

Utilizing Knowledge Bases in Text-centric Information Retrieval

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

Alexander Kotov
Wayne State University
Detroit MI, USA
kotov@wayne.edu

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

Abstract

General-purpose knowledge bases are increasingly growing in terms of depth (content) and width (coverage). Moreover, algorithms for entity linking and entity retrieval have improved tremendously in the past years. These developments give rise to a new line of research that exploits and combines these developments for the purposes of text-centric information retrieval applications. This tutorial focuses on a) how to retrieve a set of entities for an ad-hoc query, or more broadly, assessing relevance of KB elements for the information need, b) how to annotate text with such elements, and c) how to use this information to assess the relevance of text. We discuss different kinds of information available in a knowledge graph and how to leverage each most effectively.

We start the tutorial with a brief overview of different types of knowledge bases, their structure and information contained in popular general-purpose and domain-specific knowledge bases. In particular, we focus on the representation of entity-centric information in the knowledge base through names, terms, relations, and type taxonomies. Next, we will provide a recap on ad-hoc object retrieval from knowledge graphs as well as entity linking and retrieval. This is essential technology, which the remainder of the tutorial builds on. Next we will cover essential components within successful entity linking systems, including the collection of entity name information and techniques for disambiguation with contextual entity mentions. We will present the details of four previously proposed systems that successfully leverage knowledge bases to improve ad-hoc document retrieval. These systems combine the notion of entity retrieval and semantic search on one hand, with text retrieval models and entity linking on the other. Finally, we also touch on entity aspects and links in the knowledge graph as it can help to understand the entities' context.

This tutorial is the first to compile, summarize, and disseminate progress in this emerging area and we provide both an overview of state-of-the-art methods and outline open research problems to encourage new contributions.

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).

ICTIR '16 September 12-16, 2016, Newark, DE, USA

© 2016 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-4497-5/16/09.

DOI: <http://dx.doi.org/10.1145/2970398.2970441>

Presenters

Prof. Dr. Laura Dietz is a professor at University of New Hampshire, where she teaches Information Retrieval and Machine Learning. Before that she was working in the Data and Web Science group at Mannheim University, with Prof. Bruce Croft and Prof. Andrew McCallum at University of Massachusetts, and obtained her Ph.D. from the Max Planck Institute for Informatics. Her research focuses on text processing and information retrieval with knowledge bases. Her scientific contributions span from entity linking to the prediction of influences in citation graphs. In this tutorial, she will cover her seminal publication on entity query feature expansion and her work on finding relevant relations.

Prof. Dr. Alexander Kotov is an Assistant Professor in the Department of Computer Science at Wayne State University. His general research interests lie at the intersection of information retrieval, textual data mining and health informatics. Before joining Wayne State, he was a post-doctoral fellow at Emory University working with Prof. Eugene Agichtein. Dr. Kotov obtained his PhD from the University of Illinois at Urbana-Champaign, under the supervision of Professor ChengXiang Zhai. At Wayne State has been teaching graduate courses on Information Retrieval and NoSQL databases as well as undergraduate courses. This tutorial will cover his work on using semantic networks for query expansion and his recent work on entity retrieval from knowledge graphs.

Dr. Edgar Meij is a senior scientist at Bloomberg. 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 EACL 2009, SIGIR 2013, WWW 2013, and WSDM 2014. His research focuses on all applications and aspects of knowledge graphs, entity linking, and semantic search. This tutorial will cover his contributions on entity aspect mining and finding support passages for relations.

Acknowledgments

This work was in part funded by the Elitepostdoc program of the BW-Stiftung and the University of New Hampshire.