

# Knowledge-Based News Event Analysis Toolkit

Oktie Hassanzadeh<sup>1</sup>, Parul Awasthy<sup>1</sup>, Ken Barker<sup>1</sup>, Onkar Bhardwaj<sup>1</sup>,  
Debarun Bhattacharjya<sup>1</sup>, Mark Feblowitz<sup>1</sup>, Aamod Khatiwada<sup>1,3,†</sup>, Lee Martie<sup>1</sup>,  
Steve Fonin Mbouadeu<sup>1,4,†</sup>, Jian Ni<sup>1</sup>, Anik Saha<sup>1,2,†</sup>, Sola Shirai<sup>1,2,†</sup>, Kavitha Srinivas<sup>1</sup>  
and Lucy Yip<sup>1</sup>

<sup>1</sup>IBM Research, Yorktown Heights, NY, USA

<sup>2</sup>Rensselaer Polytechnic Institute, Troy, NY, United States

<sup>3</sup>Khoury College of Computer Sciences, Northeastern University, Boston, MA, USA

<sup>4</sup>St. John's University, NY, USA

## Abstract

We present an overview of our knowledge-based news event analysis toolkit. The toolkit is powered by a knowledge graph (KG) of event-related concepts and relations curated from Wikidata and enriched through knowledge extraction from text as well as a variety of link prediction methods. We describe each of the functions the toolkit provides and an overview of its various components. We present use cases in enterprise risk management, scenario planning, and media intelligence. We also discuss a number of lessons learned and directions for future research.

Businesses – large and small – can benefit tremendously from monitoring ongoing global and local newsworthy events and analyzing how recent events could impact their businesses. One mechanism of analysis is curating a rich knowledge graph (KG) of past events and their consequences, such that ongoing events can be mapped to past similar events in the KG, and one can reason about what caused them and what can happen as a result. As a simple example, back in January 8 2020, when news stories started reporting on the announcement that the World Health Organization (WHO) made on a new virus that has caused a pneumonia outbreak in Wuhan, China, one could immediately identify past similar news events, which include WHO's announcement in March 2003 that marked the onset of the 2002–2004 SARS outbreak. A business involved in tourism or oil & gas industries can then immediately start taking actions to prepare for the potential impact of a major disease outbreak on their businesses.

In this talk, we present an overview of a toolkit that enables building knowledge-based news event analysis solutions. Our goal in developing this toolkit is twofold: 1) providing news event analysis functions based on a rich curated knowledge base from publicly available sources 2) providing knowledge extraction functions used to curate our knowledge base such that users

---

*The 21st International Semantic Web Conference (ISWC2022), October 23-27, 2022*

<sup>†</sup> Work done while at IBM Research.

✉ hassanzadeh@us.ibm.com (O. Hassanzadeh); awasthy@us.ibm.com (P. Awasthy); kjbarker@us.ibm.com (K. Barker); onkarbhardwaj@ibm.com (O. Bhardwaj); debarunb@us.ibm.com (D. Bhattacharjya); mfeb@us.ibm.com (M. Feblowitz); khatiwada.a@northeastern.edu (A. Khatiwada); Lee.Martie@ibm.com (L. Martie); steve.mbouadeu19@my.stjohns.edu (S. F. Mbouadeu); nij@us.ibm.com (J. Ni); sahaa@rpi.edu (A. Saha); shiras2@rpi.edu (S. Shirai); kavitha.srinivas@ibm.com (K. Srinivas); Lucy.Yip@ibm.com (L. Yip)

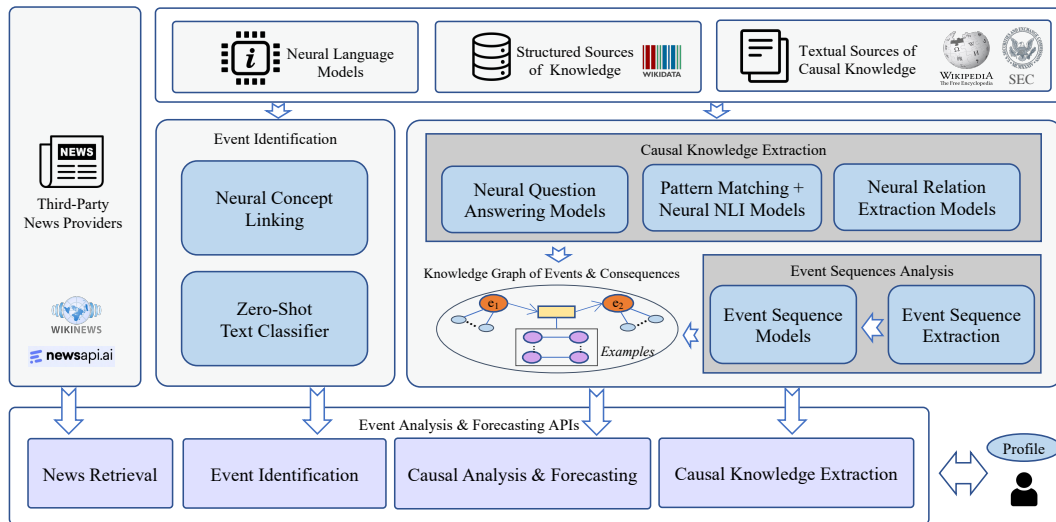
🆔 0000-0001-5307-9857 (O. Hassanzadeh); 0000-0003-0899-8904 (K. Barker); 0000-0001-5720-1207 (A. Khatiwada); 0000-0002-9137-407X (S. F. Mbouadeu); 0000-0001-6913-3598 (S. Shirai)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)



**Figure 1:** Toolkit Components and APIs

can augment the included knowledge base or curate a custom knowledge base for their domain of interest. Figure 1 presents the current architecture of our toolkit [1]. We outline several challenges we faced in applying state-of-the-art concept linking, knowledge extraction, and link prediction techniques to build our toolkit, provide a summary of lessons learned, and present a number of research challenges that need to be addressed. In particular:

- We outline the use of Wikidata as a primary source of knowledge, report on the challenges we faced with respect to the current coverage of event-related concepts in Wikidata, and how the existing knowledge in Wikidata can be enriched through automated knowledge extraction over Wikipedia articles. Our primary focus has been on weakly supervised and supervised neural models for causal relation extraction.
- We describe our solution for mapping news headlines to concepts in our KG, and report on challenges in applying existing concept linking methods to this problem.
- We report on the performance of several rule-based and knowledge graph embeddings based approaches for link prediction to enrich our KG. We also report on the challenges we faced in applying existing techniques for reasoning about potential consequences of a new event as a novel mechanism for event forecasting.
- We also report on our preliminary results on automatically extracting structured event sequences from textual corpora and applying event sequence models as a mechanism of learning complex relations between event types in our KG.

## References

- [1] O. Hassanzadeh, P. Awasthy, K. Barker, O. Bhardwaj, D. Bhattacharjya, M. Feblowitz, L. Martie, J. Ni, K. Srinivas, L. Yip, Knowledge-based news event analysis & forecasting toolkit, in: IJCAI (Demonstration Track), 2022.