

USEWOD2011 - 1st International Workshop on Usage Analysis and the Web of Data

Bettina Berendt
Katholieke Universiteit Leuven
Dept. of Computer Science
Heverlee, Belgium
bettina.berendt@cs.
kuleuven.be

Laura Hollink
Delft University of Technology
Web Information Systems
Group
Delft, The Netherlands
l.hollink@tudelft.nl

Vera Hollink
Centre for Mathematics and
Computer Science
Interactive Information Access
Group
Amsterdam, The Netherlands
V.Hollink@cwi.nl

Markus Luczak-Rösch
Freie Universität Berlin
Networked Information
Systems Workgroup
Berlin, Germany
luczak@inf.fu-berlin.de

Knud Möller
National University of Ireland
Digital Enterprise Research
Institute
Galway, Ireland
knud.moeller@deri.org

David Vallet
Universidad Autónoma de
Madrid
Escuela Politécnica Superior
Madrid, Spain
david.vallet@uam.es

ABSTRACT

The USEWOD2011 workshop investigates combinations of usage data with semantics and the web of data. The analysis of usage data may be enhanced using semantic information. Now that more and more explicit knowledge is represented on the Web, the question arises how these semantics can be used to aid large scale web usage analysis and mining.

Conversely, usage data analysis can enhance semantic resources as well as Semantic Web applications. Traces of users can be used to evaluate, adapt or personalize Semantic Web applications. Also, new ways of accessing information enabled by the Web of Data imply the need to develop or adapt algorithms, methods, and techniques to analyze and interpret the usage of Web data instead of Web pages.

The USEWOD2011 program includes a challenge to the workshop participants: three months before the workshop two datasets consisting of server log files of Linked Open Data sources were released. Participants are invited to come up with interesting analyses, applications, alignments, etc. for these datasets.

Categories and Subject Descriptors

H3.3 [Information Search and Retrieval]; H2.8 [Database applications]: Data mining; H1.2 [User/Machine Systems]: Human factors

General Terms

Experimentation, Design, Algorithms, Human Factors, Performance, Theory

Keywords

Usage data, web of data, data mining

1. INTRODUCTION

The purpose of the USEWOD2011 workshop is to investigate the synergy between semantics and semantic-web technology on the one hand and analysis and mining of usage data on the other hand. The two fields are a promising combination. First, semantics can be used to enhance the analysis of usage data. Usage logs contain information that can help to better understand users or to adapt a system to a user's needs and preferences. Now that more and more explicit knowledge is represented on the Web, in the form of ontologies, folksonomies, or linked data, the question arises how these semantics can be used to aid large scale web usage analysis and mining.

Second, usage data analysis can enhance semantic resources as well as Semantic Web applications. Traces of users can be used to evaluate, adapt or personalize Semantic Web applications. Since logs record real-life users, they provide an opportunity to create gold standards for search or recommendation tools. In addition, logs can form valuable resources from which knowledge (e.g., in the form of ontologies or thesauri) can be extracted bottom-up. Also, the emerging Web of Data demands a re-evaluation of existing usage mining techniques; new ways of accessing information enabled by the Web of Data imply the need to develop or adapt algorithms, methods, and techniques to analyze and interpret the usage of Web data instead of Web pages. An important question at this time is how the Web of Data is being used: how are datasets being accessed by human users and how by machines, what kinds of queries are being performed, and what can we learn about the usage of semantic applications? Ultimately only understanding of their usage (or non-usage) can give both summative and formative evaluations of the adequacy of resources for their final destination: their use in information processing, whether by people or by machines.

The primary goals of this workshop are to foment a new community of researchers from various fields sharing an interest in usage mining and semantics and to create a roadmap for future research in this direction. Several recent papers indicate an interest in bringing usage log analysis and se-

mantic technologies together [1, 2, 3, 4, 5, 6, 7, 8, 9], but the research communities have so far remained isolated. USEWOD2011 forms an ideal opportunity for cross-fertilisation between the different communities and research fields. The workshop is of great interest to researchers from the Semantic Web community who are working with usage data. It is also relevant for people working on log analysis and data mining from the perspective of Information Science, Human Computer Interaction and User Modeling. Also, in the Information Retrieval community there is a growing interest in both usage data and lightweight semantics, such as Linked Data or folksonomies. Finally, we encourage the participation of people from industry, as they are the data providers and therefore will ultimately benefit from the outcomes of the workshop.

2. USAGE DATA CHALLENGE

USEWOD2011 includes a data challenge to stimulate the exchange of ideas and methods between the workshop participants and to relieve some of the difficulties of obtaining real-life usage data. Three months before the workshop we released two datasets of server log files of Linked Open Data sources: DBpedia (<http://dbpedia.org>) and Semantic Web Dog Food (<http://data.semanticweb.org>). The DBpedia logs consist of several months of log data from the linked data twin of Wikipedia, one of the focal points of the Web of data. Semantic Web Dog Food is a constantly growing dataset of publications, people and organisations in the Web and Semantic Web area, covering several of the major conferences and workshops, including WWW, ISWC and ESWC. These datasets represent an exciting new category of web usage data, namely usage of the Web of Data, which is one of the topics we address in this workshop. Participants are invited to come up with interesting analyses, applications, alignments, etc. for these datasets.

3. PROGRAM

USEWOD2011 offers a full-day program starting with a keynote speaker who represents the cross-disciplinary approach USEWOD stands for. There will be presentations of regular and position papers and a semi-structured discussion on the practical steps that need to be taken to drive this interdisciplinary subfield further. Finally, participants of the challenge will have the opportunity to present their work and a prize will be awarded to the best challenge submission.

Presented papers will address the combination of the two research areas: they will show how semantic resources and techniques can be used to strengthen usage data analysis and, vice versa, how usage data can enhance semantic tools and applications. Within these boundaries, we keep the scope broad. We welcome contributions using any form of semantic information, from formal ontologies to linked data and folksonomies. All records of user actions are considered usage logs; we do not limit ourselves to any format or method of collection of usage information. This ranges from traditional content-consumption logs to various forms of content-production logs, i.e. navigation, application-related transactions, queries, tagging, editing, and similar activities.

Topics within the scope of the workshop:

- Analysis and mining of usage logs of semantic resources and applications.
- Inferring semantic information from usage logs.
- Methods and tools for semantic analysis of usage logs.
- Representing and enriching usage logs with semantic information.
- Usage-based evaluation methods and frameworks; gold standards for evaluation of semantic web applications.
- Specifics and semantics of logs for content-consumption and content-creation.
- Using semantics for recommendation, personalization and adaptation.
- Usage-based recommendation, personalization and adaptation of semantic web applications.
- Exploiting usage logs for semantic search.
- Data sharing, privacy, and privacy-protecting policies and techniques.

4. REFERENCES

- [1] R. Baeza-Yates and A. Tiberi. Extracting semantic relations from query logs. *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Jose, California*, 2007.
- [2] D. Benz, B. Krause, G. P. Kumar, A. Hotho, and G. Stumme. Characterizing semantic relatedness of search query terms. *Proceedings of the 1st Workshop on Explorative Analytics of Information Networks*, 2009.
- [3] B. Berendt, G. Stumme, and A. Hotho. Usage mining for and on the Semantic Web. *H. Kargupta, A. Joshi, K. Sivakumar, and Y. Yesha (Eds.), Data Mining: Next Generation Challenges and Future Directions*, pp. 461-480, 2004.
- [4] K. Hofmann, M. de Rijke, B. Huurnink, and E. Meij. A semantic perspective on query log analysis. *Working Notes for CLEF*, 2009.
- [5] V. Hollink, T. Tsikrika, and A. P. de Vries. The semantics of query modification. *Proceedings of 9th International Conference on Adaptivity, Personalization and Fusion of Heterogeneous Information (RIA0)*, 2010.
- [6] E. Meij, M. Bron, L. Hollink, B. Huurnink, and M. de Rijke. Learning semantic query suggestions. *Proceedings of the 8th International Semantic Web Conference, Chantilly, VA, USA.*, 2009.
- [7] P. Mika, E. Meij, and H. Zaragoza. Investigating the semantic gap through query log analysis. *Proceedings of the 8th International Semantic Web Conference, Chantilly, VA, USA.*, 2009.
- [8] K. Möller, M. Hausenblas, R. Cyganiak, and G. Grimnes. Learning from Linked Open Data usage: patterns & metrics. *Proceedings of the WebSci10: Extending the Frontiers of Society On-Line*, 2010.
- [9] T. Sakai and K. Nogami. Serendipitous search via wikipedia: a query log analysis. *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, USA*, 2009.