The 3rd Workshop on Information Credibility on the Web

(WICOW 2009)

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/wicow3

Abstract

On the 20th April, 2009 the 3rd Workshop on Information Credibility on the Web (WICOW 2009) was held as part of WWW 2009 conference in Madrid, Spain. Nine full papers and a keynote speech were presented in four sessions. This report briefly outlines the main outcomes of the workshop.

1 Introduction

Research aimed at evaluating web content credibility is recently becoming increasingly necessary since when the Web started to influence our daily lives. The abundance of content on the Web, the lack of publishing barriers and poor quality control of Web content raise credibility issues. Although the problem of establishing credibility in online environments have been recognized before by information and digital library community, the main efforts went into understanding the factors that users consider when evaluating information credibility as well as into educating users on how to successfully verify online information. However, despite the presence of ready guidelines many users are unprepared, do not possess sufficient skills or are merely unaware of credibility issue on the Web and thus cannot properly assess credibility of online information. This is partly because, in contrast to traditional publishing, the content on the Web often lacks quality indicators such as author names and their credentials, references to external, trustworthy information sources, information age and provenance and so on. Consequently, few users perform rigorous evaluation of the quality of obtained information. Therefore, effective tools for supporting users in the judgment of the credibility of Web content should be helpful and should improve user satisfaction when using the Web.

The 3rd Workshop on the Information Credibility on the Web was held on April 20th in Madrid, Spain as a part of the 18th International World Wide Web Conference[i] and was attended by about 40 participants. The aim of this workshop was 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. We have chosen World Wide Web Conference as a venue of the workshop due to the diversity of aspects and potential approaches concerned with the problem of information credibility on the Web. The multi-dimensional character of Web credibility requires the combination of solutions coming from different fields and geared towards different media types towards the common goal of improving the quality of information on the Web and towards providing users with effective tools for supporting credibility estimation of Web content.

We encouraged submissions concerned with the following topics:

- Information credibility evaluation and its applications
- Web content analysis for credibility evaluation
- Author's intent detection
- Credibility of Web search results
- Search models and applications for trustworthy content on the Web
- Conflicting opinion detection
- Online media and news credibility
- Multimedia content credibility
- Credibility evaluation of user-generated content (e.g., Wikipedia, question answering sites)
- Information credibility evaluation in social networks and Web 2.0 applications
- Analysis of information dissemination on the Web (e.g., in blogosphere)
- Spatial and temporal aspects in information credibility on the Web
- Information credibility theory and fundamentals
- Estimation of information age, provenance and validity
- Estimation of author's and publisher's reputation
- Sociological and psychological aspects of information credibility estimation
- Users study for information credibility evaluation
- Persuasive technologies
- Information credibility in online advertising and Internet monetization
- Web spam detection
- Information quality evaluation and visualization
- Data consistency and provenance
- Processing uncertain data and information

In response to the call for papers 19 submissions were received, out of which, 9 papers have been selected following a careful review process with at least three reviews for each submission. The accepted papers were grouped into four sessions: *Evaluating Credibility of Digital Resources*, *Quality of Web 2.0 Content*, *Information Aggregation and Comparison*, and *Vendor and Product Reputation*. In addition, we had a pleasure to invite *Ricardo Baeza-Yates* from Yahoo! Barcelona as a guest speaker delivering a keynote talk entitled: *User Generated Content: How Good it is?*

2   Paper Presentations
This section briefly introduces workshop papers arranged according to their sessions.

2.1   Evaluating Credibility of Digital Resources

The first session started with the presentation of a paper entitled: *Automatically Assessing Resource Quality for Educational Digital Libraries* by *P. Wetzler, S. Bethard, K. Butcher, J. H. Martin* and *T. Sumner*. The authors demonstrate machine learning-based approach to automatically assessing quality of resources in educational digital libraries. They have distinguished a set of key dimensions and indicators concerned with resource quality that educators often use. These indicators were later examined by digital library quality experts in order to develop a set of annotation guidelines for a corpus of digital resources.

Next was the presentation of a paper entitled: *A Method of Analyzing Credibility based on LOD Control of Digital Maps* by *D. Kitayama, R. Lee* and *K. Sumiya*. The authors approach the problem of

the credibility of digital maps. Their solution relies on analyzing the consistency of displayed objects in a given level of details (LOD) of a digital map. For a map to be credible, equivalent objects should be displayed together in the same scale. The authors propose a method for calculating the equivalence of geographic objects in a digital map by taking into consideration their LOD controls and the size of objects' territories.

In the last paper presented at this session entitled: *Data Quality in Web Archiving*, *M. Spaniol, D. Denev, A. Mazeika, P. Senellart* and *G. Weikum* introduce model for measuring the data quality in Web archives by identifying coherent sections in Web sites. The problem is important since content or links archived from large Web sites may become incoherent due to the politeness requirements during crawling. The authors propose also a novel crawling strategy aiming at ensuring the optimal coherence of Web archives. The preliminary experiments demonstrate the usefulness and effectiveness of the proposed strategy

2.2   Quality of Web 2.0 Content

This session started with the presentation of a paper: *QuWi: Quality Control in Wikipedia* by *A. Cusinato, V. Della Mea, F. Di Salvatore* and *S. Mizzaro*. The authors adapt the algorithm for quality control in scholarly publishing to the case of the Wikipedia by analyzing the revision history of articles with an emphasis on content steadiness. They manage to successfully identify high and low quality Wikipedia articles and demonstrate that good quality authors usually produce more long-lived contributions than the low quality authors.

Next, *Ricardo Baeza-Yates* gave a keynote talk entitled: *User Generated Content: How Good it is?* The talk introduced some of the recent activities concerned with analyzing the quality of consumer generated content. Through this survey encouraging results were demonstrated that indicated better than commonly expected quality of user generated content in popular Web 2.0 sites such as Flickr, Yahoo! Answers and Wikipedia. These conclusions provide a room for many interesting applications such as question answering or automatic generation of semantic resources.

2.3   Information Aggregation and Comparison

The presentation of the paper entitled: *The Effects of Source Credibility Ratings in a Cultural Heritage Information Aggregator* by *A. Amin, J. Zhang, H. Cramer, L. Hardman* and *V. Evers* opened the third session of the workshop. The authors investigate whether source credibility rankings have any effect on users' search and selection performance of multiple cultural heritage sources. The study assumes different scenarios in which there are many or few different information sources. Showing the credibility estimates of sources results in higher users' confidence of selection, although it does not always decrease the search time.

The paper entitled: Statement Map: A*ssisting Information Credibility Analysis by Visualizing Arguments* by *K. Murakami, E. Nichols, S. Matsuyoshi, A. Sumida, S. Masuda, K. Inui* and *Y. Matsumoto* describes a project called Statement Map for providing variety of viewpoints together with their supporting evidence to a given topic of user interest. The generated output information is mined from the Web by applying a range of NLP techniques on both Japanese and English content for detecting various semantic relations such as agreement, conflict, and evidence as well as for conducting sentiment and factual analysis.

A. Juffinger, M. Granitzer and *E. Lex* authored the paper: *Blog Credibility Ranking by Exploiting Verified Content*. It introduces a system for estimating the credibility of weblogs by comparing their

similarity to a reference news corpus. Generally, weblogs which timely track news events are deemed to be more credible than the ones that do not do so. The authors use dynamic time warping for alignment of blogs and news. In addition, the proposed system visualizes trends of topics in blogs with their distribution according to various languages.

2.4   Vendor and Product Reputation

The paper entitled: *Seller's Credibility in Electronic Markets: A Complex Network based Approach* by *A. Pereira, A. Silva, W. Meira Jr.* and *V. Almeida* began the last session of the workshop. The authors address the problem of estimating the trust levels of sellers in electronic marketplaces using complex network analysis. The network is built upon the structure of interactions between market participants. The proposed method achieves more than 80% accuracy and is meant to be a support decision mechanism for buyers in electronic markets.

The paper entitled: *Evaluating Brand Value on the Web* by *T. Kobayashi, H. Ohshima, S. Oyama* and *K. Tanaka* concluded the workshop. It introduces a method for automatically estimating brand value on the Web and for judging whether the brand names mentioned in Web pages represent products that are actually popular and reputable. The proposed method is based on discovering useful attributes of brand names using HITS algorithm and achieves 74%-85% accuracy rate.

3   Acknowledgements

Finally, we list the names and affiliations of Program Committee members:

Witold Abramowicz (Poznan University of Economics, Poland)
Sourav S Bhowmick (Nanyang Technological University, Singapore)
Yunbo Cao (Microsoft Research Asia, China)
James Caverlee (Texas A&M University, USA)
David Danielson (Stanford University, USA)
Jean-Yves Delort (Macquarie University, Australia)
Ke Deng (University of Queensland, Australia)
Pavel Dmitriev (Yahoo!, USA)
Rino Falcone (CNR, Italy)
Zhigang Hua (Yahoo!, USA)
Marta Indulska (University of Queensland, Australia)
Kentaro Inui (NAIST, Japan)
Daxin Jiang (Microsoft Research Asia, China)
Yoshikiyo Kato (NICT, Japan)
Yutaka Kidawara (NICT, Japan)
Nick Koudas (University of Toronto, Canada)
Marek Kowalkiewicz (SAP Research, Australia)
Sadao Kurohashi (Kyoto University, Japan)
Chen Li (UC Irvine, USA)

Ee-Peng Lim (Singapore Management University, Singapore)
Li Ma (IBM Research, China)
Yutaka Matsuo (University of Tokyo, Japan)
Martin Memmel (DFKI, Germany)
Miriam Metzger (UCSB, USA)
Sudha Ram (University of Arizona, USA)
Soo Young Rieh (University of Michigan, USA)
Shazia Sadiq (University of Queensland, Australia)
Kazutoshi Sumiya (University of Hyogo, Japan)
Wei Wang (University of New South Wales, Australia)
Martin Wolpers (Fraunhofer FIT, Germany)
Xiaochun Yang (Northeastern University, China)
Masatoshi Yoshikawa (Kyoto University, Japan)

---

[i] The 18th International World Wide Web Conference (WWW 2009): http://www2009.org/