The First Workshop on Knowledge Graphs and Semantics for Text Retrieval and Analysis (KG4IR)

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

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

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

ABSTRACT

Knowledge graphs have been used throughout the history of information retrieval for a variety of tasks. Technological advances in knowledge acquisition and alignment technology from the last few years gave rise to a body of new approaches for utilizing knowledge graphs in text retrieval tasks. It is therefore time to consolidate the community efforts in studying how knowledge graph technology can be employed in information retrieval systems in the most effective way. It is also time to start a dialogue with researchers working on knowledge acquisition and alignment to ensure that resulting technologies and algorithms meet the demands posed by information retrieval tasks. The goal of this workshop is to bring together a community of researchers and practitioners who are interested in using, aligning, and constructing knowledge graphs and similar semantic resources for information retrieval applications.

1 OVERVIEW

The past decade has witnessed the emergence of publicly available knowledge graphs (KGs) such as DBpedia, Freebase, and WikiData and also proprietary KGs such as Google's Knowledge Graph and Microsoft's Satori. The availability of large knowledge graphs and grounding techniques have given rise to successful approaches for many information retrieval (IR) tasks. It has been shown that heterogeneous information in knowledge graphs and entity annotations can help to significantly improve information retrieval tasks. In particular, the semantics encoded in knowledge graphs have been effectively integrated in various aspects of IR systems, including query representation [7, 8, 13, 26], retrieval models [7, 18, 21], learning-to-rank [25], and generic representations [21].

This workshop focuses on the end-to-end utilization of knowledge graphs and semantics in text retrieval and IR-related downstream applications. The scope includes suggestions for **acquisition**, **alignment**, and **utilization** of knowledge graphs and semantic resources for the purpose of optimizing end-to-end performance of information retrieval systems.

Acquisition includes (but is not limited to) knowledge graph population and semantic resource construction with a special focus on enabling IR-related techniques and applications. Examples include domain/task-specific knowledge graph construction, knowledge representation, and query-time knowledge extraction.

Alignment includes (but is not limited to) the semantic annotation process such as entity linking of short keyword queries or 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 includes (but is not limited to) using knowledge graphs and semantics in text-centric tasks. Examples are utilizing the knowledge graph to improve document retrieval, question answering, factoid search, dialogue systems, event tracking, and retrieval of complex answers.

We aim to bring together researchers and practitioners within the IR field and related communities to discuss ongoing research and best practices with the goal of addressing open research challenges of this area. The missions of KG4IR include the following.

- Facilitate meetings for researchers working on acquisition, alignment, and utilization of knowledge graphs for text retrieval and analysis.
- Serve as an incubator for long-term research on resource construction and end-to-end utilization.
- Act as a nursery for future tasks, applications, and evaluations that benefit from knowledge graphs and text retrieval.
- Provide a voice and platform to the community.

2 RELATED WORK

Acquisition. Knowledge graphs are either semi-manually constructed (Cyc [16], Freebase [4]) or machine-generated, for example using Wikipedia (DBpedia[1], Yago[2]). Supervised and unsupervised relation extraction algorithms [10, 20] provide an alternative for the construction or augmentation of knowledge graphs [9]. Through cross-references, knowledge graphs form the linked open data cloud.[3]

Alignment. A key ingredient for utilizing knowledge graphs are algorithms that align knowledge graph elements to natural language text. Given text passages, entity linking algorithms identify mentions of knowledge graph entities. While popular algorithms like "TagMe!" [11] can be applied to many documents, specialized entity linking algorithms for tweets and queries have received much attention [3, 6, 19]. A byproduct of relation extraction algorithms [10, 20] is an alignment between relation expressions in the text to an edge in the knowledge graph. A related task is to find sentences that describe a relation [24].

Entity search techniques aim to retrieve knowledge graph elements in response to an information need [14] and different variations on using fielded retrieval models or entity links in documents

SIGIR '17, August 07–11, 2017,Shinjuku, Tokyo, Japan
© 2017 Copyright held by the owner/author(s). 978-1-4503-5022-8/17/08…$15.00
DOI: http://dx.doi.org/10.1145/3077136.3084371

[1] http://www.dbpedia.org
[2] http://www.mpi-inf.mpg.de/yago
[3] http://linkeddata.org

have been developed [1, 23, 28]. Entity search is also utilized in text-centric retrieval systems as a query-to-entity alignment component [7, 25], i.e., as a retrieval task in its own right.

Utilization. The utilization of knowledge graphs in text retrieval and analysis tasks has been a recent breakthrough in information retrieval. The rich semantics stored in knowledge graphs have provided additional indicators for various components of search systems. Although semantic search traditionally focused on search within knowledge graphs [12], nowadays it is commonly generalized to include any "search with meaning" [2].

One utilization is to enrich query and document representations with entity links [21] and embedding spaces [8, 27] to derive new similarity measures. A general latent space approach is to first associate the query with relevant entities (using entity search), then use entity-centric features for document ranking [7, 18, 25, 26]. Structural graph features provide additional information for the retrieval of short documents [5].

Open areas. There are many opportunities we have only begun to study. While entity linking yields immediate success, utilizing relation extraction for text retrieval is much more difficult [15]. Treating different aspects of entities appropriately is a promising yet underexplored direction [17, 22]. Finally, widespread application of knowledge-centric retrieval techniques hinges on the advancement of knowledge graphs for new domains such as science,[4] domain-specific entity linking,[5] and complex answer retrieval.[6]

3 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 fifth-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 SIGIR, CIKM, ICTIR, and WWW.

Dr. Edgar Meij is a senior scientist at Bloomberg L.P. 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 has published 60+ peer-reviewed papers at top international venues such as SIGIR, WSDM, ISWC, and CIKM on all applications and aspects of knowledge graphs, entity linking, and semantic search.

4 ACKNOWLEDGEMENTS

This research was supported in part by National Science Foundation (NSF) grant IIS-1422676. Any opinions, findings, and conclusions in this paper are the authors' and do not necessarily reflect those of the sponsors.

REFERENCES

[1] Krisztian Balog, Marc Bron, and Maarten De Rijke. 2011. Query modeling for entity search based on terms, categories, and examples. *ACM Transactions on Information Systems (TOIS)* 29, 4 (2011), 22.

[2] Hannah Bast, Björn Buchhold, Elmar Haussmann, and others. 2016. Semantic Search on Text and Knowledge Bases. *Foundations and Trends® in Information Retrieval* 10, 2-3 (2016), 119–271.

[3] Roi Blanco, Giuseppe Ottaviano, and Edgar Meij. 2015. Fast and space-efficient entity linking for queries. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*. ACM, 179–188.

[4] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of SIGMOD 2008*. ACM, 1247–1250.

[5] Christopher Boston, Hui Fang, Sandra Carberry, Hao Wu, and Xitong Liu. 2014. Wikimantic: Toward effective disambiguation and expansion of queries. *Data & Knowledge Engineering* 90 (2014), 22–37.

[6] David Carmel, Ming-Wei Chang, Evgeniy Gabrilovich, Bo-June (Paul) Hsu, and Kuansan Wang. 2014. ERD'14: Entity recognition and disambiguation challenge. In *Proceedings of SIGIR 2014*. ACM.

[7] Jeffrey Dalton, Laura Dietz, and James Allan. 2014. Entity Query Feature Expansion using Knowledge Base Links. In *Proceedings SIGIR 2014*. ACM, 365–374.

[8] Faezeh Ensan and Ebrahim Bagheri. 2017. Document retrieval model through semantic linking. In *Proceedings of WSDM 2017*. ACM, 181–190.

[9] T. Mitchell et al. 2015. Never-Ending Learning. In *Proceedings of AAAI 2015*.

[10] Oren Etzioni, Anthony Fader, Janara Christensen, Stephen Soderland, and others. 2011. Open information extraction: The second generation. In *Proceedings of IJCAI 2011*.

[11] Paolo Ferragina and Ugo Scaiella. 2010. Fast and accurate annotation of short texts with Wikipedia pages. *arXiv preprint arXiv:1006.3498* (2010).

[12] Ramanathan Guha, Rob McCool, and Eric Miller. 2003. Semantic search. In *Proceedings of WWW 2003*. 700–709.

[13] Faegheh Hasibi, Krisztian Balog, and Svein Erik Bratsberg. 2015. Entity Linking in Queries: Tasks and Evaluation. In *Proceedings of ICTIR 2015*. ACM, 171–180.

[14] Faegheh Hasibi, Fedor Nikolaev, Chenyan Xiong, Krisztian Balog, Svein Bratsberg, E, Alexander Kotov, and Jamie Callan. 2017. Word-Entity Duet Representations for Document Ranking. In *Proceedings of SIGIR 2017*. ACM, To Appear.

[15] Amina Kadry and Laura Dietz. 2017. Open Relation Extraction for Support Passage Retrieval: Merit and Open Issues. In *Proceedings of SIGIR 2017*.

[16] Douglas B Lenat and Ramanathan V Guha. 1989. *Building large knowledge-based systems; representation and inference in the Cyc project.* Addison-Wesley Longman Publishing Co., Inc.

[17] Peng Li, Jing Jiang, and Yinglin Wang. 2010. Generating templates of entity summaries with an entity-aspect model and pattern mining. In *Proc. ACL*. 640–649.

[18] 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.

[19] Edgar Meij, Wouter Weerkamp, and Maarten de Rijke. 2012. Adding Semantics to Microblog Posts. In *WSDM*.

[20] Mike Mintz, Steven Bills, Rion Snow, and Dan Jurafsky. 2009. Distant supervision for relation extraction without labeled data. In *Proceedings of ACL 2009*. 1003–1011.

[21] Hadas Raviv, Oren Kurland, and David Carmel. 2016. Document retrieval using entity-based language models. In *Proceedings of SIGIR 2016*. ACM, 65–74.

[22] Ridho Reinanda, Edgar Meij, and Maarten de Rijke. 2015. Mining, ranking and recommending entity aspects. In *Proceedings of SIGIR 2015*. ACM.

[23] Michael Schuhmacher, Laura Dietz, and Simone Paolo Ponzetto. 2015. Ranking Entities for Web Queries Through Text and Knowledge. In *Proceedings of CIKM 2015*. ACM, 1461–1470.

[24] Nikos Voskarides, Edgar Meij, Manos Tsagkias, Maarten de Rijke, and Wouter Weerkamp. 2015. Learning to Explain Entity Relationships in Knowledge Graphs. In *Proceedings of ACL 2015*. ACL, 564–574.

[25] Chenyan Xiong and Jamie Callan. 2015. EsdRank: Connecting query and documents through external semi-structured data. In *Proceedings of CIKM 2015*. ACM, 951–960.

[26] Chenyan Xiong and Jamie Callan. 2015. Query expansion with Freebase. In *Proceedings of ICTIR 2015*. ACM, 111–120.

[27] 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.

[28] 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.

[4] http://www.springernature.com/gp/researchers/scigraph

[5] http://ir.nist.gov/feiii

[6] http://trec-car.cs.unh.edu