

# User Access Models to Event-Centric Information

Sara Abdollahi

Supervised by Prof. Dr. Wolfgang Nejdl, Prof. Dr. Elena Demidova and Dr. Simon Gottschalk  
L3S Research Center, Leibniz Universität Hannover  
Hannover, Germany  
abdollahi@L3S.de

## ABSTRACT

Events such as terrorist attacks and *Brexit* play an important role in the research of social scientists and Digital Humanities researchers. They need innovative user access models to event-centric information which support them throughout their research. Current access models such as recommendation and information retrieval methods often fail to adequately capture essential features of events and provide acceptable access to event-centric information. This PhD research aims to develop efficient and effective user access models to event-centric information by leveraging well-structured information in Knowledge Graphs. The goal is to tackle the challenges researchers encounter during their research in a workflow, from exploratory search to accessing well-defined and complete information collections. This paper presents the specific research questions and presents the approach and preliminary results.

## CCS CONCEPTS

• Information systems → Semantic web description languages; Learning to rank.

## KEYWORDS

knowledge graphs, entity recommendation, information retrieval, event collection

### ACM Reference Format:

Sara Abdollahi. 2022. User Access Models to Event-Centric Information. In *Companion Proceedings of the Web Conference 2022 (WWW '22 Companion)*, April 25–29, 2022, Virtual Event, Lyon, France. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3487553.3524193>

## 1 INTRODUCTION

The world is constantly facing events such as terrorist attacks, Brexit, and the migrant crisis, which have resulted in a vast amount of event-centric information on the web. Researchers, namely digital humanities researchers and social scientists, analyze the significant events that influence and shape our societies. They require innovative user access models that support them in exploring information, retrieving answers to their questions and efficiently obtaining an overview of relevant information. Furthermore, depending on the event's global vs. local impact, they may need help

Permission to make digital or hard copies of all or part 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 components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).  
*WWW '22 Companion*, April 25–29, 2022, Virtual Event, Lyon, France.

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.  
ACM ISBN 978-1-4503-9130-6/22/04...\$15.00  
<https://doi.org/10.1145/3487553.3524193>

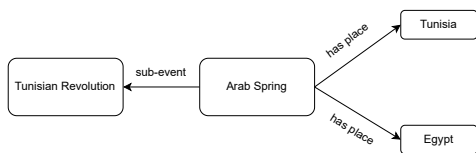
with the language barrier in studying events covered in multilingual sources.

Despite their improvements over the last years, current user access models such as recommendation and information retrieval algorithms fail to provide satisfactory results in exploring and studying event-centric multilingual information.

Some of the challenges of the current interaction methods are as follows:

- (1) The different perceptions of entities and events across language communities may significantly impact users' information needs. However, current existing entity recommendation methods do not consider language-specific context.
- (2) Events have specific temporal and spatial features and relations among each other, such as sub-event relations (e.g., *Tunisian Revolution* is a sub-event of *Arab Spring*) that distinguish them from other topics in information retrieval. Take *Arab Spring* and *Tunisian Revolution* as an example again; current methods might miss related documents to *Tunisian Revolution* if they do not contain *Arab Spring* in the text.
- (3) Web Archives reflecting the perception of events during their happening time, are primary sources for studying events. However, the constantly growing structure of Web Archives is one of the main challenges. Furthermore, the returned results of a Web Archive search are often overwhelming due to their quantity, potential redundancy, and irrelevance.
- (4) Although some Web Archives offer manually collected event-centric collections to solve this issue, not only do these collections not cover all necessary information related to a specific event, they are not updated either.

This PhD research aims to develop efficient and effective access models to event-centric information. We aim to address the mentioned challenges and provide innovative solutions. The plan of this thesis includes four steps to support researchers at different stages of their research from exploratory search to information retrieval and access to event collections. The core part of this work relies on the EventKG knowledge graph [7], a multilingual resource incorporating event-centric information. First, we create a language-specific dataset that provides relevance scores necessary for training recommendation algorithms. Next, we develop a language-specific event recommendation method that generates a list of related events considering language perspectives. We proceed to develop a document retrieval algorithm specialized for event-related queries. Finally, we use the mentioned document retrieval algorithm to improve accessing the information on Web Archives and building event collections that cover various event features.



**Figure 1: An example knowledge graph with information about two events**

## 2 PROBLEM AND RESEARCH QUESTIONS

The main goal of this thesis is to design and develop user access models to event-centric information. Such models aim at supporting users (e.g., researchers) on different steps of researching events. This work’s core hypothesis is that by leveraging the representation of events and their features in knowledge graphs, we could capture related information in different information sources like documents.

This thesis is formed around the following research questions:

- (RQ1) How can we develop event-centric and multilingual datasets to help us with training and evaluation of interaction methods?
- (RQ2) How can we develop recommendation methods that suggest events in a language-specific context?
- (RQ3) How can current document retrieval algorithms incorporate event-centric information extracted from knowledge graphs to improve final results?
- (RQ4) How can we create event collections from large sources like Web Archives that cover essential features of events as well as their temporal course?

This work largely relies on knowledge graphs that are formally defined as follows:

*Definition 2.1.* A knowledge graph is a network of nodes representing real world entities (e.g., persons, places and events) and their semantic types, properties, and relationships.

Figure 1 demonstrates an example knowledge graph representing information for the event *Arab Spring*. As shown in this figure, this event has happened in *Tunisia* and *Egypt* and *Tunisian Revolution* is its sub-event.

## 3 RELATED WORKS

This PhD research relates to several research areas, which we briefly discuss in this section.

### 3.1 Cross-lingual Analysis

Analyzing multilingual web information has been studied in several works such as [13, 23]. Wikidata, the most important publicly available multilingual source in terms of the number of languages and users, has been analyzed in many studies [2, 6, 16] which explore cross-lingual differences in it. Rogers [22] analyzes bias and linguistic points of view regarding a controversial event. Hale [9] analyzes the editing behavior in a multilingual setting and Samoilenko et al. [24] show the reflection of cross-cultural similarities in the process of collective archiving knowledge on Wikipedia.

### 3.2 Entity and Event Recommendation

Entity recommendation is the task of recommending a ranked list of entities to the user query. Blanco et al. [4] present the Spark system that provides a ranking of the entities related to the user query combining several signals from a variety of data sources. Moreover, Ni et al. [19] propose an embedding-based entity recommendation framework for Wikipedia using an architecture of multiple layered graphs. Zhang et al. [33] propose time-aware entity recommendation (TER), which allows users to restrict their interests of entities to a customized time range. Tran et al. [27] extend TER by incorporating topic and time and propose contextual relatedness among entities using embedding techniques. The user interest, preference and serendipity have also been studied in some other works [3, 12]. Unlike the approaches mentioned above, we leverage structured information from knowledge graphs and consider information needs and the specific context of language communities.

### 3.3 Document Retrieval

Document retrieval is the task of finding stored documents according to their relatedness to the user query. Using external information from knowledge graphs in document retrieval, has gained attention in recent years. Liu et al. [15] present the Entity-Duet Neural Ranking Model (EDRM), which introduces knowledge graphs to neural search systems. Xiong et al. [30] propose a framework for utilizing knowledge bases in information retrieval. This work considers words and entities from knowledge graphs equally important and represents the query and documents. Xiong et al. [31] introduce JointSem, which combines query entity linking and entity-based document ranking. This model is trained by an LTR model using document relevance labels. Xiong et al. Furthermore, Explicit Semantic Ranking [32], a ranking technique leverages knowledge graph embedding and ranks entities and documents based on their semantic connections from their knowledge graph embedding. In our work, we use information from EventKG to expand the initial query and employ it in BERT-based document retrieval methods.

### 3.4 Exploring Web Archives

Exploring Web Archives and returning related results to users’ information needs have been studied recently and many works have tried to improve the effectiveness of user experience and final results. Some works [14, 25] have focused on search interfaces. Some other approaches consider temporal aspects in searching and providing results: Holzmann et al. [11] present an approach to searching Web Archives based on temporal link graphs and corresponding anchor texts. Moreover, historian’s information seeking behavior has been analyzed in [26]. Furthermore, Holzmann et al. [10] propose a tag-based temporal search engine for Web Archives. Using non-content evidence such as file headers, links and URL strings instead of full content of documents has been studied in [28]. In terms creating document collections from Web Archives, Nanni et al. [17] use relevant concepts and entities as relevance indicators and employ embedding techniques in an LTR method to provide a final rank list of documents. Gossen et al. [5] present an approach based on focused crawling of Web Archives to create event-centric collections without requiring any full-text indices.

Compared with these event collection works, our proposed approach tries to cover important features and information of an event, especially sub-events in the final collections.

## 4 APPROACH

We intend to design a unified workflow that supports researchers at different stages of their research as follows:

- **Exploratory Search:** At the first stages of their research, researchers need to explore information to gain insight into the topic.
- **Information Retrieval:** Next, researchers seek documents relevant to their information needs.
- **Accessible Event Collections:** Finally, they need a collection of documents as a comprehensive reference covering all essential information related to an event.

To answer the research questions in the previous section and design the workflow as mentioned earlier, we will provide the following contributions:

- (C1) A multilingual dataset that reflects the language-specific relevance of events and their relations. This dataset aims to provide a reference source to train and evaluate language-specific recommendation methods and addresses RQ1.
- (C2) A language-specific event recommendation method (as a solution for RQ2) that aims to recommend events relevant to the user query in a language-specific context.
- (C3) A document retrieval algorithm that employs event-centric features from knowledge graphs into document retrieval task to get event-related documents. This component solves RQ3.
- (C4) An event collection algorithm that utilizes the former document retrieval algorithm on Web Archives to build comprehensive collections covering all essential event features. This component addresses RQ4.

Figure 2 demonstrates the components and external resources used to develop them. We start by creating a language-specific dataset (C1) in Step 1, then we develop a language-specific event recommendation method (C2) in step 2. We proceed to step 3 by creating an event document retrieval method (C3) and finally we build event collections (C4) at step 4. In terms of external resources, we utilize Wikipedia Clickstream, EventKG, Ms-Marco dataset and Web Archives. These components aim at improving results compared to their baselines regarding corresponding evaluation metrics.

## 5 METHODOLOGY

In this section, we describe the different steps involved in the overview shown in Figure 2.

### 5.1 Step 1: Language-specific Dataset

At the first step, we aim to create a language-specific dataset that provides relevance scores among entities and events in a language context. Such values are derived by normalizing click counts in the Wikipedia Clickstream of one language regarding click counts in all languages[1]. We use this dataset later to train and create interaction methods like recommendation methods. We use two data sources to create this dataset: 1) the Wikipedia clickstream

that reflects real-world user interactions with events and their relations within language-specific Wikipedia editions; and 2) the EventKG knowledge graph, a multilingual resource incorporating event-centric information extracted from several large-scale knowledge graphs.

### 5.2 Step 2: Language-specific Event Recommendation

In the second step, we develop an event recommendation method that generates a list of related events to the user’s query in a language-specific context. In order to train this model, we create language-specific embeddings of entities and events. Then, we incorporate these embeddings along with spatio-temporal and link-based features extracted from EventKG to train a Learning to Rank model that ranks events regarding their relevance to the query entity. We employ the dataset created in the previous step to train this LTR model.

In terms of evaluation in this step, we conduct two setups to assess the quality of results of language-specific recommendations. First, we assess the recommendation results using the relevance labels from the created language-specific dataset in step 1 as a ground truth. Second, we conduct a user study to assess the recommendation quality from the user’s perspective. We evaluate the language-specific event recommendation method against three baselines, which are publicly available, reproducible, and are adaptable to the novel task of language-specific event recommendation: Milne-Witten [29], which models relatedness of entities based on links, as well as the embedding-based methods such as DeepWalk [20] and Node2Vec [8].

### 5.3 Step 3: Event Document Retrieval

A search for an event such as *Arab Spring* should return documents covering all aspects of it and its temporal evolution. Accordingly, we intend to develop an event document retrieval method in step 3 that considers event-centric information from EventKG as external information to expand query and incorporates it into the document retrieval task. First, we create Ms-Marco-Event, a training dataset using EventKG and Ms-Marco [18]. Microsoft Machine Reading Comprehension (MS MARCO) - is a large-scale real-world reading comprehension dataset in which queries are collected through Bing or Cortana, and the related documents to the queries are annotated by crowd-sourcing. To create this specialized dataset, we choose queries centering around events using textual information of all events from EventKG.

This document retrieval method builds upon neural document re-ranking algorithms utilizing large pretrained language models such as BERT. The main contribution is incorporating event-centric features from EventKG, namely triples, sub-event information, temporal, and other essential information to expand initial query. Our intuition is that event-centric information extracted from EventKG plays an essential role in representing queries and would improve the final ranking results.

To evaluate the document retrieval task, we employ the previously mentioned Ms-Marco-Event dataset and will compare the results against three groups of document retrieval methods as follows:

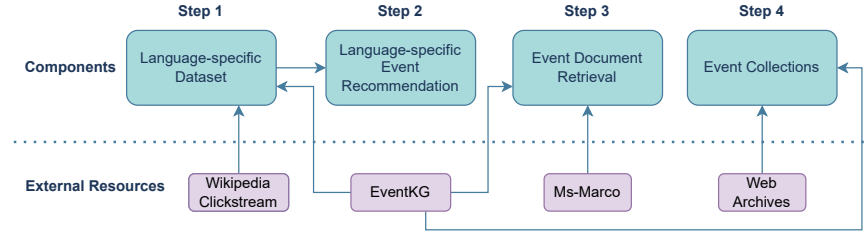


Figure 2: Steps of the developing user access models to event-centric information

- (1) Recent neural document retrieval algorithms that employ large pretrained language models like BERT,
- (2) Document retrieval algorithms that use knowledge graphs as external information,
- (3) Traditional information retrieval algorithms such as BM25 [21].

#### 5.4 Step 4: Event Collections

Finally, in the last step, we will employ event document retrieval method, developed in step 3, to build comprehensive collections from Web Archives that researchers could use to research events. This framework starts by collecting the initial document set from Web Archives using the keyword search interface of Web Archives. Since the documents of search results are mostly unavailable or inaccessible in a timely manner, we employ the snippets of each document. These snippets contain a part of the documents having query terms. After collecting these sets, we will use the previous algorithm to re-rank the initial collection and select the top most related documents. To make sure that we have a comprehensive final collection we will use various features of events as follows:

- We ensure to query all sub-events (e.g., *Arab Spring* is the main event and we retrieve documents related to all sub-events such as *Tunisian Revolution* too).
- Events have temporal and spatial features; *Arab Spring* as an example, lasted for a few years and spread to several countries. By taking temporal and spatial features into account, the final results would cover the temporal course of the events as well as their geographical movements.

To Evaluate the results of the last step, we will follow two settings:

- (1) Comparing our results with manually created event collections available in Web Archives,
- (2) User studies and expert interviews.

## 6 RESULTS

In this section, we present the current status of the PhD and the results we had so far. To answer the research questions mentioned in section 2, we evaluate the proposed approaches over baselines and calculate improvements in terms of standard metrics.

As we discussed in Section 2, creating the language-specific dataset is essential for developing interaction methods such as recommendations. We have created the EventKG+Click dataset [1] using EventKG knowledge graph and language-specific information on user interactions with events, entities, and their relations derived from the Wikipedia clickstream. The dataset includes German, French and Russian languages and covers information for

about 40,000 events in each language. Table 1 shows the statistics of this dataset. EventKG+Click is publicly available <sup>1</sup>.

Table 1: Statistics of EventKG+Click

	German	French	Russian
<b>Relevance Pairs</b>	304, 564	271, 243	254, 910
<b>Events</b>	40, 223	46, 557	33, 712
<b>Source Entities</b>	117, 281	104, 331	97, 212

We trained an LTR model using the EventKG+Click dataset to develop a language-specific event recommendation method using spatial, temporal, link-based, and latent embedding-based features. Our experiments on the EventKG+Click dataset demonstrate the effectiveness of our method in ranking events compared to the link-based and embedding-based recommendation baselines.

Table 2 compares the nDCG scores of the ranking study for our approach against the link-based and embedding based baselines. Our approach is able to outperform the baselines in all three languages.

Table 2: nDCG scores achieved by our approach and by three baselines in three languages in the ranking study

	nDCG Score			
	German	French	Russian	Avg.
<b>Milne-Witten</b>	0.900	0.904	0.915	0.906
<b>Node2Vec</b>	0.889	0.848	0.896	0.878
<b>DeepWalk</b>	0.907	0.865	0.918	0.897
<b>Our Approach</b>	<b>0.963</b>	<b>0.954</b>	<b>0.963</b>	<b>0.960</b>

Since Wikipedia Clickstream, as the primary source of multilingual user interaction, covers only a fraction of events potentially relevant to a query entity, we conducted an extensive user study. This user study evaluated our approach to determine whether it can satisfy users' information needs by considering different criteria namely general and local audience. Our approach obtains the highest MAP@5 on relevance to the language audience in all languages.

Regarding the third stage of the pipeline, we have created the Ms-Marco-Event dataset specialized for event-related queries and

<sup>1</sup><https://github.com/saraabdollahi/EventKG-Click>

documents upon Ms-Marco and using event-centric features from EventKG. We have trained a BERT model using this dataset. Next, we intend to fine-tune the base model using event-centric information. Furthermore, we will compare the results with the three groups of baselines mentioned in the Section 5.

Event document retrieval method is a starting point towards building event collections from Web Archives.

## 7 CONCLUSION

This PhD aims to develop user access models to event-centric information. We have defined a pipeline with components supporting researchers at different stages of their exploratory search to information retrieval and access comprehensive collections. So far, we have created datasets to train recommendation models. Moreover, we have developed an event recommendation method that considers the language community perspective. We aim to move forward by developing an event document retrieval method, a specialized method that incorporates event-centric information in the task of document retrieval to improve results for event queries. Furthermore, having event document retrieval, we intend to design a framework to build event collections from Web Archives as an essential source for researchers.

## 8 ACKNOWLEDGMENTS

I would like to thank my supervisors, Prof. Wolfgang Nejdl, Prof. Elena Demidova and Dr. Simon Gottschalk, for their valuable support in the realization of this work. This work has been funded by H2020-MSCA-ITN-2018-812997 under “Cleopatra”.

## REFERENCES

- [1] Sara Abdollahi, Simon Gottschalk, and Elena Demidova. 2020. EventKG+ Click: A Dataset of Language-specific Event-centric User Interaction Traces. *arXiv preprint arXiv:2010.12370* (2020).
- [2] Patti Bao, Brent Hecht, Samuel Carton, Mahmood Quaderi, Michael Horn, and Darren Gergle. 2012. Omnipedia: bridging the wikipedia language gap. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1075–1084. <https://doi.org/10.1145/2207676.2208553>
- [3] Bin Bi, Hao Ma, Bo-June Hsu, Wei Chu, Kuansan Wang, and Junghoo Cho. 2015. Learning to recommend related entities to search users. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*. 139–148. <https://doi.org/10.1145/2684822.2685304>
- [4] Roi Blanco, Berkant Barla Cambazoglu, Peter Mika, and Nicolas Torzec. 2013. Entity recommendations in web search. In *International Semantic Web Conference*. Springer, 33–48. [https://doi.org/10.1007/978-3-642-41338-4\\_3](https://doi.org/10.1007/978-3-642-41338-4_3)
- [5] Gerhard Gossen, Thomas Risse, and Elena Demidova. 2020. Towards extracting event-centric collections from Web archives. *International Journal on Digital Libraries* 21, 1 (2020), 31–45.
- [6] Simon Gottschalk and Elena Demidova. 2017. MultiWiki: Interlingual text passage alignment in Wikipedia. *ACM Transactions on the Web (TWEB)* 11, 1 (2017), 1–30. <https://doi.org/10.1145/3004296>
- [7] Simon Gottschalk and Elena Demidova. 2018. EventKG: a multilingual event-centric temporal knowledge graph. In *European Semantic Web Conference*. Springer, 272–287.
- [8] Aditya Grover and Jure Leskovec. 2016. Node2Vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. 855–864. <https://doi.org/10.1145/2939672.2939754>
- [9] Scott A Hale. 2014. Multilinguals and Wikipedia editing. In *Proceedings of the 2014 ACM conference on Web science*. 99–108. <https://doi.org/10.1145/2615569.2615684>
- [10] Helge Holzmann and Avishek Anand. 2016. Tempas: Temporal archive search based on tags. In *Proceedings of the 25th International Conference Companion on World Wide Web*. 207–210.
- [11] Helge Holzmann, Wolfgang Nejdl, and Avishek Anand. 2017. Exploring web archives through temporal anchor texts. In *Proceedings of the 2017 ACM on Web Science Conference*. 289–298.
- [12] Jizhou Huang, Shiqiang Ding, Haifeng Wang, and Ting Liu. 2018. Learning to recommend related entities with serendipity for web search users. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)* 17, 3 (2018), 1–22. <https://doi.org/10.1145/3185663>
- [13] Lucie-Aimée Kaffee, Alessandro Piscopo, Pavlos Vougiouklis, Elena Simperl, Leslie Carr, and Lydia Pintscher. 2017. A glimpse into Babel: an analysis of multilinguality in Wikidata. In *Proceedings of the 13th International Symposium on Open Collaboration*. 1–5. <https://doi.org/10.1145/3125433.3125465>
- [14] Nattiya Kanhabua, Philipp Kemkes, Wolfgang Nejdl, Tu Ngoc Nguyen, Felipe Reis, and Nam Khanh Tran. 2016. How to search the internet archive without indexing it. In *International Conference on Theory and Practice of Digital Libraries*. Springer, 147–160.
- [15] Zhenghao Liu, Chenyan Xiong, Maosong Sun, and Zhiyuan Liu. 2018. Entity-duet neural ranking: Understanding the role of knowledge graph semantics in neural information retrieval. *arXiv preprint arXiv:1805.07591* (2018).
- [16] Paolo Massa and Federico Scrinzi. 2012. ManyPedia: Comparing language points of view of Wikipedia communities. In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration*. 1–9. <https://doi.org/10.1145/2462932.2462960>
- [17] Federico Nanni, Simone Paolo Ponzetto, and Laura Dietz. 2017. Building entity-centric event collections. In *2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. IEEE, 1–10.
- [18] Tri Nguyen, Mir Rosenberg, Xia Song, Jianfeng Gao, Saurabh Tiwary, Rangan Majumder, and Li Deng. 2016. MS MARCO: A human generated machine reading comprehension dataset. In *CoCo@ NIPS*.
- [19] Chien-Chun Ni, Kin Sum Liu, and Nicolas Torzec. 2020. Layered graph embedding for entity recommendation using wikipedia in the yahoo! knowledge graph. In *Companion Proceedings of the Web Conference 2020*. 811–818. <https://doi.org/10.1145/3366424.3383570>
- [20] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. 701–710. <https://doi.org/10.1145/2623330.2623732>
- [21] Stephen E Robertson and Steve Walker. 1994. Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval. In *SIGIR'94*. Springer, 232–241.
- [22] Richard Rogers. 2013. *Digital methods*. MIT press. <https://doi.org/10.1002/asi.23594>
- [23] Dwaipayan Roy, Sumit Bhatia, and Prateek Jain. 2021. Information asymmetry in Wikipedia across different languages: A statistical analysis. *Journal of the Association for Information Science and Technology* (2021).
- [24] Anna Samoilenko, Fariba Karimi, Daniel Edler, Jérôme Kunegis, and Markus Strohmaier. 2016. Linguistic neighbourhoods: Explaining cultural borders on Wikipedia through multilingual co-editing activity. *EPJ data science* 5, 1 (2016), 9. <https://doi.org/10.1140/epjds/s13688-016-0070-8>
- [25] Jaspreet Singh, Wolfgang Nejdl, and Avishek Anand. 2016. Expedition: a time-aware exploratory search system designed for scholars. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. 1105–1108.
- [26] Jaspreet Singh, Wolfgang Nejdl, and Avishek Anand. 2016. History by diversity: Helping historians search news archives. In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*. 183–192.
- [27] Nam Khanh Tran, Tuan Tran, and Claudia Niederée. 2017. Beyond time: Dynamic context-aware entity recommendation. In *European Semantic Web Conference*. Springer, 353–368. [https://doi.org/10.1007/978-3-319-58068-5\\_22](https://doi.org/10.1007/978-3-319-58068-5_22)
- [28] Khoi Duy Vo, Tuan Tran, Tu Ngoc Nguyen, Xiaofei Zhu, and Wolfgang Nejdl. 2016. Can we find documents in web archives without knowing their contents?. In *Proceedings of the 8th ACM Conference on Web Science*. 173–182.
- [29] Ian H Witten and David N Milne. 2008. An effective, low-cost measure of semantic relatedness obtained from Wikipedia links. (2008).
- [30] Chenyan Xiong, Jamie Callan, and Tie-Yan Liu. 2017. Word-entity duet representations for document ranking. In *Proceedings of the 40th International ACM SIGIR conference on research and development in information retrieval*. 763–772.
- [31] Chenyan Xiong, Zhengzhong Liu, Jamie Callan, and Eduard 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*. 2391–2394.
- [32] Chenyan Xiong, Russell Power, and Jamie Callan. 2017. Explicit semantic ranking for academic search via knowledge graph embedding. In *Proceedings of the 26th international conference on world wide web*. 1271–1279.
- [33] Lei Zhang, Achim Rettinger, and Ji Zhang. 2016. A probabilistic model for time-aware entity recommendation. In *International Semantic Web Conference*. Springer, 598–614. [https://doi.org/10.1007/978-3-319-46523-4\\_36](https://doi.org/10.1007/978-3-319-46523-4_36)