First International Workshop on Professional Search (ProfS2018)

Co-located with SIGIR 2018

Suzan Verberne
Leiden University
s.verberne@liacs.leidenuniv.nl

Jiyin He
Centrum Wiskunde & Informatica
jiyinhe@acm.org

Udo Kruschwitz
University of Essex
udo@essex.ac.uk

Birger Larsen
Aalborg University
birger@hum.aau.dk

Tony Russell-Rose
UXLabs
tgr@uxlabs.co.uk

Arjen P. de Vries
Radboud University
arjen@acm.org

ABSTRACT

Professional search is a problem area in which many facets of information retrieval are addressed, both system-related (e.g. distributed search) and user-related (e.g. complex information needs), and the interface between user and system (e.g. supporting exploratory search tasks). Professional search tasks have specific requirements, different from the requirements of generic web search. The aim of this workshop is to bring together researchers to work on the requirements and challenges of professional search from different angles. We will have an interactive workshop where researchers not only present their scientific results but also work together on the definition of future challenges and solutions with input from information professionals. The workshop will deliver a roadmap of research directions for the years to come.

CCS CONCEPTS

• **Information systems → Specialized information retrieval**; *Users and interactive retrieval*;

KEYWORDS

professional search, domain-specific search, complex search tasks, user requirements

1 MOTIVATION

Professional search in specific domains has been addressed in IR research over the last decades. Although each domain (e.g. legal, medical, academic, governmental) has its own idiosyncrasies, professional search tasks have specific requirements in common that

are different from requirements of generic web search engines. These requirements follow directly from the context and needs of professional searchers: Searchers in different domains often exhibit particular search behavior different from general Web search [4, 7]. These unique behavioral patterns can be both a nature of the profession as well as a result of using a particular professional search tool. The differences between professional search and ad hoc web search can be summarized in three key characteristics:

Heterogeneous information sources. Often professional searchers need to collect and synthesize information from various sources, in different modalities, and perhaps in different languages. For example, academic researchers typically find information across multiple resources using multiple different search tools [2]. And assisting users to discover related materials in the right work context has shown to be useful for law professionals [8].

Highly interactive procedure. The information tasks that professional users need to accomplish often involve multiple stages. Therefore, they commonly exhibit exploratory search behavior with relatively complex queries and long sessions. The information needs are typically high-value and often recall-oriented, and include activities beyond search such as analysis, updating, and monitoring [1, 9].

Highly specialized activities. Searchers in professional contexts are often engaged with information tasks that are highly complex. Where web search engines largely rely on clickthrough data to determine relevance of a search result, clickthrough popularity is less useful as a ranking criterion in professional search because the information needs are highly specific and the amount of click data available from other users is limited [3].

The workshop addresses these requirements from multiple angles; covering many different facets of professional search (see Section 2). We envisage a highly interactive workshop (see Section 3) where researchers work with input from information professionals to their mutual benefit. The workshop will deliver a roadmap of research directions for the years to come.

Successful related workshops were organized in 2013, 2015 and 2017, co-located with ECIR and CHIIR (see Section 7). SIGIR is a key venue that attracts researchers in IR, Information Sciences, Human-Computer Interaction, and Natural Language Processing from all over the world. The workshop provides the opportunity to work

Table 1: Workshop schedule

9:15 - 9:30	Introduction
9:30 - 10:00	Presentation of survey results
10:00 - 10:30	Coffee Break I
10:30 - 11:15	Keynote 1: Paul Bennett
11:15 - 12:00	3 oral presentations
12:00 - 13:30	Lunch
13:30 - 14:15	Keynote 2: David D. Lewis
14:15 - 15:00	3 oral presentations
15:00 - 15:30	Coffee Break II
15:30 - 16:30	Breakout session
16:30 - 17:00	Reports from the breakout sessions & conclusions

with people from multiple fields (both scientific and professionally) on a common challenge: IR in professional contexts.

2 THEME AND PURPOSE

We invited submissions in one or more of the following (not exclusive) topics:

- distributed and federated search in professional search environments;
- multi-modal and multi-lingual search in professional environments;
- supporting exploratory search and complex, high-value search tasks;
- the collection and analysis of information needs in professional search domains;
- comparison of search needs, approaches and behavior patterns across different professions;
- the evaluation of longitudinal, recall-oriented tasks;
- multi-faceted evaluation of professional search tasks;
- analysis of session log data for complex search tasks;
- beyond the search engine result page: browsing, reading, analyzing and using retrieval results;
- personalization and information recommendation for professional search;
- search applications in a professional context (such as enterprise search, e-discovery and email search);
- the evaluation and improvement of specialized search tools in working environments.

These topics cover a broad range of IR technologies, all needed to address the requirements of professional search.

3 FORMAT AND PLANNED ACTIVITIES

The workshop schedule is in Table 1. The workshop will be a mix of keynotes, short oral presentations, and posters, as well as a breakout session in which there is time to focus the discussion on specific topics.

3.1 Keynote speakers

- **Paul Bennett, Microsoft Research: "Search and Recommendation in the Enterprise"**
  Paul Bennett is the Principal Research Manager of the Information & Data Sciences group in Microsoft Research AI. His published research has focused on a variety of topics surrounding the use of machine learning in information retrieval - including ensemble methods and the combination of information sources, calibration, consensus methods for noisy supervision labels, active learning and evaluation, supervised classification and ranking, crowdsourcing, behavioral modeling and analysis, and personalization. Some of his work has been recognized with awards at SIGIR, CHI, and ACM UMAP as well as an ECIR Test of Time Honorable Mention award. Prior to joining MSR in 2006, he completed his dissertation in the Computer Science Department at Carnegie Mellon with Jaime Carbonell and John Lafferty. While at CMU, he also acted as the Chief Learning Architect on the RADAR project from 2005-2006 while a postdoctoral fellow in the Language Technologies Institute.
- **Dave Lewis, Brainspace – A Cyxtera Business: "IR Research Challenges in Legal Discovery and Investigations"**
  David D. Lewis, Ph.D. is Chief Data Scientist at Brainspace, A Cyxtera Business. He leads the data science team developing new information retrieval, machine learning, and natural language processing technologies for legal, investigatory, and intelligence applications. He is a Fellow of the American Society for the Advancement of Science, and won a Test of Time Award from SIGIR in 2017 for his 1994 paper introducing the uncertainty sampling algorithm for active learning.

3.2 Involvement of information professionals

We have a large network of organizations (companies and governments) in which we work together with professionals – the target group of the professional search applications that are the object of our research. In the months before the workshop (May/June 2018) we have distributed a short survey in our network (and we asked the workshop participants to do as well) to gather the experiences of our target group with professional IR: which systems do they use, what do they like about them, what functionalities do they miss, and what aspects/challenges do they think should be on the research roadmap for professional search. The workshop will start with a presentation of the results from the survey (prepared by the organizers), and the requirements will be further addressed in the breakout session.

3.3 Presentations

Papers that are more suitable for an interactive discussion on specific topics, are selected as posters. Papers that are of general interest – such as position papers – are selected as oral presentations. Each speaker will have a 15 minutes time slot, including 5 minutes for questions and discussions.

3.4 Breakout session

In addition to plenary presentations (posters and talks), we will have a breakout session in which attendees will form and self-assign to groups addressing different requirements collected the survey held among information professionals. The goal of this session is to identify and define the major challenges in professional search and prioritise research directions. Each breakout session will be chaired and prepared by a member of the organising team. After the breakout session one participant from each group will report and summarise the conclusions of the group.

3.5 Synergy and future directions

Intended outcome, apart from publishing the proceedings online as CEUR-WS.org workshop proceedings, is a research roadmap and workshop report for SIGIR Forum in which the outcome of the survey, the presentations and discussions in this workshop are summarized. This way, we will draw the research community's attention to the importance of professional search as research domain.

4 SELECTION PROCESS

We have called for research papers and position papers, all of which have four pages ACM style as page limit. The selection of papers is based on relevance, quality and diversity. The workshop is also open for papers presenting work in progress and challenges of new projects. This way, we would like to have interesting discussions about the direction of the research area.

A limited number of research papers will be selected for short presentations (15 minutes including discussion); we allow sufficient room and time for poster presentations.

Each submission will be reviewed by three reviewers and coordinated by one meta reviewer.

5 ORGANIZERS

- Jiyin He, CWI, the Netherlands. Jiyin He is a postdoctoral researcher at the National Research Institute for Mathematics and Computer Science (CWI), the Netherlands. Her research interest is interactive information retrieval, with a focus on the understanding and modeling of user behavioral patterns with respect to diverse types of search interfaces and complex information tasks. She has co-organized several workshops at international conferences including the Search as Learning workshop series at IIiX 2014 and SIGIR 2016.
- Udo Kruschwitz, University of Essex, UK. Udo Kruschwitz is a professor at the University of Essex. His main research interest is the interface between information retrieval (IR) and natural language processing (NLP) with a particular focus on enterprise search (he is the author of "Searching the Enterprise", published in the FnTIR series in 2017 [6]). He has co-organised a range of past events including the Gamification for IR (GamifIR) workshop at SIGIR 2016 and the Industry Track at SIGIR 2012.
- Birger Larsen, Aalborg University, Denmark. Birger Larsen is professor of Information Retrieval and Information Analytics. His research interests are XML IR and user interaction,

domain specific search, understanding user intents and exploiting context in IR, as well as Informetrics/Bibliometrics, citation analysis and academic search. He was co-creator of the iSearch academic search test collection (built on arXiv documents and complex topics from physicists). He has co-organised a number of related workshops, including 'Task Based and Aggregated Search' (ECIR 2012), 'Integrating IR technologies for Professional Search' (ECIR 2013) and 'Bibliometric-enhanced Information Retrieval' (ECIR 2014).
- Tony Russell-Rose, UXlabs, Guildford, UK. Tony Russell-Rose is director of UXLabs, a user experience consultancy and innovations lab specialising in complex search and information access applications. Previously Tony has led R&D teams at Canon, Reuters, Oracle, HP Labs and BT Labs. He is author of "Designing the Search Experience: the Information Architecture of Discovery" (Elsevier, 2013) [10] and publishes widely on search, text analytics and human-computer interaction. He holds a PhD in Computer Science and a first degree in engineering, majoring in human factors.
- Suzan Verberne, Leiden University, the Netherlands. Suzan Verberne is an assistant professor at the Leiden Institute for Advanced Computer Science of Leiden University. She is involved in projects involving text mining and information retrieval in a large number of application domains, many of which in collaboration with professionals [12, 13]. She supervises projects involving information retrieval in the legal domain, the archaeological domain, and the policy-making domain. She has co-organized multiple (national and international) workshops and conferences and was local organization chair for IIiX 2012.
- Arjen P. de Vries, Radboud University, the Netherlands. Arjen P. de Vries holds the chair of Information Retrieval at the Faculty of Science at Radboud University. He coordinates the Data Science specialisation in the Computer Science Master's programme. His research aims to resolve the question how users and systems may cooperate to improve information access, with a specific focus on the value of a combination of structured and unstructured information representations. He co-founded Spinque, a small enterprise that develops search technology for information specialists in application areas including digital heritage, e-commerce and e-health.

6 PROGRAM COMMITTEE

We compiled a program committee with academic researchers as well as practitioners:

- Allan Hanbury
- Charlie Hull
- David Elsweiler
- Diane Kelly
- Elaine Toms
- Farhad Shokraneh
- Jaap Kamps
- Julie Glanville
- Krisztian Balog
- Laura Dietz
- Martin White

- Michail Salampasis
- Mihai Lupu
- Norbert Fuhr
- Peter Cotroneo
- Ray Daley
- Rene Spijker

7 RELATED WORKSHOPS

At ECIR 2013, there was a related workshop "Integrating IR technologies for Professional Search" [11]. The workshop aimed to stimulate exploratory research; to bring together various facets of IR research and to promote the discussion between researchers towards the development of a generalized framework facilitating the integration of IR technologies and search tools into next generation professional search systems. A total of 21 papers were submitted of which 13 were accepted for presentations during the workshop, including a number of position papers [11]. In addition, the workshop had Daria Sorokina (LinkedIn) and Nick Belkin (Rutgers University) as invited speakers.

ECIR 2015 and CHIIR 2017 hosted related workshops on "Supporting Complex Search Tasks" [5]. The workshops did not focus on professional search tasks only, but also covered complex search tasks in 'daily life' information seeking. The 2017 workshop had two keynotes: Mark Hall (Edge Hill University, UK) and Jussi Karlgren (Gavagai, Sweden), the latter presenting the industrial perspective. The workshop had short paper contributions, presented as posters. Eleven submissions were received, of which 9 were accepted. The workshop provided a comprehensive overview of current work on supporting complex tasks in a variety of settings, and fostered new collaboration within the field. There was great support for holding another edition of the workshop at a future conference [5].

The success of these previous workshops, even though not completely on the same topic as this workshop, shows the interests of the community in the topic of professional search. By bringing this workshop to SIGIR in 2018 we hope to reach a larger audience of interested researchers, bringing together people working on many different facets of professional search and concretely developing a roadmap of research directions in the following years.

REFERENCES

[1] Jiyin He, Marc Bron, and Arjen P de Vries. 2013. Characterizing stages of a multi-session complex search task through direct and indirect query modifications. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. ACM, 897–900.

[2] Bradley M Hemminger, Dihui Lu, KTL Vaughan, and Stephanie J Adams. 2007. Information seeking behavior of academic scientists. *Journal of the American Society for Information Science and Technology* 58, 14 (2007), 2205–2225.

[3] Zhipeng Huang, Bogdan Cautis, Reynold Cheng, and Yudian Zheng. 2016. KB-Enabled Query Recommendation for Long-Tail Queries. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. ACM, 2107–2112.

[4] Rezarta Islamaj Dogan, G. Craig Murray, Aurélie Névéol, and Zhiyong Lu. 2009. Understanding PubMedô user search behavior through log analysis. *Database* 2009 (2009). https://doi.org/10.1093/database/bap018

[5] Marijn Koolen, Jaap Kamps, Toine Bogers, Nicholas Belkin, Diane Kelly, and Emine Yilmaz. 2017. Report on the Second Workshop on Supporting Complex Search Tasks. *SIGIR Forum* 51, 1 (Aug. 2017), 58–66. https://doi.org/10.1145/3130332.3130343

[6] Udo Kruschwitz and Charlie Hull. 2017. Searching the Enterprise. *Foundations and Trends in Information Retrieval* 11, 1 (2017), 1–142. https://doi.org/10.1561/1500000053

[7] Carol Collier Kuhlthau and Stephanie L Tama. 2001. Information search process of lawyers: a call forâĂŹjust for meâĂŽinformation services. *Journal of documentation* 57, 1 (2001), 25–43.

[8] Qiang Lu and Jack G. Conrad. 2012. Bringing Order to Legal Documents An Issue-based Recommendation System via Cluster Association. (2012).

[9] Stephann Makri, Ann Blandford, and Anna L Cox. 2008. Investigating the information-seeking behaviour of academic lawyers: From EllisâĂŽs model to design. *Information Processing & Management* 44, 2 (2008), 613–634.

[10] Tony Russell-Rose and Tyler Tate. 2013. *Designing the Search Experience: The Information Architecture of Discovery*. Elsevier (Morgan Kaufmann).

[11] Michail Salampasis, Norbert Fuhr, Allan Hanbury, Mihai Lupu, Birger Larsen, and Henrik Strindberg. 2013. Integrating IR Technologies for Professional Search. In *Advances in Information Retrieval*, Pavel Serdyukov, Pavel Braslavski, Sergei O. Kuznetsov, Jaap Kamps, Stefan Rüger, Eugene Agichtein, Ilya Segalovich, and Emine Yilmaz (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 882–885.

[12] Suzan Verberne, Lou Boves, and Antal van den Bosch. 2016. Information access in the art history domain: Evaluating a federated search engine for Rembrandt research. *DHQ: Digital Humanities Quarterly* 10, 4 (2016).

[13] S. Verberne, E. D'Hondt, A. Van Den Bosch, and M. Marx. 2014. Automatic thematic classification of election manifestos. *Information Processing and Management* 50, 4 (2014), 554–567. https://doi.org/10.1016/j.ipm.2014.02.006