Report on the 2017 ACM SIGIR International Conference Theory of Information Retrieval (ICTIR'17)

Hui Fang[1] **Jaap Kamps**[2] **Evangelos Kanoulas**[2]

Maarten de Rijke[2] **Emine Yilmaz**[3]

[1] University of Delaware, USA
[2] University of Amsterdam, The Netherlands
[3] University College London, UK

Abstract

This is a report on ICTIR'17, which continued its tradition of being the premier forum for presentation of research on theoretical aspects of Information Retrieval (IR). This includes theory in a broad sense, including conceptual papers that explore key concepts, theoretical papers that model concepts and/or relations between concepts, and papers that study theory in experimental or industrial settings.

The importance of information access and information processing in industry and academia is growing in a revolutionary way, making the discussion on fundamental and long term aspects, and their relation to short term success and failure, more urgent than ever before. To highlight the increasingly strong connections between Information Retrieval and its neighboring disciplines, ICTIR'17 explicitly welcomed papers in IR areas that overlap with Human Information Access, Machine Learning, Natural Language Processing and Perception.

ICTIR'17 in Amsterdam was a memorable conference. For those who like numbers: we had 142 attendees (biggest ICTIR ever), and 97 submissions with 52 accepted (largest ICTIR ever), 450/250 euro registration for everything (cheapest ICTIR ever), and for the first time in history the organization and keynotes of a leading IR conference where majority female (but selected for being outstanding researchers, of course).

1 Introduction

This is a report on ICTIR 2017, the 7th International Conference on the Theory of Information Retrieval and the 3rd conference with that name to be fully sponsored by the ACM Special Interest Group on Information Retrieval (SIGIR). ICTIR grew from a series of SIGIR workshops on mathematical and formal methods for IR held during the years 2000 to 2005. In 2007, ICTIR became a full conference and was held held biennially in Europe (2007 Budapest, Hungary; 2009

Table 1: The ACM SIGIR ICTIR Organization Team.

General chairs	Area chairs (cont'd)
✖ Jaap Kamps, University of Amsterdam	IR+Perception track:
✖ Evangelos Kanoulas, University of Amsterdam	✖ Maria Eskevich, Radboud University
✖ Maarten de Rijke, University of Amsterdam	✖ Stefan Rueger, The Open University
PC chairs	Workshops chair
✖ Emine Yilmaz, University College London	✖ Grace Hui Yang, Georgetown University
✖ Hui Fang, University of Delaware	Tutorials chair
Short paper chairs	✖ Josiane Mothe, Institut de Recherche en
✖ Christina Lioma, University of Copenhagen	Informatique de Toulouse
✖ Katja Hofmann, Microsoft	Publicity chair
Area chairs	✖ Julia Kiseleva, University of Amsterdam &
IR+IR track:	UserSat.com
✖ Peter Bruza, Queensland University of Technology	Sponsorship chairs
✖ Lynda Tamine Lechani, Paul Sabatier University	✖ Rianne Kaptein, Crunchr
IR+HCI track:	✖ Esther Smit, University of Amsterdam
✖ Diane Kelly, University of Tennessee at Knoxville	Local organizers
IR+ML track:	✖ Hosein Azarbonyad, University of Amsterdam
✖ Qiaozhu Mei, University of Michigan	✖ Mostafa Dehghani, University of Amsterdam
✖ Alessandro Sordoni, Microsoft Research	✖ Dan Li, University of Amsterdam
IR+NLP track:	✖ Christophe Van Gysel, University of Amsterdam
✖ Alessandro Moschetti, Université de Trento	✖ Team NL

Cambridge, U.K.; 2011 Bertinoro, Italy; 2013 Copenhagen, Denmark). In 2015, ICTIR became an ACM SIGIR sponsored annual conference (2015 Northampton MA, USA; 2016 Newark DE, USA; and now, 2017 in Amsterdam, The Netherlands). This year's conference continues its tradition of being the premier forum for presentation of research on theoretical aspects of Information Retrieval (IR) including (a) conceptual papers that explore key concepts, (b) theoretical papers that model concepts and/or relations between concepts, and (c) papers that study theory in experimental or industrial settings. To highlight the increasingly strong connections between Information Retrieval and neighboring disciplines this year's conference explicitly welcomed papers in IR areas that overlap with Human Information Access, Machine Learning, Natural Language Processing and Perception.

We were honored to welcome ICTIR to Amsterdam, The Netherlands. The Netherlands has a number of strong and vibrant IR groups, e.g. at the University of Amsterdam, Twente University, Radboud University Nijmegen, Centrum Wiskunde en Informatica, Delft University of Technology, and University of Leiden and counts about 150 IR researchers, the largest number per capita in the world. These groups are successfully engaged in both theoretical and practical work. They are young groups consisting of enthusiastic people, with a track record in high quality publications.

The organization team of ICTIR'17 is summarized in Table 1. Evangelos, Jaap and Maarten acted as general chairs, and Emine and Hui as PC chairs. The PC had a number of area chairs, both ensure we attract core IR papers, as well as papers on neighboring areas dealing with the same core aspects. In particular, Lynda and Peter (Core IR area, IR & IR), Diane (IR & CHI); Alessandro S. and Qiaozhu (IR & NLP); Allessandro M. (IR & ML); and Maria and Stefan (IR & Perception). The short paper PC was chaired by Christina and Katja. The official record of the

conference is the ACM DL proceedings [6], entitled officially the *"Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval (ICTIR 2017)"*

The rest of this report will follow the program structure of the conference roughly. The workshop started with a day of tutorials and workshops (Section 2). Next, in Section 3) we summarize the three keynotes and the panel on *"the end of IR as you know it?"* Section 4 discusses the contributed papers, highlighting the winners of the best paper awards. We end with some final observations and comments in Section 5.

2 Tutorials and Workshops

There were three tutorials and two workshops at ICTIR, all held on the Sunday preceding the main conference, addressing core and emerging topics. These events generated a significant number of additional attendees of ICTIR, including about 40 workshop/tutorial day only registrations.

2.1 Differential privacy for information retrieval

Grace Hui Yang and Sicong Zhang [13] gave a tutorial on *"Differential Privacy for Information Retrieval."*

The aim of the tutorial was to offer an introduction to differential privacy, one of the most advanced techniques in privacy research, and provides necessary set of theoretical knowledge for applying privacy techniques in IR. IR research and practice have extensively utilized personalization to advance its state-of-the-art, requiring the use of users' personal information, contextual information and other sensitive and private information. In response, researchers on privacy and on IR teamed up to work on "privacy-preserving IR." Specifically, differential privacy is a technique that provides strong privacy guarantees for data protection. Theoretically, it aims to maximize the data utility in statistical datasets while minimizing the risk of exposing individual data entries to any adversary. Differential privacy has been applied across of a wide range of applications in database, data mining, and IR.

2.2 Bandit algorithms in interactive information retrieval

Dorota Glowacka [5] gave a tutorial on *"Bandit Algorithms in Interactive Information Retrieval."*

The aim of this tutorial was to provide an overview of the various applications of bandit algorithms in information retrieval as well as issues related to their practical deployment and performance in real-life systems/applications. The multi-armed bandit problem models an agent that simultaneously attempts to acquire new knowledge (exploration) and optimize his decisions based on existing knowledge (exploitation). The agent attempts to balance these competing tasks in order to maximize his total value over the period of time considered. There are many practical applications of the bandit model, such as clinical trials, adaptive routing or portfolio design. Over the last decade there has been an increased interest in the development of new bandit algorithms for specific problems in information, such as diverse document ranking, news recommendation or ranker evaluation.

2.3 Efficiency and effectiveness trade-offs in learning to rank

Claudio Lucchese and Franco Maria Nardini [8] gave a tutorial on *"Efficiency/Effectiveness Trade-offs in Learning to Rank."*

The aim of this tutorial was a hands-on introduction to efficiency/effectiveness trade-offs in Learning to Rank. In the last years, learning to rank (LtR) had a significant influence on several tasks in the Information Retrieval field, with large research efforts coming both from the academia and the industry. Indeed, efficiency requirements must be fulfilled in order to make an effective research product deployable within an industrial environment. The evaluation of a model can be too expensive due to its size, the features used and several other factors. This tutorial discusses the recent solutions that allow to build an effective ranking model that satisfies temporal budget constrains at evaluation time. For more information please visit the tutorial website at `http://learningtorank.isti.cnr.it`.

2.4 Learning next generation rankers

Nicola Ferro, Claudio Lucchese, Maria Maistro, and Raffaele Perego [4] organized a workshop on *"LEARning Next gEneration Rankers"* (LEARNER 2017).

The aim of LEARNER'17 was to investigate new solutions for Learning to Rank (LtR). It identified some research areas related to LtR which are of actual interest and which have not been fully explored yet. The workshop solicited the submission of position papers on novel LtR algorithms, on evaluation of LtR algorithms, on dataset creation and curation, and on domain specific applications of LtR. A key element was to look for novel contributions to LtR which focus on foundational and conceptual aspects, which need to be properly framed and modeled, in the coming years.

The workshop brought together a range of academic people interested in IR, ML and related application areas. More information about the LEARNER'17 workshop is found online at `http://learner2017.dei.unipd.it/`.

2.5 Search-oriented conversational AI

Mikhail Burtsev, Aleksandr Chuklin, Julia Kiseleva, and Alexey Borisov [2] organized a workshop on *"Search-Oriented Conversational AI"* (SCAI 2017).

The aim of SCAI'17 was to investigate emerging conversational search approaches. There is a gradual shift towards searching and presenting the information in a conversational form. Chatbots, personal assistants in our phones and eyes-free devices are being used increasingly more for different purposes, including information retrieval and exploration. On the other side, information retrieval empowers dialogue systems to answer questions and to get context for assisting the user in her tasks. With the recent success of deep learning in different areas of natural language processing, this appears to be the right foundation to power search conversationalization. While there is a significant progress in building goal-oriented dialogue systems and open-domain chit-chat bots, more remains to be done for theory and practice of conversation-based search and search-based dialogues.

This workshop brought together AI/Deep Learning specialists on one hand, and search/IR specialists on the other hand, to lay the ground for search-oriented conversational AI and establish

future directions and collaborations. More information about the SCAI'17 workshop is found online at `https://deeppavlov.github.io/scai17/`.

3 Keynotes and Panel

We were blessed with three great invited speakers and a panel, which greatly helped frame the discussion with an eye to move it forward.

3.1 Information retrieval meets game theory

The opening keynote was given by Oren Kurland [7] on "*Information Retrieval Meets Game Theory.*"

Oren Kurland discussed how (algorithmic) game theory can be used to analyze some aspects of the competitive search setting. In competitive search settings such as the Web, authors of documents may have an incentive to have their documents highly ranked for certain queries. This can drive corpus dynamics as documents may be manipulated in response to induced rankings (e.g., by applying search engine optimization). Such post-ranking corpus effects are not directly modeled in ad hoc retrieval models and, more generally, are not accounted for by the formal foundations of retrieval paradigms. In his talk, Kurland first discussed the probability ranking principle which is the theoretical underpinning of most ad hoc retrieval methods. As it turns out, the PRP is sub-optimal in competitive settings. In addition, he discussed some initial theoretical and empirical results regarding the strategic behavior of document authors in competitive retrieval settings, specifically with respect to the foundations of classical ad hoc retrieval models.

3.2 The evolution of computational advertising

The second keynote was given by Suju Rajan [10] on "*The Evolution of Computational Advertising.*"

Suju Rajan highlighted the recent research challenges in the field of computational advertising, and how the field is evolving to incorporate ideas from areas such as reinforcement learning, econometrics, deep learning, and large-scale recommender systems. Machine learning literature on computational advertising typically tends to focus on the simplistic click-through-rate prediction problem which while being relevant is the tip of the iceberg in terms of the research challenges in the field. There have been several recent efforts, shaped by the realities of a complex ad ecosystem, to develop models that try to better encapsulate the journey of an ad from its impression to possibly leading to a purchase.

3.3 How to exploit relationships to improve predictions

The third keynote was given by Jennifer Neville [9] on "*How to Exploit Relationships to Improve Predictions.*"

Jennifer Neville outlined some of the algorithmic and statistical challenges that arise due to partially-observed, large-scale networks, and describe methods for semi-supervised learning, latent-variable modeling, and active sampling to address the challenges. The popularity of social

networks and social media has increased the amount of information available about users' behavior online—including current activities, and interactions with followers, friends, and family. This rich relational information can be used to improve predictions even when individual data is sparse, since the characteristics of friends are often correlated. Although this type of network data offer several opportunities to improve predictions about users, the characteristics of online social network data also present a number of challenges to accurately incorporate the network information into machine learning systems.

3.4   The end of IR as we know it?

Finally, there was a panel on "the end of IR as we know it?" that raised considerable excitement and follow up discussion.

The panel was chaired by Arjen de Vries, with Norbert Fuhr, Leif Azzopardi, Grace Yang, Fiana Raiber, and Jimmy Lin as panelists. The panel started with initial five minutes statements from each panelist, followed with a free discussion the rest of the audience, as well as between the panelists. Arguments covered a lot ground—too much to summarize in a paragraph. Norbert reflected on the state of current IR, and the need for methodological improvements in order to make it a first class citizen. Leif posited the changes needed when a large fraction of the users of information access systems were "intelligent agents." Grace rebutted the "revolutionary" novelty of AI applications, pointing out how many new approaches are based on the same principles studied for long. Fiana reminded of earlier revolutions that would have changed the world, such as TV replacing existing media when introduced, but existing media adapted and still exist. Jimmy embraced the new AI approaches, while pointing out the idiocy of taken them to the extreme (learn to learn to learn to ... to learn). Whilst the panel didn't reach a singular consensus solution, the exposed views and animated discussion were very interesting, and generally optimistic about the future of IR in the current move to data-driven "AI" and deep learning.

4   Contributed Papers

We requested the submission of full and short papers. Overall, ICTIR received a record number of 97 submissions, and accepted 52 (a 54% acceptance rate). Broken down over submission types, we received 54 full papers and accepted 27 (50%) and 43 short papers and accepted 25 (58%).

While we aspired to be inclusive and explicitly encouraged the submission of papers beyond the traditional IR topics, we did also uphold usual high quality standards with a strong PC consisting of senior researchers supplemented with the young and hopeful. That is, the resulting accept/reject threshold is similar to other main IR conferences, and the relatively high acceptance rate is due to the selective nature of the submissions to the conference, with the overwhelming majority of submissions coming from leading groups in the field.

Every accepted paper made an interesting contribution to the program, and space restrictions force us to limit ourselves in the remainder of this section to those recognized as the "best papers."

4.1 Best paper award

The *Best Paper Award at the 2017 ACM SIGIR International Conference Theory of Information Retrieval* (ICTIR'17) was given to Marco Ferrante, Nicola Ferro, and Silvia Pontarollo [3] for their paper entitled "*Are IR Evaluation Measures on an Interval Scale?*"

Marco Ferrante, Nicola Ferro, and Silvia Pontarollo formally investigate whether, or not, IR evaluation measures are on an interval scale, which is needed to safely compute the basic statistics, such as mean and variance, we daily use to compare IR systems. They face this issue in the framework of the representational theory of measurement and it relies on the notion of difference structure, i.e. a total equi-spaced ordering on the system runs. They found that the most popular set-based measures, i.e. precision, recall, and F-measure are interval-based. In the case of rank-based measures, using a strongly top-heavy ordering, and found that only RBP with p = 1/2 is on an interval scale while RBP for other p values, AP, DCG, and ERR are not. Moreover, using a weakly top-heavy ordering, they found that none of RBP, AP, DCG, and ERR is on an interval scale.

The ICTIR Best Paper Award comes with a cash prize of $ 1,000 graciously funded by the ACM SIGIR.

4.2 Best student paper award

The *Best Student Paper Award at the 2017 ACM SIGIR International Conference Theory of Information Retrieval* (ICTIR'17) was given to Zhiwen Tang and Grace Hui Yang [11] for their paper entitled "*Investigating per Topic Upper Bound for Session Search Evaluation.*"

Zhiwen Tang and Grace Hui Yang investigate session search, which is a complex IR task. As a result, its evaluation is also complex. A great number of factors need to be considered in the evaluation of session search. They include document relevance, document novelty, aspect-related novelty discounting, and user's efforts in examining the documents. Due to increased complexity, most existing session search evaluation metrics are NP-hard. Consequently, the optimal value, i.e. the upper bound, of a metric highly varies with the actual search topics. In Cranfield-like settings such as TREC, scores for systems are usually averaged across all search topics. With undetermined upper bound values, however, it could be unfair to compare IR systems across different topics. This paper addresses the problem by investigating the actual per topic upper bounds of existing session search metrics. Through decomposing the metrics, we derive the upper bounds via mathematical optimization. They show that after being normalized by the bounds, the NP-hard session search metrics are then able to provide robust comparison across various search topics. The new normalized metrics are experimented on official runs submitted to the TREC 2016 Dynamic Domain Track.

The ICTIR Best Student Paper Award also comes with a cash prize of $ 1,000 graciously funded by the ACM SIGIR.

4.3 Best short paper award

The best paper committee found the two top ranked papers of equal merit and decided to announce both papers as shared winners of the *Best Short Paper Award at the 2017 ACM SIGIR International Conference Theory of Information Retrieval* (ICTIR'17). The two papers receiving

the award are, in arbitrary order, Gaurav Baruah and Jimmy Lin [1] for their paper entitled "*The Pareto Frontier of Utility Models as a Framework for Evaluating Push Notification Systems*," and Julián Urbano and Mónica Marrero [12] for their paper entitled "*The Treatment of Ties in AP Correlation.*"

Gaurav Baruah and Jimmy Lin propose a utility-based framework for the evaluation of push notification systems that monitor document streams for users' topics of interest. Our starting point is that users derive either positive utility (i.e., "gain") or negative utility (i.e., "pain") from consuming system updates. By separately keeping track of these quantities, we can measure system effectiveness in a gain vs. pain trade-off space. The Pareto Frontier of evaluated systems represents the state of the art: for each system on the frontier, no other system can offer more gain without more pain. Our framework has several advantages: it unifies three previous TREC evaluations, subsumes existing metrics, and provides more insightful analyses. Furthermore, our approach can easily accommodate more refined user models and is extensible to different information-seeking modalities.

Julián Urbano and Mónica Marrero investigate the Kendall tau and AP correlation coefficients, which are very commonly used to compare two rankings over the same set of items. Even though Kendall tau was originally defined assuming that there are no ties in the rankings, two alternative versions were soon developed to account for ties in two different scenarios: measure the accuracy of an observer with respect to a true and objective ranking, and measure the agreement between two observers in the absence of a true ranking. These two variants prove useful in cases where ties are possible in either ranking, and may indeed result in very different scores. AP correlation was devised to incorporate a top-heaviness component into Kendall tau, penalizing more heavily if differences occur between items at the top of the rankings, making it a very compelling coefficient in Information Retrieval settings. However, the treatment of ties in AP correlation remains an open problem. This paper fills this gap by providing closed analytical formulations of AP correlation under the two scenarios of ties contemplated in Kendall tau. In addition, we developed an R package that implements these coefficients.

5   Conclusions

The conference took place from Sunday October 1st (Workshops and Tutorial day) to Wednesday October 4th, in the cozy CASA Hotel in Amsterdam, combining the conference rooms and as well as accommodation for attendees at the edge of the historic inner-city of Amsterdam. The organizers made particular effort to have an inclusive conference by keeping the registration fees as low as 250 Euro for students and 450 Euro for seniors (ACM members, early rates). This included the full conference package, including workshops/tutorials, the banquet and receptions, all lunches, and—of course—the main conference days. There was a total of over 140 attendees over the whole event, with over a 100 attendees on every single day.

The strong technical program and stimulating keynotes and panel, as discussed above, are the cornerstone of the conference. However, the main point of bringing 100+ researchers from all over the globe together in a single place, is to stimulate discussion, and we devoted particular efforts on this. First, one decision is to have 20 minutes full paper presentations, freeing generous time in the schedule for more interactive discussion. Second, we spent particular effort on the social events to match the quality of the technical program—much in the spirit of the original Greek *symposion* or

symposium[1]—to stimulate continued discussion. Just to give some of the highlights: On Sunday, there were welcome drinks at the end of the workshop/tutorial day. On Monday, there was the welcome reception at the Booking.com HQ. On Tuesday, there was the conference banquet with five courses of exception food in *De Kas* (the greenhouse). On Wednesday, there were further drinks and snacks at the collocated "Search Engines Amsterdam" meeting—particularly for those who didn't go on the free canal boat tour offered by the City of Amsterdam. And these are only the highlights...

There are two special thanks we want to give in the main text rather than hide in an acknowledgment section, as they were crucial for making the conference it's great success. First, we thanks all the sponsors supporting ICTIR, not only for their support in financial support allow more students (and seniors) to attend, but in particular for their participation in the technical program. Thanks to ACM SIGIR, Bloomberg, Booking.com, Facebook, City of Amsterdam, Google, Microsoft, Netherlands Organization for Scientific Research (NWO), Netherlands Research School for Information and Knowledge Systems (SIKS), University of Amsterdam, and the Information Science Netherlands Community (WGI). Special thanks are to Melanie Müller and `http://Booking.com/` for hosting us in their HQ for amazing hospitality, amazing views, and amazing tech talks on "challenges in online recommendation" and "conversational search at Booking.com." Second, we were fortunate to be supported by the large group of IR PhD students "Team NL" that helped in every stage of the planning and the conference. We want to mention in particular: Alex Olieman, Ana Lucic, Chuang Wu, Chang Li, Dan Li, Dat Tien Nguyen, Harrie Oosterhuis, Hosein Azarbonyad, Julia Kiseleva, Kaspar Beelen, Petra Best, and Ziming Li. And we probably forgot some of the other volunteers who helped in great number to make ICTIR'17 such a memorable experience.

References

[1] G. Baruah and J. Lin. The pareto frontier of utility models as a framework for evaluating push notification systems. In Kamps et al. [6], pages 253–256. URL `http://doi.acm.org/10.1145/3121050.3121089`.

[2] M. Burtsev, A. Chuklin, J. Kiseleva, and A. Borisov. Search-oriented conversational AI

---

[1] `https://en.wikipedia.org/wiki/Symposium`

(SCAI). In Kamps et al. [6], pages 333–334. URL `http://doi.acm.org/10.1145/3121050.3121111`.

[3] M. Ferrante, N. Ferro, and S. Pontarollo. Are IR evaluation measures on an interval scale? In Kamps et al. [6], pages 67–74. URL `http://doi.acm.org/10.1145/3121050.3121058`.

[4] N. Ferro, C. Lucchese, M. Maistro, and R. Perego. Learning next generation rankers (LEARNER 2017). In Kamps et al. [6], pages 331–332. URL `http://doi.acm.org/10.1145/3121050.3121110`.

[5] D. Glowacka. Bandit algorithms in interactive information retrieval. In Kamps et al. [6], pages 327–328. URL `http://doi.acm.org/10.1145/3121050.3121108`.

[6] J. Kamps, E. Kanoulas, M. de Rijke, H. Fang, and E. Yilmaz, editors. *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval, ICTIR 2017, Amsterdam, The Netherlands, October 1-4, 2017*, 2017. ACM. URL `http://dl.acm.org/citation.cfm?id=3121050`.

[7] O. Kurland. Information retrieval meets game theory. In Kamps et al. [6], page 1. URL `http://doi.acm.org/10.1145/3121050.3121079`.

[8] C. Lucchese and F. M. Nardini. Efficiency/effectiveness trade-offs in learning to rank. In Kamps et al. [6], pages 329–330. URL `http://doi.acm.org/10.1145/3121050.3121109`.

[9] J. Neville. How to exploit relationships to improve predictions. In Kamps et al. [6], page 167. URL `http://doi.acm.org/10.1145/3121050.3121081`.

[10] S. Rajan. The evolution of computational advertising. In Kamps et al. [6], page 99. URL `http://doi.acm.org/10.1145/3121050.3121080`.

[11] Z. Tang and G. H. Yang. Investigating per topic upper bound for session search evaluation. In Kamps et al. [6], pages 185–192. URL `http://doi.acm.org/10.1145/3121050.3121069`.

[12] J. Urbano and M. Marrero. The treatment of ties in AP correlation. In Kamps et al. [6], pages 321–324. URL `http://doi.acm.org/10.1145/3121050.3121106`.

[13] G. H. Yang and S. Zhang. Differential privacy for information retrieval. In Kamps et al. [6], pages 325–326. URL `http://doi.acm.org/10.1145/3121050.3121107`.