.: Coprivacy: towards a theory of sustainable privacy. In: Domingo-Ferrer, J. (ed.) PSD 2010. LNCS, vol. 6344, pp. 258–268. Springer, Heidelberg (2010)
9. Domingo-Ferrer, J., Bras-Amorós, M., Wu, Q., Manjón, J.: User-private information retrieval based on a peer-to-peer community. *Data and Knowledge Engineering* 68(11), 1237–1252 (2009)
10. Domingo-Ferrer, J., Solanas, A., Castellà-Roca, J.:  $h(k)$ -Private information retrieval from privacy-uncooperative queryable databases. *Online Information Review* 33(4), 720–744 (2009)
11. Erola, A., Castellà-Roca, J., Navarro-Arribas, G., Torra, V.: Semantic microaggregation for the anonymization of query logs. In: Domingo-Ferrer, J. (ed.) PSD 2010. LNCS, vol. 6344, pp. 127–137. Springer, Heidelberg (2010)
12. Howe, D.C., Nissenbaum, H.: TrackMeNot: Resisting surveillance in web search. In: Kerr, I., Lucock, C., Steeves, V. (eds.) *Lessons from the Identity Trail: Privacy, Anonymity and Identity in a Networked Society*. Oxford University Press, Oxford (2009), <http://www.mrl.nyu.edu/~dhowe/trackmenot/>
13. Internet World Stats, <http://www.internetworldstats.com/>
14. iProspect Blended Search Results Study (April 2008), [http://www.iprospect.com/premiumPDFs/researchstudy\\_apr2008\\_blendedsearchresults.pdf](http://www.iprospect.com/premiumPDFs/researchstudy_apr2008_blendedsearchresults.pdf)
15. Nash, J.: Non-cooperative games. *Annals of Mathematics* 54, 289–295 (1951)
16. Netcraft June 2010 Web Server Survey, <http://news.netcraft.com/archives/2010/06/16/june-2010-web-server-survey.html>
17. Nisan, N., Roughgarden, T., Tardos, É., Vazirani, V.V. (eds.): *Algorithmic Game Theory*. Cambridge University Press, Cambridge (2007)
18. Open Directory Project, <http://www.dmoz.org>
19. Ogata, W., Kurosawa, K.: Oblivious keyword search. *Journal of Complexity* 20(2–3), 356–371 (2004)
20. Ostrovsky, R., Skeith III, W.E.: A survey of single-database PIR: techniques and applications. In: Okamoto, T., Wang, X. (eds.) PKC 2007. LNCS, vol. 4450, pp. 393–411. Springer, Heidelberg (2007)
21. Reiter, M.K., Rubin, A.D.: Crowds: anonymity for web transactions. *ACM Transactions on Information Systems Security* 1(1), 66–92 (1998)
22. Saint-Jean, F., Johnson, A., Boneh, D., Feigenbaum, J.: Private web search. In: *Proc. of the 2007 ACM Workshop on Privacy in Electronic Society*, pp. 84–90 (2007)
23. Shen, X., Tan, B., Zhai, C.X.: Privacy protection in personalized search. *ACM SIGIR Forum* 41(1), 4–17 (2007)
24. The Tor Project, Inc. “Tor: Overview”, <http://torproject.org/overview.html.en>
25. Torbutton 1.2.5, <https://addons.mozilla.org/ca/firefox/addon/2275/>
26. Viejo, A., Castellà-Roca, J.: Using social networks to distort users’ profiles generated by web search engines. *Computer Networks* (to appear)
27. Princeton University. WordNet: A lexical database for English, <http://wordnet.princeton.edu/>