# EVIA 2016: The Seventh International Workshop on Evaluating Information Access

Charles L. A. Clarke
University of Waterloo
*claclark@gmail.com*

Emine Yilmaz
University College London
*emine.yilmaz@ucl.ac.uk*

## 1 Introduction

Information access technologies provide the interface between human information needs and digital information resources. The reliable evaluation of these technologies has been recognized for decades as central to the advancement of the field. As information retrieval technologies become more pervasive, the forms of retrieval more diverse, and retrieval tools richer, the importance of effective, efficient, and innovative evaluation grows as well.

The International Workshop on Evaluating Information Access (EVIA) provides a refereed forum for the publication of novel research across the field of information retrieval evaluation, including test collections, metrics, statistics, user studies, implicit user feedback, and crowdsourcing. Every year-and-a-half EVIA is held in conjunction with the NTCIR conference in Tokyo, Japan at the National Institute of Informatics (NII). The seventh EVIA workshop was held on June 7, 2016.

The EVIA 2016 program consisted of one invited talk, presentation of seven refereed research papers, and a panel discussing the future of information access evaluation. The remainder of this report identifies the EVIA 2016 workshop program committee and briefly summarizes the invited keynote talk, accepted research papers, and the panel discussion. Proceedings of the workshop can also be found online [1].

## 2 Program Committee

Atsushi Fujii, Tokyo Institute of Technology, Japan
Donna Harman, NIST, USA
Hideo Joho, University of Tsukuba, Japan
Gareth Jones, Dublin City University, Ireland
Noriko Kando, National Institute of Informatics, Japan
Evangelos Kanoulas, University of Amsterdam, The Netherlands
Diane Kelly, University of North Carolina, USA

[1] `http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings12/EVIA/toc_evia.html`

Liadh Kelly, Trinity College Dublin, Ireland
Filip Radlinski, Microsoft Research, UK
Tetsuya Sakai, Waseda University, Japan
Mark Sanderson, RMIT University, Australia
Falk Scholer, RMIT University, Australia
Anne Schuth, University of Amsterdam, The Netherlands
Mark Smucker, University of Waterloo, Canada
Ian Soboroff, NIST, USA
Ruihua Song, Microsoft Research Asia, China
Adith Swaminathan, Cornell University, USA
Paul Thomas, CSIRO, Australia

# 3 Keynote Talk

The workshop started with a keynote talk titled *"Understanding and Predicting Search Satisfaction in a Heterogeneous Environment"*, by Dr. Yiqun Liu from Tsinghua University, China.

Traditional search performance evaluation can be performed using metrics based on result relevance or alternative measures based on users search experience. However, it was shown that these metrics may not be perfectly correlated with the satisfaction of end users. In his talk, Dr. Liu described some of the recent progresses they have made in the understanding and effective prediction of search satisfaction. He started his talk by analysing the relationship between relevance, usefulness and satisfaction. Then, he focused on understanding users satisfaction perception in a heterogeneous search environment and discussed how the different vertical results on SERPs may affect users satisfaction. Finally, he described a satisfaction prediction framework which relies on users mouse movement patterns (motifs) to identify satisfied or unsatisfied search sessions.

# 4 Refereed Papers

Each submitted paper was reviewed by at least three members of the program committee. As a result of this process, the EVIA 2016 programme commitee accepted seven refereed papers for presentation at the workshop.

The first paper presented at the workshop titled *"An Easter Egg Hunting Approach to Test Collection Building in Dynamic Domains"*, by Seyyed Hadi Hashemi, Charles L. A. Clarke, Adriel Dean-Hall, Jaap Kamps and Julia Kiseleva. The paper described a novel approach to reusable test collection construction, where they inject judged pages into an existing corpus, and have systems retrieve pages from the extended corpus with the aim to create a reusable test collection.

Tetsuya Sakai presented the second paper *"On Estimating Variances for Topic Set Size Design"*, which he co-authored with Lifeng Shang. The paper focused on the problem of determining appropriate number of topics in constructing a test collection and using some pilot data analysed how many teams and topics are actually necessary in the pilot data for obtaining accurate variance estimates.

The third paper presented at the workshop, *"A Laboratory-Based Method for the Evaluation of Personalised Search"* was by Camilla Sanvitto, Debasis Ganguly, Gareth J. F. Jones and Gabriella

Pasi. The paper described a method that enables the creation of publically available extended test collections to allow repeatable laboratory-based evaluation of personalised search.

On the paper titled *"Promoting Repeatability Through Open Runs"*, Ellen Voorhees, Shahzad Rajput and Ian Soboroff focused on the concept of "Open Runs" in response to the increasing focus on repeatability of information retrieval experiments. They explored possible difficulties associated with using open runs for repeatability and made suggestions for how to address the deficiencies.

The fifth paper of the workshop, *"Evaluating Search Among Secrets"*, by Douglas W. Oard, Katie Shilton and Jimmy Lin argued for the need for creating a new class of search algorithms designed to effectively search among secrets by balancing the user's interest in finding relevant content with the provider's interest in protecting sensitive content and descibed some approaches that can be used for evaluating the quality of such systems.

Liana Ermakova presented the paper titled *"Automatic Sentence Ordering Assessment Based on Similarity"*, focused on the problem of evaluation of sentence ordering and presented an automatic evaluation framework for analysis of the metrics of sentence order that requires only a text collection.

The final paper presented at the workshop was titled *"Two-layered Summaries for Mobile Search: Does the Evaluation Measure Reflect User Preferences?"* by Makoto P. Kato, Virgil Pavlu, Tetsuya Sakai, Takehiro Yamamoto and Hajime Morita. The paper focused on two-layered summarization for mobile search, and proposed an evaluation framework for such summaries.

# 5 Panel Discussion

The panel consisted of Gareth J.F. Jones (Dublin City University, Ireland), Yiqun Liu (Tsinghua University, China), Mark Sanderson (RMIT University, Australia), and Ian Soboroff (NIST, USA), The panel discussed the future of the Cranfield paradigm, the future of EVIA, and the future of information retrieval evaluation. Other topics included the "live labs" approach to evaluation, simulation of interaction, and crowdsourcing.