



Special issue on knowledge graphs and semantics in text analysis and retrieval

Laura Dietz¹ · Chenyan Xiong^{2,3} · Jeff Dalton⁴ · Edgar Meij⁵

© Springer Nature B.V. 2019

1 Topical overview

Knowledge graphs are an effective way to store semantics in a structured format that is easily used by computer systems. In the past few decades, work across different research communities led to scalable knowledge acquisition techniques for building large-scale knowledge graphs. The result is the emergence of large publicly available knowledge graphs (KGs) such as DBpedia (Lehmann et al. 2014), Freebase (Bollacker et al. 2008), and others. While knowledge graphs are designed to support a wide set of different applications, this special issue focuses on the use case of text retrieval and analysis.

Utilizing knowledge graphs for text analysis requires effective alignment techniques that associate segments of unstructured text with entries in the knowledge graph, for example using entity extraction and linking algorithms (Carmel et al. 2014; Mendes et al. 2011; Blanco et al. 2015). A wide range of approaches that combine query-document representations and machine learning repeatedly demonstrate significant improvements for such tasks across diverse domains (Dalton et al. 2014; Liu and Fang 2015; Hasibi et al. 2015; Xiong and Callan 2015; Raviv et al. 2016; Ensan and Bagheri 2017; Xiong et al. 2017). The goal of this special issue is to summarize recent progress in research and practice in

✉ Laura Dietz
dietz@cs.unh.edu

Chenyan Xiong
Chenyan.Xiong@Microsoft.com

Jeff Dalton
jeff.dalton@glasgow.ac.uk

Edgar Meij
emeij@bloomberg.net

¹ University of New Hampshire, 33 Academic Way, Durham, NH 03824, USA

² Microsoft Research AI, One Microsoft Way, Redmond, WA 98052, USA

³ Carnegie Mellon University, 5705 GHC, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA

⁴ University of Glasgow, School of Computing Science, Glasgow G12 8QQ, UK

⁵ Bloomberg L.P., City Gate House, 39-45 Finsbury Square, London EC2A 1HD, UK

constructing, grounding, and utilizing knowledge graphs and similar semantic resources for text retrieval and analysis applications. The scope includes *acquisition*, *alignment*, and *utilization* of knowledge graphs and other semantic resources for the purpose of optimizing end-to-end performance of information retrieval systems.

2 Articles in this special issue

Extraction errors, missing relations, and yet unobserved entity-facts lead to never-complete knowledge graphs. Entity linking and missing link prediction methods make use of entity set expansion as an underlying component. Rastogi et al. (2018) suggest a neural auto-encoder framework for entity set expansion, which particularly excels in difficult situations when entities appear in multiple sentences and many entities are contained in the query.

For similar reasons, large knowledge graphs undergo continuous improvement cycles. With the increased coverage of entities and facts it is likely that the ontological type structure needs to be modified. To compare and assess different versions of an evolving knowledge graph, Nayak et al. (2018) develop a similarity measure for knowledge hierarchies.

Several works focus on search queries that ask for entities. Garigliotti et al. (2018) focus on how to best exploit type information that is present queries for entity retrieval. Different approaches to matching type information to entities are explored ranging between top-level types and most specific types in the hierarchy.

Sawant et al. (2019) develop a method to exploit entity mentions and type hints to reason about the target entity. The approach combines information from a knowledge graph with information from a full text corpus with entity link annotations. These heterogeneous pieces of information are integrated through multiple convolutional networks to gracefully handle cases where exact word matches fail.

Jimmy et al. (2018) explore KG-aware retrieval models in the context of consumer health search. A key challenge of the consumer health domain is the vocabulary mismatch problem. To this end, Jimmy et al. (2018) using knowledge graphs to match consumer queries to the medical terminology used in documents.

Addressing the population of knowledge graphs with text, where each entity is associated with several facets (i.e., headings on Wikipedia pages), MacAvaney et al. (2018) suggest an approach for identifying relevant text passages. Where a traditional information retrieval approach would include such facets into the keyword query, MacAvaney et al. (2018) address the problem of low-utility facets, these are frequent headings whose words are unlikely to appear in relevant passages.

These six contributions discuss important aspects that arise when knowledge graphs are populated and used in the context of information retrieval. Each of them makes a strong case for integrating unstructured text and structured knowledge across a wide range of domains and applications.

3 Selection process and community

For this special issue we selected six articles out of 23 submissions. Each article was reviewed by at least three reviewers and underwent at least one revision.

More literature on how to effectively use of knowledge graphs in information retrieval can be found in the proceedings of the KG4IR Workshop series. Inquiries about methods and topical discussions can be directed to the discussion forum kg4ir@googlegroups.com.

References

- Blanco, R., Ottaviano, G., & Meij, E. (2015). Fast and space-efficient entity linking for queries. In *Proceedings of the 8th ACM international conference on web search and data mining* (pp. 179–188). ACM.
- Bollacker, K., Evans, C., Paritosh, P., Sturge, T., & Taylor, J. (2008). Freebase: A collaboratively created graph database for structuring human knowledge. In *Proceedings of SIGMOD 2008* (pp. 1247–1250). ACM.
- Carmel, D., Chang, M. W., Gabrilovich, E., Hsu, B. J. P., & Wang, K. (2014). ERD'14: Entity recognition and disambiguation challenge. In *Proceedings of SIGIR 2014*. ACM.
- Dalton, J., Dietz, L., & Allan, J. (2014). Entity query feature expansion using knowledge base links. In *Proceedings SIGIR 2014* (pp. 365–374). ACM.
- Ensan, F., & Bagheri, E. (2017). Document retrieval model through semantic linking. In *Proceedings of WSDM 2017* (pp. 181–190). ACM.
- Garigliotti, D., Hasibi, F., & Balog, K. (2018). Identifying and exploiting target entity type information for ad hoc entity retrieval. *Information Retrieval Journal*. <https://doi.org/10.1007/s10791-018-9346-x>.
- Hasibi, F., Balog, K., & Bratsberg, S. E. (2015). Entity linking in queries: Tasks and evaluation. In *Proceedings of ICTIR 2015* (pp. 171–180). ACM.
- Jimmy, Z. G., & Koopman, B. (2018). Payoffs and pitfalls in using knowledge-bases for consumer health search. *Information Retrieval Journal*. <https://doi.org/10.1007/s10791-018-9344-z>.
- Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P. N., et al. (2014). DBpedia—A large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web Journal*, 6, 167.
- Liu, X., & Fang, H. (2015). Latent entity space: A novel retrieval approach for entity-bearing queries. *Information Retrieval Journal*, 18(6), 473–503.
- MacAvaney, S., Yates, A., Cohan, A., Soldaini, L., Hui, K., Goharian, N., et al. (2018). Overcoming low-utility facets for complex answer retrieval. *Information Retrieval Journal*. <https://doi.org/10.1007/s10791-018-9343-0>.
- Mendes, P. N., Jakob, M., García-Silva, A., & Bizer, C. (2011). DBpedia spotlight: Shedding light on the web of documents. In *Proceedings of the 7th international conference on semantic systems* (pp 1–8). ACM.
- Nayak, G., Dutta, S., Ajwani, D., Nicholson, P., & Sala, A. (2018). Automated assessment of knowledge hierarchy evolution: Comparing directed acyclic graphs. *Information Retrieval Journal*. <https://doi.org/10.1007/s10791-018-9345-y>.
- Rastogi, P., Poliak, A., Lyzinski, V., & Van Durme, B. (2018). Neural variational entity set expansion for automatically populated knowledge graphs. *Information Retrieval Journal*. <https://doi.org/10.1007/s10791-018-9342-1>.
- Raviv, H., Kurland, O., & Carmel, D. (2016). Document retrieval using entity-based language models. In *Proceedings of SIGIR 2016* (pp. 65–74). ACM.
- Sawant, U., Garg, S., Chakrabarti, S., & Ramakrishnan, G. (2019). Neural architecture for question answering using a knowledge graph and web corpus. *Information Retrieval Journal*. <https://doi.org/10.1007/s10791-018-9348-8>.
- Xiong, C., & Callan, J. (2015). EsdRank: Connecting query and documents through external semi-structured data. In *Proceedings of CIKM 2015* (pp. 951–960). ACM.
- Xiong, C., Callan, J., & Liu, T. Y. (2017). Word-entity duet representations for document ranking. In *Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval* (pp. 763–772). ACM.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.