

DiversiWeb 2011

First International Workshop on Knowledge Diversity on the Web

Elena Simperl
KIT
Karlsruhe, Germany
elena.simperl@kit.edu

Denny Vrandečić
KIT
Karlsruhe, Germany
denny.vrandecic@kit.edu

Devika P. Madalli
Indian Statistical Institute
Bangalore, India
devika@drtc.isibang.ac.in

Enrique Alfonseca
Google
Zurich, Switzerland
ealfonseca@google.com

ABSTRACT

The workshop provides an interdisciplinary forum for researchers and practitioners to present and discuss their ideas related to the challenges posed by diversity on the Web. We address a wide array of interdisciplinary questions, which need to be tackled in order to preserve the fragile balance between a world that is continually converging and growing together, the rich diversity of the global society, and the dangers of fragmentation and splintering.

Categories and Subject Descriptors

H.3.5 [Information Storage and Retrieval]: On-line Information Services—*Web-based services*; D.5.4 [Information Interfaces and Presentation]: Hypertext/Hypermedia; K.4.2 [Computers and Society]: Social Issues

General Terms

Human Factors

Keywords

Diversity

1. INTRODUCTION

Almost 20 years after its introduction, the Web provides a platform for the publication, use and exchange of information, at planetary scale, on virtually every topic, and representing an amazing diversity of opinions, viewpoints, mindsets and backgrounds. The success of the Web can be attributed to several factors, most notably to its principled scalable design, but also to a number of subsequent ICT developments such as smart user-generated content, mobile devices, and most recently cloud computing.

The first two of these have dramatically lowered the last barriers of entry when it comes to producing and consuming information online, leading to an unprecedented growth and mass collaboration. They are responsible for hundreds of millions of users all over the globe creating high-quality encyclopedias, publishing Terabytes of multimedia content,

contributing to world-class software, and lively taking part in defining the agenda of many aspects of our society by raising their voices, and publicly expressing and sharing their ideas, viewpoints, and resources.

The other side of the coin in this unique success story is, nevertheless, the great challenges associated with managing the sheer amounts of information continuously being published online, whilst allowing for purposeful use, and leveraging the diversity inherently unfolding through global-scale collaboration. In this context, diversity includes different opinions, sentiments, preferences, or worldviews that are reflected in the way information is expressed on the Web. These challenges are still to be solved at many levels, from the infrastructure to store and access the information, through the methods and techniques to make sense out of it, to the paradigms underlying the processes of Web-based information provision and consumption.

As an example, when searching for blog posts, state-of-the-art technology – be that popularity-based algorithms, recommendation engines or collaborative filters – tends to return either the most popular posts, or those which correspond with a personal profile and therefore with the known opinions and tastes of the reader. Alternative points of view, and new unexpected content, are not taken into account as they are not highly ranked, and posts expressing different opinions are sometimes even discarded.

This behavior has particularly negative consequences when dealing with information that is expected and intended to be subject to diverse opinion – as is the case with news reports, ratings of products or media content, customer reviews, or any other type of subjective assessment. The same negative effects apply in a community-driven environment that is designed for collaboration – the most obvious example here being Wikipedia and the blogosphere. The information diversity exposed in such an environment, impressive both with respect to scale and the richness of opinions and viewpoints expressed, cannot be handled without adequate computer support in an economically feasible manner. In the long run, maintaining the current state-of-affairs will change the ways and the extent to which people are informed (or not) on a particular topic, tremendously influencing how they look into that topic, what they find about it and what they think about it.

On top of all this, it is meanwhile acknowledged that the

current state of affairs hampers true collaboration. For example, Wikipedia is a tremendous success, but it is also a largely meritocratic system with a decreasing number of active contributors, whereas the blogosphere has to deal with the limited attention of the blog authors. What is needed are novel concepts, methods and tools that allow humans and machines to leverage the huge amounts of information created by a community, based on interaction models that support expressing, communicating and reasoning about divergent models simultaneously. This would not only enhance true collaboration, but would also significantly improve various aspects of the information management life cycle, thus addressing information overload in sectors which rely on opinions-driven information sources and mass participation – news, ratings, reviews, and social and information sharing portals of any kind.

2. GOALS OF THE WORKSHOP

The overall aim of this workshop is to provide an interdisciplinary forum for researchers and practitioners to present and discuss their ideas related to the challenges posed by diversity on the Web. We aim to address a wide array of interdisciplinary questions, which need to be tackled in order to preserve the fragile balance between a world that is continually converging and growing together, the rich diversity of the global society, and the dangers of fragmentation and splintering. This includes but is not limited to questions such as *'How to model diversity?'*, *'How to discover bias and opinion in blog posts, tweets, forum items, wiki edits, etc.'*, *'How to rank, aggregate, summarize, and exploit information in a diversity-aware manner?'*, *'What are the applications of diversity-rich information sources?'*, *'How can we use diversity as an asset instead of regarding it as a barrier?'*.

3. TOPICS

In particular we welcome submissions that

- Analyze the capabilities of current information management models, algorithms and technologies to leverage knowledge diversity,
- Extend existing models, methods, techniques and tools to accommodate the requirements arising from paying a proper account to diversity-expressed information sources and communication and collaboration environments characterized by a rich variety of opinions and viewpoints.
- Discuss the foundations of knowledge diversity on the Web and propose alternative paradigms,
- Propose novel evaluation strategies, methods and techniques to assess the impact of diversity-minded information management.

Topics of the workshop include, but are not limited to:

- Risks and advantages of diversity and diversification on the Web
- Facets of knowledge diversity and conceptual and formal models for representing and understanding diversity

- Discovery and mining of corpora for diversity-related information
- Use of Natural Language Processing techniques for diversity mining
- Classifying Web 2.0 content items such as blog posts, videos, tweets, and wiki-edits by their biases
- Usage and benefits of diversity in the corporate context, e.g. in order to understand feedback and communication with the customer
- Enabling or improving communication and collaboration over barriers induced by diversity
- Extensions to Web applications taking diversity into account
- Exposing and explaining diversity to end users
- User experiences avoiding the radicalization of groups by exposing them to alternatives
- User interfaces allowing the explicit annotation of content with diversity markers
- Studies on the acceptance by end-users of diversified applications.

We explicitly invite experience reports on topics related to diversity from the Web as it is used in India.

4. SUBMISSION TYPES

We aim at four different kind of submission:

1. research papers of the length of 8 pages presenting mature work, prototypes and methodologies,
2. position papers of the length of 4 pages presenting early work and elaborated ideas,
3. demo outlines of the length of 2 pages, and
4. experience reports of up to 6 pages about dealing and managing with diversity, especially within the usage of the Web in India.

5. ACKNOWLEDGMENTS

The workshop was partially funded by EU IST FP7 projects RENDER¹ and Living Knowledge.² We also thank the program committee for the indispensable help in the selection process for the workshop program.

¹<http://render-project.eu>

²<http://livingknowledge-project.eu/>