HIA'15: Heterogeneous Information Access Workshop at WSDM 2015

Ke Zhou
Yahoo Labs
London, U.K.
kezhou@yahoo-inc.com

Roger Jie Luo
Yahoo Labs
Sunnyvale, USA
luojie@yahoo-inc.com

Djoerd Hiemstra
Department of Computer Science
University of Twente
hiemstra@cs.utwente.nl

Joemon M. Jose
School of Computing Science
University of Glasgow
Joemon.Jose@glasgow.ac.uk

ABSTRACT

The HIA'15 workshop aims to bring together information retrieval practitioners from industry and academic researchers concerned with heterogeneous information access and search federation. We would like to create a forum to encourage discussion and exchange of ideas on heterogeneous information access in different contexts. To facilitate the discussion, we encourage submissions on ideas and results from different aspects of heterogeneous information access including aggregated search, composite retrieval, personal search, structured search, etc. Another objective of the workshop is to encourage submissions with novel ideas (e.g. new applications) on heterogeneous information access and potential future directions of this area.

Categories and Subject Descriptors

H.3.3 [**Information Systems**]: Information Search and Retrieval

Keywords

federated search, aggregated search, composite retrieval

1. INTRODUCTION

Information access is becoming increasingly heterogeneous. Especially when the user's information need is for exploratory purpose, federating a set of diverse results from different resources [6] could benefit the user. For example, when a user is planning a trip to China, retrieving and showing results from vertical search engines like travel, flight information, map and Q2A sites could satisfy the user's rich and diverse information need. It is also common that users search their own desktop (e.g. Splotlight) to re-find or browse different types of files (source codes, documents, emails, etc.) when aiming to fulfill a given work task. This heterogeneous search paradigm is useful in many contexts and brings many new challenges.

Aggregated search [1] and composite retrieval [3] are two instances of this new search paradigm that aims to be suitable for such information needs. Federation could also be useful in several other scenarios: a user aims to re-find comprehensive information about his query in his personal search [4]; or a user searches and gathers different nugget information (e.g. an entity) from a set of RDF Web datasets [2] (e.g., DBpedia, IMDB, etc.); or a user searches a set of different files in a peer-to-peer online file sharing systems [5]. This is an emerging area as different services provided are becoming more heterogeneous and complex. Therefore, there are a number of directions that might be interesting for both the research and industrial community. How to select the most relevant resources and present them concisely in order to best satisfy the user? How to model the complex user behavior in this search scenario? How can we evaluate the performance of these systems? Those are a few key interesting research questions to study for heterogeneous information access. The main objective of the workshop is to draw researchers' attention to this rising field as well as to inspire them to discuss and collaborate in order to further define new directions for heterogeneous information access.

2. WORKSHOP TOPICS

The workshop topics are within the context of heterogeneous information access. They include but are not limited to:

- User Modelling for Heterogeneous Information Access
  - Short and Long-term User Modelling
  - Personalization
  - Diversification
  - Coherence
- Metrics and Measurements
  - Metrics based on test collection

- Controlled laboratory study
- Online metrics
- Test Collection

- Optimization
  - Resource/Vertical Selection
  - Result Presentation Visualization Strategies
  - Presentation Optimization
  - Result Diversification
- Applications
  - Aggregated/Federated Search
  - Composite Retrieval
  - Personal Search
  - Structured/Semantic Search
  - Peer-to-peer information retrieval

3. WORKSHOP ORGANIZERS

- **Ke Zhou** is a research scientist working in Yahoo Labs London. He was previously a research associate in Language Technology Group in University of Edinburgh working on text mining and information retrieval. He has conducted his PhD research on evaluation of aggregated search at the Information Retrieval Group in University of Glasgow. He has published in reputable conferences and journals (SIGIR, WWW, CIKM, TOIS) and served as PC member for SIGIR, CIKM, ECIR and AIRS. He also served as a co-organizer for NTCIR-11/12 IMine task and TREC FedWeb 2014 task.

- **Roger Jie Luo** is a Research Scientist and an engineering & science lead at Yahoo Labs. At Yahoo, he lead the efforts to improve the ranking of aggregated search results on the Yahoo! search result page and on understanding the users' intents given a query. He obtained his PhD in computer science on machine learning from Swiss Federal Institute of Technology in Lausanne (EPFL), Switzerland.

- **Djoerd Hiemstra** is an associate professor of search engine and database technology at the University of Twente. He wrote an often cited PhD thesis on the use of statistical language models for information retrieval. His research interests include information retrieval, natural language processing and probabilistic graphical models. He co-authored over 200 research papers. Djoerd co-organized several workshops and conferences including ACM SIGIR 2007, several editions of the Dutch-Belgian Information Retrieval Workshop (DIR), the SIGIR 2010 workshop on Accessible search and the ECIR 2013 workshop on Group membership and search.

- **Joemon M. Jose** is a full Professor at the School of Computing Science, University of Glasgow. He is a fellow of the BCS, IET and a chartered information technology professional (CITP), member of the ACM and IEEE. He has a well-established reputation in research on multimedia information retrieval, developing advanced retrieval models, studying the role of emotion in search, personalization and adaptive retrieval. He has published over 150 journal and conference articles and leads a team of 9 PhD students and 2 post-doctoral researchers. He successfully, as local coordinator, completed 4 major EU funded projects on multimedia retrieval and multi-modal interaction (SALERO, MIAUCE, SEMEDIA, K-SPACE) and now participates in the LiMOSiNe project. He has organized number of information retrieval related events like ICMR 2014, ECDL 2010, workshops at previous SIGIR conferences, IR Fest in Glasgow 2005 etc. He has been senior PC member of the SIGIR, CIKM, ECIR conferences.

4. WORKSHOP PROGRAM

4.1 Invited Speakers

- Yiqun Liu, Associate Professor, Tsinghua University
- Maarten de Rijke, Professor, University of Amsterdam
- Milad Shokouhi, Senior Applied Science Lead, Microsoft Research Cambridge
- Gui-rong Xue, Senior Director, Aliyun.com; Associate Professor, Shanghai Jiao-Tong University

4.2 Program Committee

- Jaime Arguello (University of North Carolina at Chapel Hill, USA)
- Roi Blanco (Yahoo! Lab Barcelona, Spain)
- Mohand Boughanem (University of Toulouse, France)
- Jamie Callan (Carnegie Mellon University, USA)
- Zhicheng Dou (Microsoft Research Asia, China)
- Jin Young Kim (Microsoft Bing, USA)
- Yiqun Liu (Tsinghua University, China)
- Ilya Markov (University of Amsterdam, Netherlands)
- Maarten de Rijke (University of Amsterdam, Netherlands)
- Pavel Serdyukov (Yandex Moscow, Russia)
- Milad Shokouhi, (Microsoft Research Cambridge, U.K.)
- Paul Thomas (CSIRO, Australia)
- Arjen P. de Vries (CWI, Netherlands)
- Hengshuai Yao (Yahoo Labs Sunnyvale, USA)
- Qian You (Amazon, USA)

5. REFERENCES

[1] J. Arguello, F. Diaz, J. Callan, and J.-F. Crespo. Sources of evidence for vertical selection. In *SIGIR '09*, pages 315–322, New York, NY, USA, 2009. ACM.

[2] D. Herzig, and T. Thanh. Heterogeneous web data search using relevance-based on the fly data integration. In *WWW '12*, pages 141–150, New York, NY, USA, 2012. ACM.

[3] H. Bota, K. Zhou, J. M. Jose, and M. Lalmas. Composite retrieval of heterogeneous web search. In *WWW'14*, pages 119–130, International World Wide Web Conferences 2014.

[4] K. Jinyoung, and W. B. Croft. Ranking using multiple document types in desktop search. In *SIGIR'10*, pages 50–57, Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval. ACM, 2010.

[5] T. Almer S., D. Hiemstra, and D. Trieschnigg. Peer-to-peer information retrieval: An overview. In *TOIS 2012*. ACM Transactions on Information Systems (TOIS) 30.2 (2012): 9.

[6] M. Shokouhi and S. Luo. Federated search. In *FTIIR*. Foundations and Trends in Information Retrieval 5.1 (2011): 1-102.