

# **OSIR 2006**

Second Workshop On

**Open Source Information Retrieval**

In Conjunction with the 2006 ACM SIGIR Conference

August 10, 2006

Seattle, WA

Edited by Michel Beigbeder, Wray Buntine and Wai Gen Yee

# OSIR 2006 Organization

## Workshop chairs

Michel Beigbeder  
G2I Department  
École Nationale Supérieure des Mines de Saint-Étienne, France

Wray Buntine  
Complex Systems Computation Group  
Helsinki Institute for Information Technology, Finland

Wai Gen Yee  
Department of Computer Science  
Illinois Institute of Technology, USA

## Program Committee

Antonio Badia, University of Louisville, USA  
Michel Beigbeder, École Nationale Supérieure des Mines de Saint-Étienne, France  
Wray Buntine, Helsinki Institute for Information Technology, Finland  
Abdur Chowdhury, America Online, USA  
Bruce Croft, University of Massachusetts, USA  
Doug Cutting, Yahoo, USA  
Ophir Frieder, Illinois Institute of Technology, USA  
Nazli Goharian, Illinois Institute of Technology, USA  
Donald Kraft, Louisiana State University, USA  
Olfa Nasraoui, University of Louisville, USA  
Iadh Ounis, University of Glasgow, UK  
Michael Stack, Internet Archive, USA  
Wai Gen Yee, Illinois Institute of Technology, USA  
Clement Yu, University of Illinois at Chicago, USA

## Workshop Web site

[www.emse.fr/OSIR06](http://www.emse.fr/OSIR06)

# OSIR 2005

## Open Source Information Retrieval

The World Wide Web has grown to be a primary source of information for millions of people. Due to the size of the Web, search engines have become the major access point for this information. However, “commercial” search engines use hidden algorithms that put the integrity of their results in doubt, collect user data that raises privacy concerns, and target the general public thus fail to serve the needs of specific search users. Open source search, like open source operating systems, offers alternatives.

The goal of the Open Source Information Retrieval Workshop (OSIR) is to bring together practitioners developing open source search technologies in the context of a premier IR research conference to share their recent advances, and to coordinate their strategy and research plans. The intent is to foster community-based development, to promote distribution of transparent Web search tools, and to strengthen the interaction with the research community in IR.

A workshop about Open Source Web Information Retrieval was held last year in Compigne, France as part of WI 2005. The focus of this workshop is broadened to the whole open source information retrieval community.

We want to thank all the authors of the submitted papers, the members of the program committee: Antonio Badia, Abdur Chowdhury, Bruce Croft, Doug Cutting, Ophir Frieder, Nazli Goharian, Donald Kraft, Olfa Nasraoui, Iadh Ounis, Michael Stack, Clement Yu, and the several reviewers whose contributions have resulted in these high quality proceedings.

Michel Beigbeder, Wray Buntine and Wai Gen Yee



# Table of contents

Low Latency Index Maintenance in Indri <i>Trevor Strohman and W. Bruce Croft</i>	7
PF/Tijah: text search in an XML database system <i>Djoerd Hiemstra, Henning Rode, Roel van Os and Jan Flokstra</i>	12
Terrier: A High Performance and Scalable Information Retrieval Platform <i>Iadh Ounis, Gianni Amati, Vassilis Plachouras, Ben He, Craig Macdonald and Christina Lioma</i>	18
Deploying Lucene on the Grid <i>Edgar Meij and Maarten de Rijke</i>	25
IR-Wire: A Research Tool for P2P Information Retrieval <i>Shefali Sharma, Linh Thai Nguyen and Dongmei Jia</i>	33
Tagging in Peer-to-Peer Wikipedia – A Method to Induce Cooperation <i>Jenneke Fokker, Wray Buntine and Johan Pouwelse</i>	39
Web Recommender System Implementations in Multiple Flavors: Fast and (Care-)Free for All <i>Olfa Nasraoui, Zhiyong Zhang and Esin Saka</i>	46
An Effectiveness Measure for Evaluating Open Retrieval Systems <i>Hsieh-Chang Tu and Jieh Hsiang</i>	54
SHOW AND TELL: A Seamlessly Integrated Tool For Searching with Image Content And Text <i>Zhiyong Zhang, Carlos Rojas, Olfa Nasraoui and Hichem Frigui</i>	60
Standards for Open Source Information Retrieval <i>Wray Buntine, Michael P. Taylor and François Lagunas</i>	68
Panel: IR Research and Open Source	73

