# Report on the Models of Trust for the Web Workshop (MTW'06)

Tim Finin
University of Maryland, Baltimore County
Maryland MD 21250 USA
finin@umbc.edu

Lalana Kagal
Massachusetts Institute of Technology
Cambridge MA 02139 USA
lkagal@csail.mit.edu

Daniel Olmedilla
L3S Research Center and Hannover University
30539 Hannover Germany
olmedilla@l3s.de

## 1   Introduction

We live in a time when millions of people are adding information to the Web through a growing collection of tools and platforms. Ordinary citizens publish all kinds of content on Web pages, blogs, wikis, podcasts, vlogs, message boards, shared spreadsheets, and new publishing forums that seem to appear almost monthly. As it becomes easier for people to add information to the Web, it is more difficult, and also more important, to distinguish reliable information and sources from those that are not.

Search engines excel at finding results that are relevant to a user's query, but many are outdated, biased, inaccurate, and/or from unreliable sources. Popularity based metrics such as Google's PageRank help, but users are still forced to filter the results to select the most reliable information based on their particular trust requirements. With the introduction of web services, the problem is further exacerbated as users have to come up with a new set of requirements for trusting web services and web services themselves require a more automated way of trusting each other. Apart from inaccurate or outdated information, we also need to anticipate *Semantic Web spam* (SWAM) – where malicious sources publish false facts and scams to deliberately mislead software agents and programs.

The *Models of Trust for the Web* workshop was held in conjunction with the 15th International World Wide Web Conference (WWW2006) on 22nd May, 2006 in Edinburgh, Scotland. The goal of the workshop was to bring together researchers and experts from different communities (e.g., Information Systems, Database, Semantic Web, Web Services) who have been working on topics like trust, provenance, privacy, security, reputation, and spam, in order to understand the challenges associated with facilitating trust on the Web, to deliver a state-of-the-art overview in the area, and to identify guidelines for future research. The workshop built on several related workshops, including the *Workshop on Policy Management for the Web* [1] held at the 2005 World Wide Web conference and *Semantic Web and Policy Workshop* [2] held in conjunction with 2005 International Semantic Web Conference.

The MTW'06 workshop was attended by over thirty researchers for a full day of presentations, panels and spirited discussions. The eleven papers that were presented covered a wide spectrum of topics from inferring trust, to using trust to prevent spam, and the role of social networks in calculating trust [3].

## 2   Presented Papers

The keynote speaker, Ricardo Baeza-Yates (Yahoo! Research), discussed how social networks can be exploited to provide social and economic deterrents for spamming. There are several kinds of spam that need to be monitored: scraper scam that copies good data from other sites and adds monetization, synthetic text that provides boilerplate text built around key phrases, query-targeted spam in which each page targets a single tail query, DNS spam where many domains use the same servers, and blog spam. Using Flickr as an example, Ricardo showed how the "wisdom of crowds can be used to search" as Flickr users collaboratively search and tag each other's photos and the anchor text is collective knowledge used to create a search. At Yahoo!, spam is detected and characterized using a combination of algorithmic and editorial techniques in order to prevent it from distorting the rank-

ing of web pages.

## 2.1 Session I : Trust Networks

David Brondsema and Andrew Schamp described their work in using social trust networks to filter spam in "Konfidi: Trust Networks Using PGP and RDF". They proposed that spam can be filtered by reasoning over trust relationships in RDF. These relationships include who (both identity and public key) is trusted, value of trust, and with respect to what topic.

In "Using Trust and Provenance for Content Filtering on the Semantic Web", Jennifer Golbeck and Aaron Mannes showed that annotating relationships with binary trust values in web-based social networks allowed trust values to be inferred between unrelated entities. They discussed the FilmTrust project, which is a movie recommendation system developed using their approach.

Patricia Victor et al. defined their billatice trust model that takes trust, distrust, lack of data, and contradictory data into consideration while calculating trust in "Towards a Provenance-Preserving Trust Model in Agent Networks".

## 2.2 Session II : Inferring Trust

The paper "Propagating Trust and Distrust to Demote Web Spam" by Baoning Wu, Vinay Goel, and Brian Davison addressed the problem of web spam also known as search engine spam in which a target page gets undeserved ranking. They described different methods that a parent page can use to divide its trust or distrust among its child pages. They also defined mechanisms for calculating trust (using outgoing links) and distrust (using incoming links) - accumulation, maximum share, and maximum parent.

L. Jean Camp, Cathleen McGrath, and Alla Genkina approached human trust behavior from a social science perspective. They described their results in "Security and Morality: A Tale of User Deceit" in which they present how users "consider failures in benevolence more serious than failures in competence".

Deborah McGuinness et al. reported in "Investigations into Trust for Collaborative Information Repositories: A Wikipedia Case Study" that both provenance of information and revision details are required to improve the trustworthiness of collaborative information systems such as Wikipedia. They discussed citation-based trust, which is derived from citation relationships among articles, and revision-based trust, which is derived from the original

article, revision operators, and revision authors, as mechanisms for inferring trustworthiness of a Wikipedia article. They also presented a mockup of a Wikipedia version marked up with trust.

## 2.3 Session III : Trust Models

Santtu Toivonen, Gabriele Lenzini, and Ilkka Uusitalo explored the role of context in trust determination in their paper "Context-aware Trust Evaluation Functions for Dynamic Reconfigurable Systems". They distinguished between quality attributes, which are static attributes of the trustee and context attributes, which are optional attributes that can change dynamically such as location, and device type. They discussed how context-aware trust is calculated from quality attributes, reputation, and recommendations within a certain trust scope at a certain time.

In "How Certain is Recommended Trust-Information", Uwe Roth and Volker Fusenig suggested that trust information given by a recommender may not be reliable and could negatively affect the trust decision. They proposed a strategy for making trust decisions based on a converting a network of relations of direct and recommended trust information into a decision tree by choosing the path as well as whether to trust the recommended information along the path at random.

## 2.4 Session IV : Trust in Applications

The paper titled "Quality Labeling of Web Content: The Quatro approach" by Vangelis Karkaletsis et al. reports on a common machine-readable vocabulary for labelling web content that will be represented by user friendly icons for ease of understanding. The vocabulary includes several kinds of labels: page-specific such as whether the page uses clear language, whether it includes a privacy statement, content provider specific such as whether his credentials have been verified, business-specific such as whether it complies with rules and regulations of e-business, and label-specific such as when the label was issued, and when it was last viewed.

Ing-Xiang Chen and Cheng-Zen Yang examined biased search engine results and their deviations in "A Study of Web Search Engine Bias and its Assessment". An example of search engine bias is results in China for the keyword "Falun Gong". The authors proposed a two-dimensional scheme by adopting both indexical bias (differences in the sets of URLs retrieved) and content bias (deviations of content).

Alex Tsow suggested that users are prone to attacks from malicious software embedded on their routers in

2

"Phishing with Consumer Electronics - Malicious Home Routers" and hinted that trust is not a software only matter but there is an implicit trust in hardware vendors.

# 3 Future Directions / Open Research Issues

The *Models of Trust for the Web* workshop helped to understand current state-of-the-art research and to provide a discussion forum for researchers working on trust issues. It highlighted the importance of existing lines of research and brought up some salient emerging problems. Some open questions and problems include:

- *Trust modeling:* Is trust just boolean or does it have some certainty associated? Is it transitive? May we infer that the enemy of my enemy is my friend? Is it context or time dependent?

- *Trust awareness:* Users' trust on computers highly depends on previous experiences. They are initially too trustful and typically unaware of the risks associated with their computer usage. Improving user interfaces and increasing user awareness on privacy and security issues remains an open issue.

- *Trustworthy information:* Whether information retrieved from unknown sources is correct or not is a major question in current Web. Furthermore, SWAM makes this task even more difficult.

- *Database Security:* Trust in data and systems over time (compliance storage) is currently a major issue (specially after latest laws in which companies need to store data for longer time). The current winner of the best paper award at the Very Large Database Conference "Trustworthy Keyword Search for Regulatory-Compliant Record Retention" [4] demonstrates the importance of this line of research.

- *Access control & Trust Management:* Specification of conditions under which service access is granted or private information is released in an open distributed world has been and is still one of the most important lines of research. Work on policies is currently a hot topic in areas such as the Semantic Web [2, 5] for example.

# 4 Conclusion

The Web and its evolving infrastructure have made it easy to access virtually all of the world's knowledge and is the first source to which most of us turn when we need to know something. Search engines and other tools have focused on finding information *relevant* to users' queries with results ranked at best by their popularity. As it becomes easier to publish information on the Web, it is increasingly important to develop good frameworks for evaluating the trustworthiness of the information found. The papers from the WWW'06 *Models of Trust for the Web* workshop addressed these issues, identified important issues, and offered some partial solutions.

# References

[1] Lalana Kagal, Tim Finin, and James Hendler, editors. *Proceedings of the WWW'05 Workshop on Policy Management for the Web*, Chiba JP, May 2005. http://ebiquity.umbc.edu/paper/html/id/221/.

[2] Lalana Kagal, Tim Finin, and James Hendler, editors. *Proceedings of the ISWC 2005 Semantic Web and Policy Workshop*, Galway IE, November 2005. http://ebiquity.umbc.edu/paper/html/id/268/.

[3] Tim Finin, Lalana Kagal, and Daniel Olmedilla, editors. *Proceedings of the WWW'06 Workshop on Models of Trust for the Web*, volume 190 of *CEUR Workshop Proceedings*, Edinburgh, Scotland, May 2006. CEUR-WS.org. http://ceur-ws.org/Vol-190.

[4] Soumyadeb Mitra, Windsor W. Hsu, and Marianne Winslett. Trustworthy keyword search for regulatory-compliant record retention. In *Proceedings of the 32nd International Conference on Very Large Data Bases*, pages 1001–1012, Seoul, Korea, Sep 2006.

[5] Piero A. Bonatti, Li Ding, Tim Finin, and Daniel Olmedilla, editors. *Proceedings of the ISWC'06 2nd International Semantic Web Policy Workshop (SWPW)*, Athens GA USA, November 2006.

3