

The Second Workshop on Knowledge Graphs and Semantics for Text Retrieval, Analysis, and Understanding (KG4IR)

Laura Dietz
University of New Hampshire
Durham, NH, USA
dietz@cs.unh.edu

Jeff Dalton
University of Glasgow
Glasgow, United Kingdom
jeff.dalton@glasgow.ac.uk

Chenyan Xiong
Carnegie Mellon University
Pittsburgh, PA, USA
cx@cs.cmu.edu

Edgar Meij
Bloomberg
London, United Kingdom
edgar.meij@acm.org

ABSTRACT

Semantic technologies such as controlled vocabularies, thesauri, and knowledge graphs have been used throughout the history of information retrieval for a variety of tasks. Recent advances in knowledge acquisition, alignment, and utilization have given rise to a body of new approaches for utilizing knowledge graphs in text retrieval tasks and it is therefore time to consolidate the community efforts and study how such technologies can be employed in information retrieval systems in the most effective way. It is also time to start and deepen the dialogue between researchers and practitioners in order to ensure that breakthroughs, technologies, and algorithms in this space are widely disseminated. The goal of this workshop is to bring together and grow a community of researchers and practitioners who are interested in using, aligning, and constructing knowledge graphs and similar semantic resources for information retrieval applications.

CCS CONCEPTS

• **Information systems** → *Presentation of retrieval results*;

KEYWORDS

Knowledge graphs, Information Retrieval, Entity Linking, Entity Retrieval, Entity-Oriented Search

ACM Reference Format:

Laura Dietz, Chenyan Xiong, Jeff Dalton, and Edgar Meij. 2018. The Second Workshop on Knowledge Graphs and Semantics for Text Retrieval, Analysis, and Understanding (KG4IR). In *SIGIR '18: 41st International ACM SIGIR Conference on Research and Development in Information Retrieval, July 8-12, 2018, Ann Arbor, MI, USA*. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3209978.3210196>

1 OVERVIEW

The past decade has witnessed the emergence of publicly available knowledge graphs (KGs) such as DBpedia, Freebase, WikiData, and

also proprietary KGs such as Google's Knowledge Graph and Microsoft's Satori. The availability of large, heterogeneous knowledge graphs and grounding techniques such as entity linking gave rise to successful improvements for many information retrieval (IR) tasks. Entity retrieval from KGs has been a long-studied IR research topic for many years [1, 4, 12, 26]. The utilization of entities and KG information for text-centric retrieval tasks has also drawn wide attention recently.

In particular, the semantics encoded in KGs are effectively integrated in query representation [5, 8, 11, 22], retrieval models [5, 15, 18], learning-to-rank [21], and generic representations [18]. Recent top-tier IR conferences feature many other papers on search systems are exploiting knowledge graphs [9, 13, 14, 19, 20, 23–25]. Some of these systems have also participated in TREC evaluations [16] or led to development of new TREC tracks [6]. Despite these successes, the utilization of knowledge graphs and semantics in information retrieval is still in its infancy and the community is actively working on how the different kinds of semantics in KGs can be utilized to improve end-to-end IR tasks.

In addition, many questions around improving text retrieval, analysis, and understanding remain unanswered, including the following: (i) How do we most effectively (and efficiently!) identify and align entities and structured information in KGs to unstructured text? (ii) What is the role of entities and semantics in modern intelligent systems? (iii) How can neural information systems effectively leverage knowledge graphs?

Answering these open questions will lead to the next generation of more intelligent IR systems. These questions require researchers with varying backgrounds to come together from separate communities working on different aspects of knowledge graphs, semantics, and IR. The research on knowledge acquisition and alignment often overlooks their applications in text-centric information retrieval systems. Similarly, in IR research knowledge graph acquisition and alignment technologies are often treated as black boxes. This mismatch in research goals leads to combined systems that do not live up to their full potential. The main goal of this workshop is to bring together researchers on knowledge graph population and alignment with IR researchers. Together we will work towards the ultimate goals of: (i) acquiring semantics for focused application needs, (ii) aligning efforts around KGs and texts towards end-to-end usage, and (iii) building IR systems that more effectively utilize KGs, semantics, and alignment techniques.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
SIGIR'18, July 8–12, 2018, Ann Arbor, MI, USA
© 2018 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-5657-2/18/07.
<https://doi.org/10.1145/3209978.3210196>

Our workshop at last year's SIGIR was a success with more than 50 registered attendees [7]. The program included six great keynote speakers, diverse panel speakers, and eight accepted paper presentations. The participants spanned both academia and industry and provided extremely positive feedback. Given the open research topics that remain, we aim to continue the workshop series. Moreover, there have been several key advances in the field of KG4IR in the past year. As advocated by our previous workshop, more integrated systems that perform the acquisition, alignment, and utilization of KG jointly have been developed [9, 13, 14, 17, 19, 20, 24]. We are also encouraged to see that the utilization of knowledge graphs and semantics influenced a broader set of IR applications, for example, dialog systems [10] and recommendation systems [2, 3]. The TREC Complex Answer Retrieval [6] track provides a basis for evaluating both entity centric and text centric approaches on an openly available large-scale benchmark. With the rapid growth in this area, it is important to build momentum in this field by hosting the second KG4IR workshop that will facilitate the exchange of new ideas, encourage deeper dialogue, and also to expand the community of IR researchers involved in utilizing knowledge graphs.

2 THEME AND PURPOSE

This workshop focuses on the end-to-end utilization of knowledge graphs and semantics in text retrieval and IR applications. The scope includes the **acquisition, alignment, and utilization** of knowledge graphs and semantic resources for the purpose of optimizing end-to-end performance of a system that responds to a user's information need.

- **Acquisition** refers to knowledge graph population and semantic resource construction with a particular emphasis on enabling IR-related techniques and applications. Examples include domain/task-specific knowledge graph construction, knowledge representation, and query-time knowledge extraction.
- **Alignment** refers to the semantic annotation of text. This includes entity linking of keyword queries as well as query-focused relation extraction for satisfying information needs. It also includes information integration, ontology matching, entity search, and knowledge graph selection based on an information need.
- **Utilization** refers to using knowledge graphs and semantics in text-centric retrieval tasks. Examples include utilizing the knowledge graph to improve ad hoc document retrieval, factoid question answering, understanding user queries, knowledge-focused diversification and summarization of SERPs, tracking of events, and retrieval of complex answers.

This workshop aims to not only bring together researchers and practitioners in related fields, but more importantly experts on graphs, semantics, and information extraction with experts on text retrieval, analysis, and query understanding. The goal is to discuss ongoing research and best practices with the aim of addressing open research challenges. Our mission includes: (i) facilitating discussions between researchers working on acquisition, alignment, and utilization of KG technologies for IR applications, (ii) serving as an incubator for long-term research on resource construction and end-to-end utilization, and (iii) collecting datasets, tasks, and evaluations to study IR algorithms in the context of KGs.

3 UPDATES FROM LAST WORKSHOP

We hosted the first KG4IR workshop as a full-day workshop at SIGIR 2017 [7], attracting more than 50 registrants, 6 invited talks, 8 contributed papers, and a joint discussion panel with the OKBQA workshop. This year we aim to continue our vision of promoting KG4IR and serve as a platform for widening participation in the community. In addition, there are several major upgrades and differences that motivated us to propose this workshop for a second iteration.

New Applications — There are new advances in utilizing KGs for many new applications, e.g., dialog systems, recommendation systems, and reading comprehension systems. The development of task-focused knowledge acquisition and alignment technologies are also promising trends.

New Techniques — The development of new techniques and the continued growth of deep learning approaches to learning the relevance of output from relation extraction and subgraph prediction for KGs. These techniques lead to new approaches to extract, align, and utilize knowledge and semantics while raising different needs for structured semantics. Our workshop will provide a platform for researchers to communicate and discuss the recent advancements of these new technologies and their influence on KG4IR.

New Modalities — The importance of multi-modal retrieval has continued to grow. There are now rich KGs that go beyond text to also include music, images, and video. These have very different retrieval mechanisms as ways of utilizing KGs. Advances in acquisition and alignment using deep learning approaches are allowing significant new applications to explore in KG4IR.

Cross-fertilization with the Semantics track at SIGIR — This year SIGIR has broadened its scope with a multi-track system that features a track on "AI, Semantics, and Dialog." This is a great opportunity for our workshop as one of our goals is to facilitate discussion and build a community around incorporating semantics in IR systems.

IRJ Special Issue on Knowledge Graphs and Semantics in Text Analysis and Retrieval — The Information Retrieval Journal is soliciting submissions to the Special Issue on Knowledge Graphs and Semantics in Text Analysis and Retrieval. The submission deadlines between the workshop papers and special issue papers are synchronized to offer appropriate venues for research in all stages. Authors of the accepted special issue papers will be speaking about their work at the workshop.

Tutorial on Utilizing Knowledge Graphs for Text-centric Information Retrieval — Furthermore, we are offering a SIGIR conference tutorial on foundations and best practices on utilizing knowledge graphs for text-centric information retrieval. The goal is to jump-start any information retrieval researcher who is interested in contributing to this line of work in the future.

4 RELATED WORKSHOPS

While a large number of workshops and tutorials on KGs are held in different venues, they focus either on KG construction, semantic search, or other applications; none of them have the focus on IR that we advocate. Our main focus lies on the use of KGs for retrieving, analyzing, and understanding text. By focusing on text as

the main target, our workshop differentiates itself from the general semantic search and knowledge graph topics. Nevertheless, we aim for a broadly inclusive community, encouraging participation from researchers with expertise in information extraction, semantic search, and text retrieval with the goal of promoting joint research.

Exploiting Semantic Annotations for IR (ESAIR) is the most closely related workshop as it focuses on the exploitation of semantic annotations in information retrieval. It was successfully hosted at CIKM for many years, but the series is currently on hold. Our workshop builds upon the results from ESAIR with a broader scope on graph-structured knowledge with respect to construction, alignment, and utilization for IR tasks. Jeff Dalton was one of the organizers of this series.

Automated Knowledge Base Construction (AKBC) focuses on population techniques using information extraction and information integration. It is a very successful workshop and has been held in multiple conference such as NAACL, NIPS, and CIKM. Its focus on construction is complementary to our focus on utilization.

Entity Recognition and Disambiguation (ERD) focused on entity linking in long (document) and short (query) texts. It has been held only once, at SIGIR 2014. In fact, many of the query entity linking systems used and presented in the ERD workshop have served as a critical step in KG4IR research. We also note that many representatives from this community are included in our workshop's program committee.

KB Construction, Reasoning and Mining (KBCOM) is a new workshop co-located with WSDM 2018 with a broad scope that includes knowledge graph construction, reasoning, and applications. The scope is similar to AKBC, but also includes reasoning and applications.

Open Knowledge Base and Question Answering (OKBQA) and Question Answering over Linked Data (QALD) are workshops on question answering using the structure of knowledge graphs. In contrast, the KG4IR workshop focuses on longer texts and natural language queries.

Open Knowledge Network (OKN) is a US-based network created by Yolanda Gill, RV Guha, Sharat Israni, Andrew Moore, and Louiqa Rashid with the aim to further the development of KGs available to everyone. Both Laura Dietz and Edgar Meij are active members of this network.

5 WORKSHOP PROGRAM

The full-day workshop features a combination of talks, poster sessions, and panel discussion.

Keynote talks are selected that are well-known in the field, maintaining a healthy balance between industry and academia and providing a fresh perspective on the topic. Presentations on contributed position papers will give an overview of ongoing efforts. Furthermore, we solicited talk abstracts, inviting community members to share their ideas and perspectives. Authors of papers in our "KG4IR" special issue in the Information Retrieval Journal are encouraged to give an overview of their work.

Table 1: Confirmed program committee members.

Name	Affiliation
Esraa Ali	ADAPT research centre
Mohammad Aliannejadi	University of Lugano
Avishek Anand	L3S Research Center
Bogdan Arsintescu	LinkedIn
Marc Bron	Schibsted
Tongfei Chen	Johns Hopkins University
Bhavana Dalvi	Allen Institute for AI
Arjen de Vries	Radboud University and Spinque
John Foley	University of Massachusetts
Ingo Frommholz	University of Bedfordshire
Faegheh Hasibi	NUST
Xiangnan He	National University of Singapore
Johannes Hoffart	Ambiverse and Max Planck Institute
Ioana Hulpus	Mannheim University
Rose Catherine Kanjirathinkal	Carnegie Mellon University
Alexander Kotov	Wayne State
Huang Lifu	Rensselaer Polytechnic Institute
Kwan Hui Lim	University of Melbourne
Xitong Liu	Google
Bhaskar Mitra	Microsoft
Federico Nanni	Mannheim University
Giulio Ermanno Pibiri	University of Pisa and ISTI-CNR
Jay Pujara	University of Southern California
Pushpendre Rastogi	Johns Hopkins University
Hadas Raviv	Technion
Achim Rettinger	Karlsruhe Institute of Technology
Benjamin Roth	Ludwigs Maximilians Universitaet
Pedro Saleiro	University of Chicago
Bahareh Sarrafzadeh	University of Waterloo
Michael Schuhmacher	BASF
Yu Su	University of California
Camilo Thorne	University of Stuttgart
Salvatore Trani	ISTI-CNR
Suzan Verberne	Leiden University
Nikos Voskarides	University of Amsterdam
Lydia Weiland	Mannheim University
Arie Wahyu Wijayanto	Tokyo Institute of Technology
Jun Xu	Chinese Academic of Science
Hai-Tao Yu	University of Tsukuba
Hamed Zamani	University of Massachusetts
Yuan Zhang	Peking University

Finally, the presentations are complemented with a poster session and a panel discussion. The panel discussion will facilitate dialogue on a diverse range of the issues raised in this workshop, possibly including the following:

- Lessons learned using knowledge graphs and semantics in IR
- Demands for KG and alignment techniques for utilization
- Mechanisms to encourage task-oriented development of KG's and corresponding techniques
- Future benchmarks for KG-aware alignment and retrieval models
- Possible ways to facilitate future research and reproducibility in this area

The final workshop program is available at <https://kg4ir.github.io>.

6 ORGANIZERS

Prof. Dr. Laura Dietz is an Assistant Professor at University of New Hampshire. Before that she was a research scientist at Mannheim University and University of Massachusetts after graduating from the Max Planck Institute for Informatics. Her research focuses on information retrieval on knowledge-centric information needs. Her scientific contributions span from query expansion

with entities to the prediction of influences in citation graphs. She coordinates the TREC Complex Answer Retrieval track.

Chenyan Xiong is a final-year Ph.D. student at the Language Technologies Institute, Carnegie Mellon University. His research focuses on using knowledge graphs and semantics to improve text understanding in search engines. He has published many KG4IR papers in top venues SIGIR, CIKM, ICTIR, and WWW.

Dr. Jeff Dalton is Lecturer in Information Retrieval at the University of Glasgow. Previously, he was a Software Engineer Google, where projects included the Google Assistant Natural Language Understanding and the Knowledge Vault efforts on automatic knowledge graph construction. He completed his Ph.D. at the University of Massachusetts Amherst with James Allan in the Center for Intelligent Information Retrieval. His research focuses on the intersection of Information Retrieval and Natural Language Processing. He publishes papers on using knowledge graphs for IR at leading conferences including SIGIR, CIKM, and others.

Dr. Edgar Meij is a senior scientist at Bloomberg where he also leads a team focusing on graph analytics and semantic technologies. Before this, he was a research scientist at Yahoo Labs and a postdoc at the University of Amsterdam, where he also obtained his Ph.D. He regularly teaches at the (post-)graduate level, including university courses and conference tutorials, e.g., at EAACL, ICTIR, SIGIR, WWW, and WSDM. His research focuses on all applications and aspects of knowledge graphs, entity linking, and semantic search.

Acknowledgements

C. Xiong is funded on NSF solicitation IIS-1422676. L. Dietz is supported by a scholarship of the Eliteprogramm for Postdocs of the Baden-Württemberg Stiftung. Any opinions, findings expressed in this material are the authors' and do not necessarily reflect those of the sponsor.

REFERENCES

- [1] Krisztian Balog and Robert Neumayer. 2013. A test collection for entity search in dbpedia. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval (SIGIR 2013)*. ACM, 737–740.
- [2] Rose Catherine and William Cohen. 2016. Personalized recommendations using knowledge graphs: A probabilistic logic programming approach. In *Proceedings of the 10th ACM Conference on Recommender Systems*. ACM, 325–332.
- [3] Rose Catherine, Kathryn Mazaitis, Maxine Eskenazi, and William Cohen. 2017. Explainable Entity-based Recommendations with Knowledge Graphs. *arXiv preprint arXiv:1707.05254* (2017).
- [4] Jing Chen, Chenyan Xiong, and Jamie Callan. 2016. An Empirical Study of Learning to Rank for Entity Search. In *Proceedings of SIGIR 2016*. ACM, 737–740.
- [5] Jeffrey Dalton, Laura Dietz, and James Allan. 2014. Entity Query Feature Expansion using Knowledge Base Links. In *Proceedings SIGIR 2014*. ACM, 365–374.
- [6] Laura Dietz, Manisha Verma, Filip Radlinski, and Nick Craswell. 2018. TREC Complex Answer Retrieval Overview. In *Text REtrieval Conference (TREC)*.
- [7] Laura Dietz, Chenyan Xiong, and Edgar Meij. 2017. The First Workshop on Knowledge Graphs and Semantics for Text Retrieval and Analysis (KG4IR). In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2017)*. ACM, 1427–1428.
- [8] Faezeh Ensan and Ebrahim Bagheri. 2017. Document retrieval model through semantic linking. In *Proceedings of WSDM 2017*. ACM, 181–190.
- [9] Dario Garigliotti, Faegheh Hasibi, and Krisztian Balog. 2017. Target Type Identification for Entity-Bearing Queries. (2017).
- [10] Marjan Ghazvininejad, Chris Brockett, Ming-Wei Chang, Bill Dolan, Jianfeng Gao, Wen-tau Yih, and Michel Galley. 2017. A knowledge-grounded neural conversation model. *arXiv preprint arXiv:1702.01932* (2017).
- [11] Faegheh Hasibi, Krisztian Balog, and Svein Erik Bratsberg. 2015. Entity Linking in Queries: Tasks and Evaluation. In *Proceedings of ICTIR 2015*. ACM, 171–180.
- [12] Faegheh Hasibi, Krisztian Balog, and Svein Erik Bratsberg. 2016. Exploiting Entity Linking in Queries for Entity Retrieval. In *Proceedings of the 2016 ACM on International Conference on the Theory of Information Retrieval*. ACM, 209–218.
- [13] Faegheh Hasibi, Krisztian Balog, and Svein Erik Bratsberg. 2017. Dynamic Factual Summaries for Entity Cards. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 773–782.
- [14] Amina Kadry and Laura Dietz. 2017. Open Relation Extraction for Support Passage Retrieval: Merit and Open Issues. In *Proceedings of SIGIR 2017*.
- [15] Xitong Liu and Hui Fang. 2015. Latent entity space: A novel retrieval approach for entity-bearing queries. *Information Retrieval Journal* 18, 6 (2015), 473–503.
- [16] Xitong Liu, Peilin Yang, and Hui Fang. 2014. Entity came to rescue - Leveraging entities to minimize risks in web search. In *Proceedings of The 23rd Text Retrieval Conference, (TREC 2014)*. NIST.
- [17] Federico Nanni, Simone Paolo Ponzetto, and Laura Dietz. 2017. Building entity-centric event collections. In *Digital Libraries (JCDL), 2017 ACM/IEEE Joint Conference on*. IEEE, 1–10.
- [18] Hadas Raviv, Oren Kurland, and David Carmel. 2016. Document retrieval using entity-based language models. In *Proceedings of SIGIR 2016*. ACM, 65–74.
- [19] Rebecca Sharp, Mihai Surdeanu, Peter Jansen, Marco A Valenzuela-Escárcega, Peter Clark, and Michael Hammond. 2017. Tell Me Why: Using Question Answering as Distant Supervision for Answer Justification. In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*. 69–79.
- [20] Nikos Voskarides, Edgar Meij, and Maarten de Rijke. 2017. Generating descriptions of entity relationships. In *European Conference on Information Retrieval*. Springer, 317–330.
- [21] Chenyan Xiong and Jamie Callan. 2015. ESDRank: Connecting query and documents through external semi-structured data. In *Proceedings of CIKM 2015*. ACM, 951–960.
- [22] Chenyan Xiong and Jamie Callan. 2015. Query expansion with Freebase. In *Proceedings of ICTIR 2015*. ACM, 111–120.
- [23] Chenyan Xiong, Jamie Callan, and Tie-Yan Liu. 2016. Bag-of-Entities representation for ranking. In *Proceedings of the sixth ACM International Conference on the Theory of Information Retrieval (ICTIR 2016)*. ACM, 181–184.
- [24] Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Eduard H. Hovy. 2017. JointSem: Combining Query Entity Linking and Entity based Document Ranking. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (CIKM 2017)*. 2391–2394.
- [25] Chenyan Xiong, Russell Power, and Jamie Callan. 2017. Explicit semantic ranking for academic search via knowledge graph embedding. In *Proceedings WWW 2017*. ACM, 1271–1279.
- [26] Nikita Zhiltsov, Alexander Kotov, and Fedor Nikolaev. 2015. Fielded Sequential Dependence Model for Ad-Hoc Entity Retrieval in the Web of Data. In *Proceedings of SIGIR 2015*. ACM, 253–262.