Foreword

These proceedings contain the contributed papers of the Seventh Workshop on Exploiting Semantic Annotations in Information Retrieval (ESAIR 2014), held at CIKM 2014 in Shanghai, on November 7, 2014. After successful workshops at ECIR’08 in Glasgow, WSDM’09 in Barcelona, CIKM’10 in Toronto, CIKM’11 in Glasgow, CIKM’12 at Maui, and CIKM’13 in San Francisco, this year’s workshop will remain to advance the general research agenda on this core problem, with an explicit focus on one of the most challenging aspects to address in the coming years. The main remaining challenge is on the user’s side—the potential of rich document annotations can only be realized if matched by more articulate queries exploiting these powerful retrieval cues—and a more dynamic approach is emerging by exploiting new forms of query autosuggest. How can the query suggestion paradigm be used to encourage searcher to articulate longer queries, with concepts and relations linking their statement of request to existing semantic models? How do entity results and social network data in “graph search” change the classic division between searchers and information and lead to extreme personalization—are you the query? How to leverage transaction logs and recommendation, and how adaptive should we make the system? What are the privacy ramifications and the UX aspects—how to not creep out users?

These and other related questions will be discussed at this open format workshop -- the aim is to provide paths for further research to change the way we understand information access today!

The workshop will consist of three main parts:

1. Two keynotes to help us frame the problem, and create a common understanding of the challenges: Peter Mika (Yahoo Labs) and Silviu-Petru Cucerzan (Microsoft Research).

2. A boaster and poster session with 11 papers selected by the program committee from 15 submissions (a 73% acceptance rate). Each paper was reviewed by at least two members of the program committee.

3. Breakout groups on different aspects of exploiting semantic annotations, with reports being discussed in the final session.

These papers represent the ideas and opinions of the authors who are trying to stimulate debate. It is the combination of these papers and the debate that will make the workshop a success! We thank Google for sponsoring the workshop, which helped us build a program with great keynotes speakers and allowed us to continue the tradition of organizing social events after the workshop. We thank the ACM and CIKM for hosting this workshop, and Huan Liu, Xiaofeng Meng, and Lisa Tolles for their outstanding support in the organization. Thanks also go to the program committee, the authors of the papers, and all participants without whom there would be no workshop.

Omar Alonso
ESAIR 2014 Co-chair
Microsoft

Jaap Kamps
ESAIR 2014 Co-chair
University of Amsterdam

Jussi Karlgren
ESAIR 2014 Co-chair
Gavagai & KTH
Stockholm
# Table of Contents

ESAIR 2014 Workshop Organization .............................................................................................................. vii

## Keynote Address I
Session Chair: Jaap Kamps *(University of Amsterdam)*

- **Semantic Search at Yahoo!** ...................................................................................................................... 1
  Peter Mika *(Yahoo Labs)*

## Keynote Address II
Session Chair: Omar Alonso *(Microsoft)*

- **Linking to Web Knowledge Bases and Applications to Web Search** .................................................. 3
  Silviu Petru Cucerzan *(Microsoft)*

## Boaster Session
Session Chair: Jaap Kamps *(University of Amsterdam)*

- **Documents Search Using Semantics Criteria** ........................................................................................... 5
  Santiago Cotelo, Alejandro Makowski, Luis Chiruzzo, Dina Wonsever *(Universidad de la República, Montevideo, Uruguay)*

- **Towards Named-Entity-based Similarity Measures: Challenges and Opportunities** ....................... 9
  Tom De Nies *(Ghent University-iMinds, Belgium)*, Christian Beecks *(RWTH Aachen University, Germany)*,
  Wesley De Neve *(Ghent University & KAIST)*, Thomas Seidl *(RWTH Aachen University)*,
  Erik Mannens, Rik Van de Walle *(Ghent University)*

- **Can Corpus Similarity-Based Self-Annotation Assist Information Retrieval?** ................................. 13
  Vinay Deolalikar *(Hewlett-Packard Research)*

- **AIDA-Social: Entity Linking on the Social Stream** ................................................................. 17
  Yusra Ibrahim, Mohamed Amir Yosef, Gerhard Weikum *(Max Planck Institute for Informatics, Germany)*

- **A Probabilistic Concept Annotation for IT Service Desk Tickets** .................................................. 21
  Ea-Ee Jan, Kuan-Yu Chen, Tsuyoshi Ide *(IBM T.J. Watson)*

- **Semantic Annotation with RescoredESA: Rescoring Concept Features Generated from Explicit Semantic Analysis** ........................................................................................................... 25
  Zhuoren Jiang *(Dalian Maritime University, China)*,
  Miao Chen, Xiaozhong Liu *(Indiana University Bloomington)*

- **Using Semantic Role Labeling to Predict Answer Types** .......................................................... 29
  Zuyao Li *(University of Southern California)*, Peter Exner, Pierre Nugues *(Lund University, Sweden)*

- **Leverage the Associations between Documents, Subject Headings and Terms to Enhance Retrieval** ......................................................................................................................... 33
  Jin Mao *(Wuhan University, China)*, Kun Lu *(University of Oklahoma)*

- **Bringing Head Closer to the Tail with Entity Linking** ........................................................................ 37
  Manisha Verma *(University College London)*, Diego Ceccarelli *(ISTI CNR, Pisa)*

- **A Fragment-Based Similarity Measure for Concept Hierarchies and Ontologies** ......................... 41
  Hui Yang *(Georgetown University)*

- **Exploiting Inference from Semantic Annotations for Information Retrieval: Reflections from Medical IR** ......................................................................................................................... 43
  Guido Zuccon *(Queensland University of Technology)*, Bevan Koopman *(CSIRO, Australia)*,
  Peter Bruza *(Queensland University of Technology)*

## Author Index ................................................................................................................................................. 46
ESAIR 2014 Workshop Organization

General Chairs:  Omar Alonso (Microsoft, Mountain View CA, USA)
Jaap Kamps (University of Amsterdam, The Netherlands)
Jussi Karlgren (Gavagai Stockholm, Sweden)

Program Committee:  Omar Alonso (Microsoft Bing, USA)
Krisztian Balog (University of Stavanger, Norway)
Shlomo Geva (QUT, Australia)
Djoerd Hiemstra (University of Twente, The Netherlands)
Jaap Kamps (University of Amsterdam, The Netherlands)
Noriko Kando (National Institute of Informatics, Japan)
Jussi Karlgren (UC Berkeley, USA)
Ray Larson (Microsoft Research, USA)
Liz Liddy (Syracuse University, USA)
Maarten Marx (University of Amsterdam, The Netherlands)
Peter Mika (Yahoo! Research Barcelona, Spain)
Patrick Pantel (Yahoo! Research, USA)
Livia Polanyi (Stanford University, USA)
Ralf Schenkel (Universität Passau, Germany)
Andrew Trotman (University of Otago, New Zealand)
Roman Yangarber (University of Helsinki, Finland)

Sponsors:

In cooperation with: