

Introduction to the special issue of the World Wide Web journal on “Social Media Preservation and Applications”

Alexandra I. Cristea · Dimitrios Katsaros ·
Yannis Manolopoulos

Received: 10 February 2014 /
Accepted: 13 February 2014 / Published online: 1 March 2014
© Springer Science+Business Media New York 2014

Social media have become firmly established as a new publishing and communication tool with significant value for all the aspects of contemporary society, culture and science. Social media feature a dynamic and continuously evolving nature, fostering the creation and exchange of concepts and ideas in an unprecedented scale.

The leading role of social media coupled with the inherent volatility and the ephemeral nature of this medium, compared with other forms of electronic communication, has rendered social media preservation as necessary. It is critical to preserve social media as they are an essential part of our heritage and they can prove valuable for current and future generations.

Social media preservation is an advanced topic of digital preservation, encompassing the capture, management, preservation and availability of social media for future use and research. To achieve this goal, we need to advance the state of the art in a wide range of diverse topics such as social media modeling and analysis, classification, information retrieval, content migration, replication, emulation and web preservation strategies. Furthermore, we need to focus not only on theoretical advances but also on innovative applications and systems. In this special issue, we are aiming to exchange the latest fundamental advances in the state of the art and practice of social media preservation and important related areas.

In this issue we have included 7 papers, which have been selected out of 17 papers that were submitted in response to our call. Some information about these papers follows.

The paper by L. Akritidis and P. Bozanis entitled “Improving Opinionated Blog Retrieval Effectiveness with Quality Measures and Temporal Features” deals with the problem of retrieval of blog entries, which contain opinions about entries by paying special attention to the importance of such entries. The proposed ranking method takes into account the influence of the blogger who authored an opinion, the reputation of the specific blog site and the impact of the blog post itself. Experiments with TREC Blogs08 dataset shows enhancement of retrieval precision.

The paper by G. Gkotsis, K. Stepanyan, A. Cristea and M. Joy entitled “Entropy-based Automated Wrapper Generation for Web Data Extraction” focus on the process of weblog data extraction. The approach includes a model for generating a wrapper that exploits web feeds for deriving probabilistically (based on entropy) a set of extraction rules in an automatic way. The

A. I. Cristea · D. Katsaros · Y. Manolopoulos (✉)
Aristotle University, Thessaloniki, Greece
e-mail: manolopo@csd.auth.gr

evaluation of the method shows significant prediction accuracy and thus can be used for robust extraction of weblog properties. Successful extraction of weblog properties into structured data can be useful in not only the analysis and search, but also in the preservation of such resources.

N. Kasioumis, V Banos and H. Kalb have co-authored the paper under the title “Towards Building a Blog Preservation Platform”. As mentioned in the title, the authors are describing a general architecture for a platform to enable blog preservation. In particular, several issues are examined to make such a platform as efficient, effective and complete as possible. Issued related to user, interoperability and performance requirements, to the blog spider, and finally to the digital repository (including metadata, storage and indices) have been investigated.

The paper by P. Korenek and M. Simko presents a method to extract prevailing sentiments and opinions from microblogs, such as Twitter, which may provide preservation tools the appropriate knowledge of what information is essential to be preserved for future generations—a discussion that is ongoing between specialists. Their type of information may be filling gaps in user models, which could further be preserved to capture the specific sentiments on a given topic.

The paper by N. Pellas and I. Kazanidis is indirectly related to topic of the special issue. That is, under the framework of Second Life, they examine the impact of computer self-efficacy, situational interest and academic self-concept in virtual communities in inquiry during distance learning procedures. The purpose of the paper is to optimize the benefits of the use of this medium as a learning one, by taking into account students’ interactions.

Y. Rao, X. Quan, L. Wenying, Q. Li and M Chen have co-authored the paper entitled “Building Emotional Dictionary for Sentiment Analysis of Online News”, which reports on an applications for blogs and microblogs. In particular, they propose an algorithm to automatically build a world-level emotional dictionary for social emotion detection. The method is validated by experiments with real-world data. The generated dictionary can be used to predict the emotional distribution of news articles, and is related to the paper by Korenek and Simko both in area as well as application to future preservation of essential features of social media in general.

Finally, the paper by P. Symeonidis and E. Tiakas elaborates on on-line social networks and is entitled “Transitive Node Similarity: Predicting and Recommending Links in Signed Social Networks”. The paper proposed a new technique to provide recommendations in a social network by extending previous approach that take into account only local features of the network. The proposed algorithm outperforms previous approaches in terms of accuracy and efficiency. In large networks, where long term digital preservation of social networks is necessary, such algorithms are more scalable over time.

Finally, the guest editors would like to thank the reviewers, which contributed to come up with this issue. Thus, thank are due to:

N. Agarwal (University of Arkansas at Little Rock)

V. Bamos (Aristotle University of Thessaloniki)

S. Bao (IBM Watson Research Center)

P. Basile (University of Bari Aldo Moro)

D. Bell (University of Leeds)

Y.C. Chen (National Chiao Tung University)

J. Foss (University of Warwick)

F. Frasincar (Erasmus School of Economics)

M. Freire (Universidad Autonoma de Madrid)

A. Giokas (University of Warwick)

G. Gkotsis (University of Warwick)

C. Hava Muntean (Dublin City University)
M. Hedstrom (University of Michigan)
M. Hendrix (University of Coventry)
B. Karagiannidis (University of Thessaly)
F. Kokkoras (TEI of Thessaly)
K. Kotropoulos (Aristotle University of Thessaloniki)
V. Koutsonikola (Aristotle University of Thessaloniki)
I. Liccardi (University of Paris-Sud)
F. Luo (Clemson University)
S. Moebs (Dublin City University)
A. Molnar (National College Ireland)
P. Mylonas (National Technical University of Athens)
A. Nanopoulos (University of Eichstätt-Ingolstadt)
G. Pallis (University of Cyprus)
S. Papadopoulos (Institute of Telematics and Informatics, Thessaloniki)
E. Petrakis (Technical University of Crete)
D. Qudah (University of Warwick)
F. Schweitzer (ETH Zurich)
J. Scotton (University of Warwick)
A. Sidiropoulos (TEI of Thessaloniki)
D. Spiliotopoulos (National and Kapodistrian University of Athens)
K. Stepanyan (University of Warwick)
P. Symeonidis (Aristotle University of Thessaloniki)
J. Tank (Chinghua University)
T. Tsiatsos (Aristotle University of Thessaloniki)
M. Tzagarakis (University of Patras)
A. Vakali (Aristotle University of Thessaloniki)
I. Varlamis (Harokopio University)