

Report of the 4th Workshop on Information Credibility on the Web (WICOW 2010)

Adam Jatowt and Katsumi Tanaka

Graduate School of Informatics, Kyoto University

Yoshida-honmachi, Sakyo-ku, Kyoto, Japan

{adam, tanaka}@dl.kuis.kyoto-u.ac.jp

Workshop Website: <http://www.dl.kuis.kyoto-u.ac.jp/wicow4>

Abstract

The 4th Workshop on Information Credibility on the Web (WICOW 2010) was held in conjunction with the WWW 2010 conference in Raleigh, NC, USA on the 27th April, 2010. Ten full papers and a keynote speech were presented in four sessions. This report briefly outlines the main outcomes of the workshop.

1 Introduction

Credibility and quality of online information are key factors for guaranteeing high user satisfaction when relying on Web information in their daily lives. However, with the advent of Web 2.0 and its nearly unrestricted models of content generation it has become difficult for users to correctly evaluate the extent to which they could trust encountered information. Therefore it is important to continue research aiming at offering tools for assisting users in this process. Clearly, solutions for separating credible content from incorrect, biased or obsolete information as well as finding trustworthy information sources on the Web should be very useful. Such approaches should be based on the fundamental research into the way in which humans find, compare, connect, aggregate and evaluate information in both real and digital worlds.

The 4th Workshop on the Information Credibility on the Web was held on April 27th in Raleigh, NC, USA as a part of the 19th International World Wide Web Conferenceⁱ. The aim of WICOW workshop is to provide a platform for exchanging novel ideas and research outcomes as well as to promote discussions on multiple aspects of information credibility among both researchers and practitioners. The previous workshops were organized in conjunction with the 21st Annual Conference of The Japanese Society for Artificial Intelligenceⁱⁱ (JSAI 2007) in 2007 at Miyazaki, Japan, in conjunction with the ACM 17th Conference on Information and Knowledge Managementⁱⁱⁱ (CIKM 2008) in 2008 at Napa Valley, USA and together with the 18th World Wide Web Conference^{iv} (WWW 2009) in 2008 in Madrid, Spain.

In response to the call for papers 22 full paper submissions were received, out of which, 10 papers have been selected following a careful review process with at least three reviews for each submission. The acceptance rate of the workshop was thus 45%. The accepted papers were grouped into three sessions: *Wikipedia Credibility*, *Studies of Web Information Credibility* and *Evaluating Information Credibility*. This year we had presentations concerned with diverse aspects of information credibility on the Web related to varying documents types and genres. For example, several presentations

demonstrated methods for evaluating credibility and quality of Wikipedia articles; others aimed at assessing the process of selecting news articles, estimation of the quality of online videos or automatically finding disputed claims in Web pages. Other papers demonstrated the results of studies on the adversarial behavior in online advertisements or link generation, the overlap between online and real-world identities of users and a novel method for capturing user trust evaluations of Web content.

In addition, we had a pleasure to invite *Miriam Metzger* from the University of California, Santa Barbara as a guest speaker delivering a keynote talk entitled: *Understanding Credibility across Disciplinary Boundaries*. In her talk, Miriam called for more integration between different disciplines towards achieving the common goal of assisting users with the evaluation of credibility of Web information. Researchers from social information, computer science, library studies, psychology, education and communication areas could benefit more from the exchange of information on both the theoretical and methodological standpoints. For example, it would be interesting to study the diversity in fundamental views and approaches to online information credibility across different disciplines.

2 Paper Presentations

This section briefly introduces all the workshop papers.

2.1 Wikipedia Credibility

The session started with the presentation of a paper entitled: “*Detecting Wikipedia Vandalism with Active Learning and Statistical Language Models*” by *Si-Chi Chin, W. Nick Street, Padmini Srinivasan* and *David Eichmann*. The authors introduce an active learning approach using statistical language model for categorizing vandalism in Wikipedia into seven major types. The statistical language models are built from the distributions of words in revision history of Wikipedia articles. An active learning model approach is used for solving the problem of noisy and incomplete labeling of vandalism in Wikipedia.

Next, was the presentation of a paper entitled: “*On Measuring the Quality of Wikipedia Articles*” by *Gabriel de la Calzada* and *Alex Dekhtyar*. The authors propose applying different information quality assessment models for different categories of Wikipedia articles such as “stabilized” and “controversial” article categories. For each category, first a classifier is developed to assign an article to a given broad class. Then the article is evaluated by a quality model corresponding to the particular class. The results of the quality estimation were then evaluated by a large number of human participants.

In the last paper presented at this session entitled: “*Trust in Wikipedia: How Users Trust Information from an Unknown Source*”, *Teun Lucassen* and *Jan Maarten Schraagen* describe the results of the study of the features that Wikipedia users consider when evaluating the trustworthiness of encountered articles. By presenting 15 university students with a set of Wikipedia articles with varying quality levels and degrees of topic familiarity, the authors found that textual features, references and images play the major role in the process of credibility estimation.

2.2 Studies of Web Information Credibility

The next session started with the presentation of a paper: “*Smart Marketing or Bait & Switch? Competitors' Brands as Keywords in Online Advertising*” by *Mark Rosso* and *Bernard Jansen*. The

authors investigate the occurrence of piggybacking in online advertisements. *Bait & Switch* means an advertising scenario in which consumer searching with a query composed of a brand name of one company is presented with the advertisements of its competitors. Through the study of advertisements returned for 100 queries consisting of top brand names the authors found that competitive piggybacking is not a widespread phenomenon.

The paper entitled: “*The Credibility of Digital Identity Information on the Social Web: a User Study*” by *Matthew Rowe* describes the study of the credibility associated with digital identity information of users in social networks services. The author compares characteristics of online identity representations such as the relationships and the frequency of human interactions with the ones in the real world. The results indicate that social Web users construct digital identity representations that generally mirror their real-world identities.

The last paper presented at this session was “*TIME: a Method of Detecting the Dynamic Variances of Trust*” by *Laurian Vega, Yeong-Tay Sun, Scott McCrickard and Steve Harrison*. The authors investigate a dynamic approach for trust evaluation called Trust Incremental Measure Evaluation (TIME). It consists of repeated measurements of trust levels that users assign to particular elements of pages in Web sites. The paper provides also review of other trust evaluation methods.

2.3 Evaluating Information Credibility

The presentation of the paper entitled: “*Identifying Spam Link Generators for Monitoring Emerging Web Spam*” by *Young-joo Chung, Masashi Toyoda and Masaru Kitsuregawa* opened the last session of the workshop. The authors propose classification-based approach for identifying hosts which are spam link generators. The features used for classification comprise link-based features including modified PageRank scores based on white and spam seeds and the scores of neighbouring hosts. The experiments conducted on a three years old Japanese Web archive demonstrate nearly 56% to 73% precision rate and F-measure from 0.54 to 0.68.

Thomas Largillier, Guillaume Peyronnet and Sylvain Peyronnet authored the paper: “*SpotRank: a Robust Voting System for Social News Websites*”. It describes a robust voting system for news articles whose objective is to demote the effect of vote manipulation by malicious users. The system uses statistical filters, a collusion detection mechanism and the computation of user and news pertinence. The authors present the system implementation and demonstrate its advantage over similar competitor systems through statistical analysis and user study.

The paper entitled: “*What is Disputed on the Web?*” by *Rob Ennals, Dan Byler, John Mark Agosta and Barbara Rosario* describes a method for automatically acquiring a corpus of disputed claims from Web pages. Disputed claims are controversial statements about which there is no consensus on the Web. Such claims are harvested using patterns such as “falsely claimed that X” and employing classifier for selecting text that composes the claims. The authors presented also the companion paper at the main conference which describes *Dispute Finder* system. It indicates disputed claims in browsed pages based on the constructed corpus.

The presentation of the paper entitled: “*Modulating Video Credibility via Visualization of Quality Evaluations*” by *Nicholas Diakopoulos and Irfan Essa* concluded the workshop. The talk was concerned with a system which presents visual clues for viewers of informational videos in order to convey aggregated evaluations of multiple users. Such clues consist of user activity levels as well as the polarity of user-generated annotations.

3 Acknowledgements

We convey our sincere thanks to *Xiaofang Zhou* and *Min Zhang*, the other co-organizers of WICOW 2010, and to the Program Committee members for their effort in reviewing the submitted papers. We would also like to thank the WWW 2010 organizers for letting us to organize this workshop. We express our gratitude to the sponsors of WICOW 2010, National Institute of Information and Communications Technology and Kyoto University Global COE Program: Informatics Education and Research for Knowledge-Circulating Society.

ⁱ The 19th International World Wide Web Conference (WWW 2010): <http://www2010.org/>

ⁱⁱ The 21st Annual Conference of The Japanese Society for Artificial Intelligence (JSAI 2007): http://www.ai-gakkai.or.jp/jsai/conf/2007/index_en.html

ⁱⁱⁱ The ACM 17th Conference on Information and Knowledge Management (CIKM 2008): <http://www.cikm2008.org/>

^{iv} The 18th World Wide Web Conference (WWW 2009): <http://www2009.org/>