

Temporal Question Answering in News Article Collections

Adam Jatowt
University of Innsbruck
Innsbruck, Austria
adam.jatowt@uibk.ac.at

ABSTRACT

The fields of automatic question answering and reading comprehension have been rapidly advancing recently [6]. Open-domain question answering, in particular, assumes answering arbitrary user questions from a large document collection. The existing QA approaches are however generally working on synchronic document collections such as Wikipedia, Web data or short-term news corpora. This talk is about our latest efforts in automatic question answering over temporal news collections which can contain millions of news articles published during several decades' long time frames. Temporal aspects of both news articles and user questions form an additional challenge for this kind of a temporal question answering task. To correctly answer questions over such collections one usually needs first to find candidate documents that are likely to contain these answers. We will first discuss a re-ranking approach [3] for news articles which works by utilizing temporal information embedded in questions and in the underlying document collection, thus combining methods from Temporal Information Retrieval [1, 2] and Natural Language Processing. Next, we will discuss a dedicated solution for answering "When" type questions which require finding occurrence dates of events described in input questions based on the underlying news archive [5]. Finally, we will introduce ArchivalQA [4] - a large-scale question answering dataset which has been automatically created from a two decades' long news article collection, and which contains over 500k question-answer pairs. The dataset has been processed to remove temporally ambiguous questions for which more than one correct answer exist.

CCS CONCEPTS

• **Information systems** → **Question answering; Content analysis and feature selection.**

KEYWORDS

question answering, news archives, temporal IR, question generation

ACM Reference Format:

Adam Jatowt. 2022. Temporal Question Answering in News Article Collections. In *Companion Proceedings of the Web Conference 2022(WWW '22 Companion)*, April 25–29, 2022, Virtual Event, Lyon, France. ACM, New York, NY, USA, 1 page. <https://doi.org/10.1145/3487553.3526023>

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).

WWW '22 Companion, April 25–29, 2022, Virtual Event, Lyon, France

© 2022 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-9130-6/22/04.

<https://doi.org/10.1145/3487553.3526023>

1 BIOGRAPHY

Adam Jatowt is a Professor at the Department of Computer Science and Digital Science Center at the University of Innsbruck. He is also a visiting researcher at the National Institute of Advanced Industrial Science and Technology in Tokyo. Before moving to Austria, he worked at Kyoto University as an Assistant and later as an Associate Professor. He received his Ph.D. in Information Science and Technology from the University of Tokyo, Japan in 2005. His research interests include knowledge extraction, information retrieval, natural language processing, digital libraries and digital history. Adam has published over 150 research papers in international conferences and journals. He was also a co-organizer of three NTCIR evaluation tasks and a co-organizer of over 20 international workshops. He is an editorial board member of IP&M, JASIST, IJDL, JIS and IEEE JoSC journals. Adam has received the Vannevar Bush Best Paper Award at JCDL2021, the best short paper award at ECIR2018, and the best demo award at ECIR2019.



REFERENCES

- [1] Ricardo Campos, Gaël Dias, Alípio M Jorge, and Adam Jatowt. 2014. Survey of temporal information retrieval and related applications. *ACM Computing Surveys (CSUR)* 47, 2 (2014), 1–41.
- [2] Nattiya Kanhabua, Roi Blanco, and Kjetil Nørkvåg. 2015. Temporal Information Retrieval. *Foundations and Trends® in Information Retrieval* 9, 2 (2015), 91–208. <https://doi.org/10.1561/15000000043>
- [3] Jiexin Wang, Adam Jatowt, Michael Färber, and Masatoshi Yoshikawa. 2021. Improving question answering for event-focused questions in temporal collections of news articles. *Information Retrieval Journal* 24, 1 (2021), 29–54.
- [4] Jiexin Wang, Adam Jatowt, and Masatoshi Yoshikawa. 2021. ArchivalQA: A Large-scale Benchmark Dataset for Open Domain Question Answering over Historical News Collections. *SIGIR* (2021).
- [5] Jiexin Wang, Adam Jatowt, and Masatoshi Yoshikawa. 2021. Event Occurrence Date Estimation based on Multivariate Time Series Analysis over Temporal Document Collections. In *SIGIR '21*. ACM, 398–407.
- [6] Fengbin Zhu, Wenqiang Lei, Chao Wang, Jianming Zheng, Soujanya Poria, and Tat-Seng Chua. 2021. Retrieving and Reading: A Comprehensive Survey on Open-domain Question Answering. *arXiv preprint arXiv:2101.00774* (2021).