

WikiEvents - A Novel Resource for NLP Downstream Tasks

Lars Michaelis^{1,2,*}, Junbo Huang^{1,2,*} and Ricardo Usbeck^{1,2,*}

¹University of Hamburg, Germany

²Hamburger Informatik Technologie-Center (HITeC) e.V., Germany

Abstract

Efficient Natural Language Processing (NLP) models require large amounts of training data. Manually creating training data is time-consuming. We present WikiEvents, an automatically curated dataset based on Wikipedia's Current Events portal. WikiEvents is a novel knowledge graph that aims to provide data for various event-centric NLP tasks, such as event-related location extraction and entity linking. Therefore, WikiEvents includes event summaries with linked entities and locations. WikiEvents also provides spatial and temporal information about extracted events for various use case analyses. We leverage the NLP Interchange Format (NIF) ontology and an event-specific novel ontology - CoyPu. We evaluate the suitability regarding NLP tasks by (1) training three BERT models on event-related location extraction with data queried from WikiEvents and (2) comparing WikiEvents to the existing entity linking dataset AIDA-YAGO2. Qualitative, event-related research capabilities are explored by querying data from WikiEvents for multiple use cases and visualizing it.

Keywords

Events, NLP, Knowledge Graph, Dataset, Event Extraction, Event Detection, NIF, CoyPu

1. Introduction

With the rise of machine learning came an increasing need for large-scale training data. Event-centric NLP comprises several subtasks. The NLP task of event-related location extraction (LE) concentrates on only annotating the named location where an event is taking place within a document. Multiple LE datasets were made using human annotation of different sources. Lingad et al. [1] and Ji et al. [2] annotated Tweets, Gupta and Nishu [3] annotated news articles, and Wang et al. [4] annotated Wikipedia articles. The NLP task of entity linking (EL) involves linking named entities in texts to their entities in knowledge bases. Existing EL datasets include AIDA-YAGO2 [5] and SAT-300 [6] do not focus on (crisis) event reports but rather news or encyclopedic articles in general. For more works w.r.t. EL we refer the interested reader to Möller et al. [7].

The creation of datasets through human annotation is time and cost-intensive. With WikiEvents, we leverage an automatic extraction based on an existing source of summarized events with manually linked named entities. That is, our system extracts content from

SEMMEs 2023: Semantic Methods for Events and Stories workshop, ESWC 2023, May 28-29, 2023, Heraklion, Greece

*Corresponding author.

✉ lars.michaelis@hitec-hamburg.de (L. Michaelis); junbo.huang@uni-hamburg.de (J. Huang);

ricardo.usbeck@uni-hamburg.de (R. Usbeck)



© 2023 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)

Wikipedia’s Current events portal¹ to create a novel KG-based dataset for LE and EL tasks. Furthermore, our system uses Wikipedia articles, Wikidata² resources, the Nominatim API³ and the Falcon 2.0 entity linker⁴ [8] to extract additional temporal, spatial and event information about events and entities mentioned in the Current events portal.

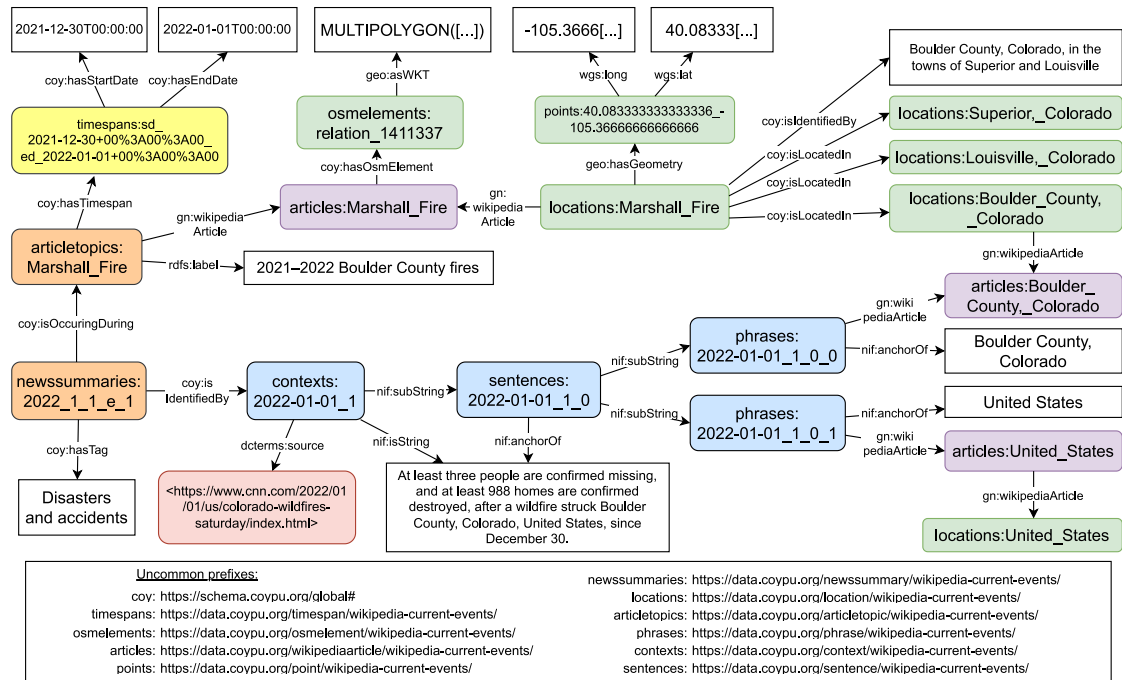


Figure 1: Example simplified graph generated for our running example event annotated in Figure 2.

Our goal is to use WikiEvents for the NLP tasks of EL and event-related LE, as well as provide the extracted event data for event-related research. For a sub-graph example based on the Marshall Fires see Figure 1. To the best of our knowledge, no comprehensive dataset for these tasks existed before.

We evaluate WikiEvents on the mentioned NLP tasks by (1) comparing it to the existing EL dataset AIDA-YAGO2 using performances of current EL models on both datasets as well as (2) training and evaluating event-related transformer-based location extractors. Finally, we explore several use cases regarding possible event-related research questions.

The source code and a dataset sample are available at:

- The extractor of the dataset:
<https://github.com/semantic-systems/current-events-to-kg>
- The machine learning code for EL and LE:
<https://github.com/semantic-systems/coypu-current-events-for-ml>

¹https://en.wikipedia.org/wiki/Portal:Current_events

²<https://www.wikidata.org/>

³<https://nominatim.openstreetmap.org/ui/search.html>

⁴<https://labs.tib.eu/falcon/falcon2/>

- A sample of the dataset including training and test samples for EL and LE:
<https://www.fdr.uni-hamburg.de/record/11447>

2. WikiEvents Knowledge Graph

The *WikiEvents* knowledge graph is automatically generated from the Wikipedia Current event portal as its primary data source. It is saved as *Resource Description Format* (RDF)⁵ and serialized as JSON-LD⁶.

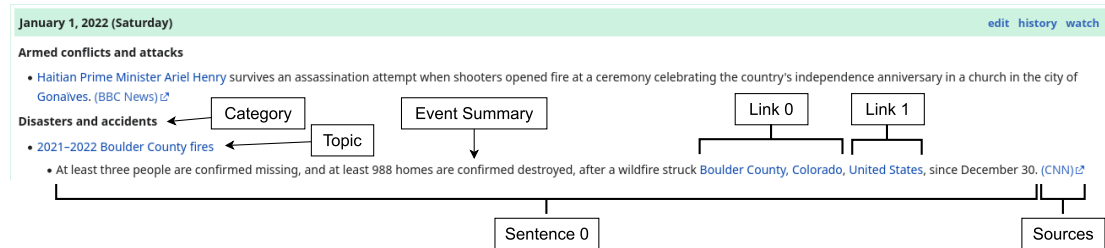


Figure 2: The first events in the Current event portal of January 1, 2022. Elements are annotated to illustrate the naming scheme used in this paper. (Screenshot taken 28 December, 2022. This modified version is licensed under CC BY-SA 3.0.)

First, we give an overview of the general structure of WikiEvents. Afterward, the following subsections will provide details about each kind of included information.

Two types of events are extracted from the Current events portal: *Event summaries* and *topics*. Both event types link to Wikipedia articles. Articles linked to topics further describe the topic event, while articles in event summaries describe named entities. The event summaries are grouped into sections of different *categories*. Entities for locations are created if an article is identified as describing a location, see Figure 2.

2.1. Event Information

The *CoyPu* ontology⁷ is used to encode the event-related data of topics and event summaries. It was developed as part of the CoyPu⁸ project which aims to increase the resilience of companies during crises. We acknowledge the existence of ontologies which could replace the CoyPu Ontology, such as CIDOC-CRM [9] or Simple Event Model (SEM) [10]. However, we decided to use a niche ontology because using established ontologies would contradict the project requirements.

Event summaries are short summaries of significant real-world events. They include hyperlinks from named entities to Wikipedia articles. The event summary entities are linked to the category under which they were extracted which can be used to roughly classify them (e.g. *Armed conflicts and attacks* or *Disasters and accidents*). Providing more specific event types

⁵<https://www.w3.org/TR/rdf11-concepts/>

⁶<https://www.w3.org/TR/json-ld11/>

⁷<https://schema.coypu.org/global/>

⁸<https://coypu.org/>

for event summaries (flood, election, ...) is accomplished by using event types linked to the Wikidata entity of its parent topics.

The *NLP Interchange Format* (NIF) ontology [11] is being used to encode the link from named entities in event summaries to their entities, i.e., Wikipedia articles. The event summary is first split into sentences, to which included named entities are linked. These entities of named entities link to the Wikipedia article's entity which their hyperlink originally referenced. The news sources mentioned in the event summaries are linked to the respective event summary.

Topics can but do not need to reference a Wikipedia article and are therefore mapped to two classes. The class for topic entities with a linked article is a subclass of the second class since only the possible properties are extended through the additional article. Topics mentioned at different points are mapped to identical entities if they either link to the same article or have the same headline when no article is linked.

Both topics and event summaries link to the event under which they are listed in the Current events portal as their parent event. Events summaries and topics can only be sub-events of topics since event summaries are leaves within a tree-structured list of mentioned events.

2.2. Wikipedia Articles

The Geonames ontology⁹ is used to generate entities for Wikipedia articles referenced by topics and linked in event summaries. It is linked to the metadata from the Wikipedia article's schema graph and its Wikidata entity. The one-hop graph extracted around the Wikidata entity is additionally included to eliminate simple queries to Wikidata endpoints.

2.3. Spatial Information

WikiEvents includes location entities, coordinates of locations, and WKT-encoded boundaries of locations. We create a location entity next to marking it as a Wikipedia link/entity, if a Wikipedia article is identified as describing a location or if an article is referenced by a topic.

Our method of identifying location articles follows the method of identifying toponyms from Wang et al. [4]. They concluded that the identification of Wikipedia articles about locations should be done by determining if an article uses specific infobox templates¹⁰. Additionally, we check the infobox's HTML table element, if the class attribute includes location-related template classes, e.g., `ib-island`. This addition increased the recall of the location identification process from 93.3% to 94.3% when the evaluation identified locations for January 2022.

We encode hierarchical information about the locations by linking location entities together. This is useful for event-related LE, so to only label the most specific location in an event summary as the location of the event. We extract location hierarchies through (1) links in article infoboxes under location-describing keys, (2) entities linked by Falcon 2.0 under these keys, and (3) querying Wikidata for parent locations of entities.

Boundaries of locations are queried from the OpenStreetMaps database using the Nominatim geocoding service. Additionally, we check the Wikidata entity of the Wikipedia article for linked spatial information from OpenStreetMaps.

⁹<https://www.geonames.org/ontology/documentation.html>

¹⁰Listed here: https://en.wikipedia.org/wiki/Wikipedia:List_of_infoboxes/Place

2.4. Temporal Information

All events are linked to dates on which they have been mentioned in the Current events portal. Additionally, we employ the infoboxes of topic articles to extract more specific temporal information about the topic event. We developed a pattern-based parser to parse the values of specific infobox keys. The parser extracts (1) dates, (2) times, (3) spans of date and/or time, and (4) UTC timezones. Similar to the CIDOC-CRM ontology, the temporal information extracted from the infobox is saved to a timespan entity since it enables timespan-timespan relation inference. We also include the parsed source strings from the infobox to enable supervised machine learning on timespan extraction in the future. In WikiEvents from January 2022 to December 2022 are 3683 unique topics with 1303 of them having a linked timespan entity from a total of 1167 unique timespan entities.

3. Evaluation

We evaluated the usefulness of WikiEvents w.r.t. NLP downstream task applicability via event-related location extraction and entity linking tasks.

3.1. Event-Related Location Extraction

Event-related location extraction is a special subtask of LE, where only the event location is extracted. For example, the following sentence “The United States Embassy in Kyiv calls on Russia to ”fully comply” with the ceasefire in Donbas after pro-Russian forces shelled the strategic Hnutove entry-exit checkpoint and a humanitarian road corridor.”¹¹ contains multiple location mentions. Event-related LE only extracts where the event took place (Kyiv) and ignores other location mentions (United States, Russia, Donbas, Hnutove).

Since WikiEvents is a KG, various information can be retrieved using custom SPARQL queries. Therefore, we created a query to retrieve training data for event-related LE and tested it by fine-tuning multiple transformer-based models and evaluated their performance.

3.1.1. Querying Training Samples

Our query has three logical steps: (1) selecting named location entities for each event summary (candidate locations), (2) filtering out less specific candidate locations, e.g., Germany in *Hamburg, Germany*, and (3) choosing the location of the event from remaining candidate locations. The first step is done via a SPARQL query. The second step (filtering) is facilitated by the extracted hierarchy between locations. This reduced the number of event summaries with multiple location candidates by 31.7% in the queried data. The third step is performed using a heuristic that takes the first location candidate as the event location.

To analyze the quality of our heuristics, we looked at the event summaries from January 2023 (n=276). In 77.5% of event summaries, the first link was a correct event location entity. In 85.5% of event summaries, the first link was part of a set of correct event location entities. We are aware that choosing only one location as the assumed true event location is a systematic

¹¹Source: https://en.wikipedia.org/wiki/Portal:Current_events/February_2022

problem with this approach since multiple affected locations can be mentioned. Thus, further research into selecting the right candidates is required to be able to utilize this dataset’s full potential.

3.1.2. Fine-Tuning Setting

Three different uncased BERT models (DistilBERT, BERT base, and large) were fine-tuned on the data to evaluate how model size affects performance. The uncased models were chosen since we assumed that the case of the text holds limited value to the detection of location information. We modeled the task as a token classification task. Hyperparameters were taken from ones suggested by Devlin et al. [12] (learning rate = $3e-5$, batch size = 16, AdamW optimizer). No epoch limit in combination with early stopping was used to assure full training. Following Mosbach et al. [13], a warm-up phase of 10% of total training steps was employed (taking 4 training epochs as a prediction of total training epochs).

The training samples from WikiEvents were queried from January 2020 to December 2022 resulting in 16451 samples. We based the following experiments on a 80/10/10 train-eval-test split for fine-tuning.

3.1.3. Results

The performance metrics of each model are shown in Table 1. From the minor differences in performance between the models, you can conclude that model size is not majorly limiting the performance.

Table 1

Performance metrics of BERT models trained on the location extraction data queried from the WikiEvents dataset between 01/2020 and 12/2022. The best values are bold.

Model	F1	Precision	Recall
distilbert-base-uncased	83.0	84.5	81.6
bert-base-uncased	83.4	84.2	82.6
bert-large-uncased	83.2	84.5	82

Additionally, we tried to evaluate the DistilBERT model on the event summaries from January 2023 which were previously used to evaluate the heuristic for selecting the best location link. One limitation of reusing this evaluation dataset is that the correct location is not always the most specific in cases where the most specific was not hyperlinked (e.g., “Hamburg” could not be annotated in “Hamburg, Germany” when only “Germany” is hyperlinked). To counteract this, wrong location predictions were manually checked for location annotation limitations (13 were found). The model identified all event locations in 70.3% of event summaries while at least one location was identified in 79% of event summaries (65.6% and 75.4% without manual reevaluation).

In the future, the next step will be improving the quality of the data samples for training. In particular, we suspect that we need a better third step, i.e., identifying the correct event location.

3.2. Entity Linking

The second analyzed NLP task for the WikiEvents dataset is entity linking. The hypothesis is that entity linking data samples can be queried with enough quality and quantity to train entity linking models. To test this hypothesis, a query was constructed to get data samples usable for entity linking tasks. Moreover, the constructed dataset, together with AIDA-YAGO2 (AIDA) dataset, was evaluated with two existing EL models.

3.2.1. Experimental Setting

The used EL models are BLINK [14] and ELQ [15]. BLINK is an entity linking model based on a two-stage fine-tuned BERT architecture. ELQ uses an end-to-end (one-stage) entity linking BERT model for linking entities, primarily in short texts such as questions. These models were used since no openly accessible trained models were found for other entity linking models (DeepType [16], BERT-Entity [17]) and training would have required considerable resources.

The data samples representing WikiEvents were queried from a WikiEvents dataset extracted from 01/2020 to 12/2022¹². In total, we created 20630 samples with a list of mentions each. Since AIDA consists of longer news articles as source texts, it only has 1392 samples but with more mentions per sample. The queried WikiEvents entity linking data has 70241 mentions with on average 3.5 mentions in each sample. The AIDA dataset has only 27812 mentions but with 20 mentions per sample on average.

3.2.2. Experimental Results

Table 2

Performance metrics of BLINK and ELQ on AIDA and the queried WikiEvents entity linking data from the WikiEvents dataset between 01/2020 and 12/2022. The best accuracy values are bold.

Dataset	BLINK			ELQ				
	Acc.	Recall@10	Supp.	Acc.	F1	Prec.	Recall	Supp.
WikiEvents	75.8	82.5	75151	42.5	59.6	60.6	58.7	103900
AIDA	81.6	92.6	27812	21.6	35.5	66.9	24.1	31106

As shown in Table 2, BLINK clearly outperforms ELQ. One limiting factor of ELQ in prediction mode is the truncation of longer input texts. Since AIDA has many longer articles, entity mentions are cut off which results in a lower recall. Since BLINK receives mentions individually with the surrounding context, it has a significantly higher recall on AIDA. An additional explanation for the higher recall of BLINK on AIDA could be unlinked entities in WikiEvents. A third explanation is that entities that are unknown to both models are included in WikiEvents. Since both models were trained on Wikipedia dumps from August 2019, they do not include the events from 2020 onward present in this WikiEvents dataset. The significantly better precision of ELQ on AIDA provides evidence for this. Our evaluation shows that WikiEvents can be

¹²<https://doi.org/10.25592/uhhfdm.11447>

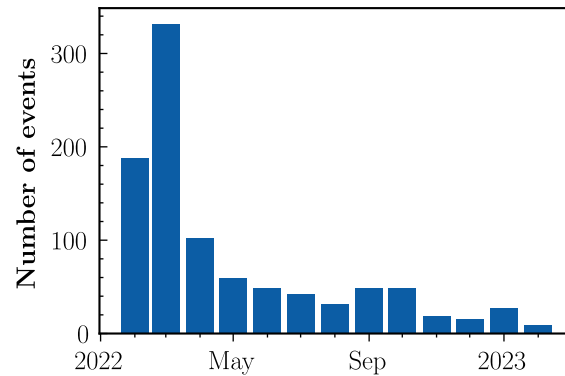


Figure 4: The amount of sub-events of the 2022 Russian invasion of Ukraine with event summaries.

4.2. Analyzing Event Development

To estimate how an event develops, you can analyze the number of sub-events over time. The Current events portal often relies on media coverage of the event. This coverage is then summarized by volunteers. Thus, the frequency and amount of sub-events regarding one event could indicate the coverage and presence of this event in people’s lives. Following the previous example event, Figure 4 shows the number of event summaries linked to the 2022 Russian invasion of Ukraine. You can observe (1) a substantial decrease in the number of event summaries after the start of the invasion and (2) slight temporary increases around starting September when the Ukrainian counteroffensive took place and at the beginning of 2023.

4.3. Locating Events Geographically

Using the included spatial information in WikiEvents, events with a linked location correspond to a specific area. WikiEvents’ spatial data also enables filtering for area-specific events and, thus, entities in this area, e.g., companies in cities under attack.

Following the previous example again, you could visualize the number of events mentioning specific areas at different during February and August 2022, shown in Figure 5.

From these maps, you can observe that WikiEvents is able to link larger areas to events and vice versa. The detail (size of areas) is dictated by how much the authors summarized the original events in the Current events portal. The shown areas match information from the Institute for the Study of War about both months of the invasion [18, 19, 20].

5. Related Work

Multiple event-related datasets have been created for different goals. The GDELT Project [21] generates event data from news articles to study human societal-scale behavior and beliefs across the world. The ACLED Project [22] collects crisis and political events to map and analyze them. EM-DAT [23] is a dataset of mass disaster events to improve disaster related decision-making at the (inter-)national level. EventWiki [24] manually identified Wikipedia articles about major

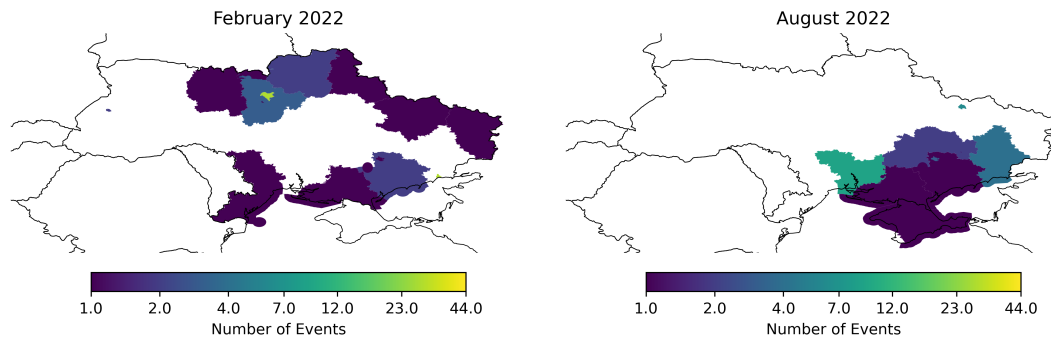


Figure 5: The number of sub-events of the 2022 Russian invasion of Ukraine mentioning specific areas during February and August 2022.

events. They extracted event-related data from infoboxes and article texts while classifying each event into 95 event types based on the used infobox template. The first three use a relational data model for storing extracted events while this can only be assumed for EventWiki. In contrast, EventKG [25] uses the knowledge graph data model to consolidate event data from multiple sources into a common format. EventKG focuses on the completeness of temporal information regarding events. Intended use cases are Digital Humanity and NLP tasks like question answering, timeline generation, and language- or community-specific cross-cultural studies. Table 3 shows an overview of the mentioned datasets.

The closest dataset to WikiEvents is EventKG. Both are knowledge graphs and include events with temporal and spatial information. The main differences between WikiEvents over EventKG are: (1) the inclusion of links between event summaries and mentioned entities, (2) the inclusion of boundary data of locations, (3) only identifying named entities as being locations but not for being actors, (4) not being multilingual by only including information in English, (5) having a less abstract ontology to include source-specific information. Figure 6 shows a comparison of graph structure between WikiEvents and EventKG.

Table 3

Overview of general aspects of mentioned existing event datasets.

Dataset	# of Events	Specific Types	Event	Languages	Licence/Terms of Use
GDELT	250 mil.	CAMEO types		>100	Terms of Use ¹³
ACLED	>1 mil.	political & crisis events		>100	Terms of Use ¹⁴
EM-DAT	18.000	mass disasters		English	Data Access Policy ¹⁵
EventWiki	21.275	major events		English	-
EventKG	690.000	-		English, German, French, Portuguese	CC BY-SA 4.0

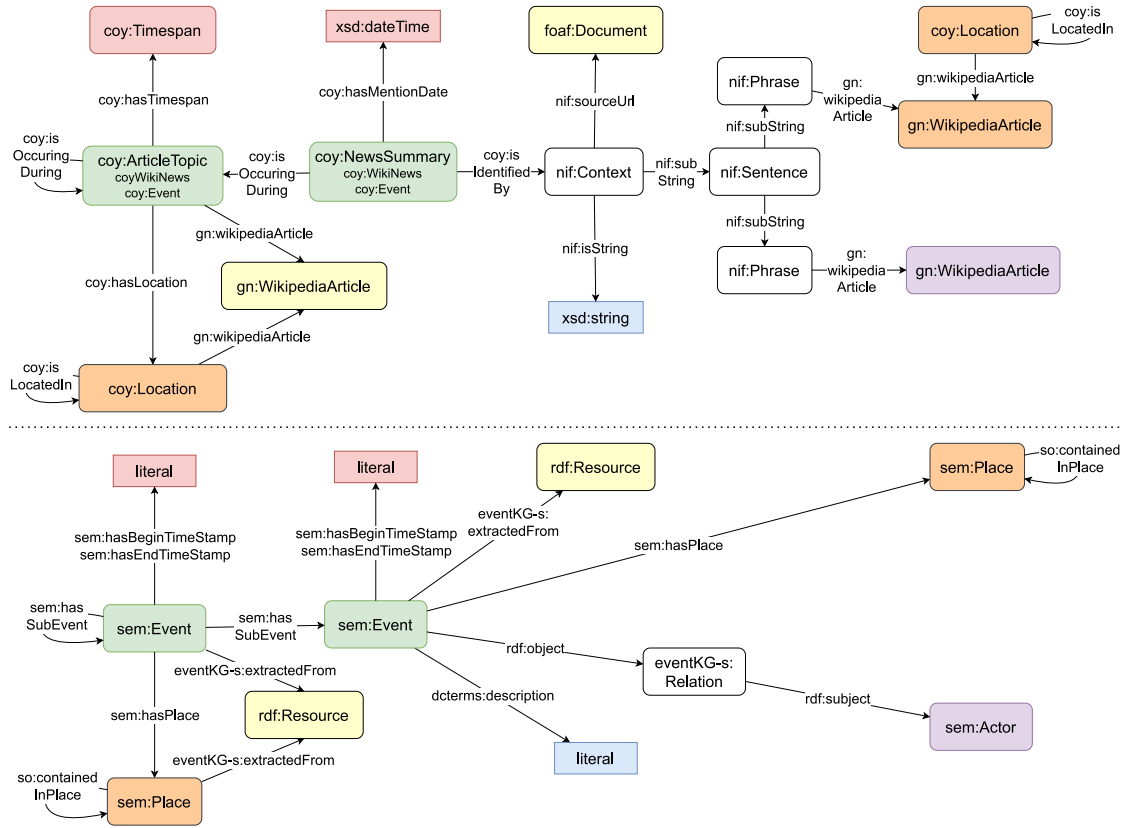


Figure 6: A comparison of the graph structure of WikiEvents (top) and EventKG (bottom). The similarity is expressed through node location and node color (white nodes have no counterpart).

6. Summary and Future Work

We presented the novel WikiEvents knowledge graph that is extracted automatically from the Wikipedia Current events portal and other data sources. Its main aim are NLP tasks such as EL and event-related LE. We evaluated these capabilities by (1) comparing it to the existing human-annotated EL dataset AIDA-YAGO2 and (2) training three BERT models on event-related LE. Also, we describe multiple use cases of WikiEvents regarding event-related research exploring the 2022 Russian invasion of Ukraine. The use cases included sub-event analysis, event development analysis, and event localization. Finally, to highlight differences, we compared the dataset to the existing event knowledge graphs such as EventKG.

In the near future, we will improve the event location identification heuristic in news summaries and continuously improve the temporal and spatial information extractor since Wikipedia’s website constantly evolves. To foster machine learning research, we will create larger task-specific datasets with dedicated train-validation-test splits. Our approach to training event-related LE BERT models using WikiEvents can be further evaluated by comparing the model performances to models trained on comparable datasets.

Acknowledgments

This research was supported by grants from NVIDIA and utilized NVIDIA 2 x RTX A5000 24GB. Furthermore, we acknowledge the financial support from the Federal Ministry for Economic Affairs and Energy of Germany in the project CoyPu (project number 01MK21007[G]).

References

- [1] J. Lingad, S. Karimi, J. Yin, Location extraction from disaster-related microblogs, in: WWW (Companion Volume), International World Wide Web Conferences Steering Committee / ACM, 2013, pp. 1017–1020.
- [2] Z. Ji, A. Sun, G. Cong, J. Han, Joint recognition and linking of fine-grained locations from tweets, in: WWW, ACM, 2016, pp. 1271–1281.
- [3] S. Gupta, K. Nishu, Mapping local news coverage: Precise location extraction in textual news content using fine-tuned BERT based language model, in: Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science, Association for Computational Linguistics, Online, 2020, pp. 155–162. URL: <https://aclanthology.org/2020.nlpcss-1.17>. doi:10.18653/v1/2020.nlpcss-1.17.
- [4] J. Wang, Y. Hu, K. Joseph, Neurotrp: A neuro-net toponym recognition model for extracting locations from social media messages, Trans. GIS 24 (2020) 719–735.
- [5] J. Hoffart, M. A. Yosef, I. Bordino, H. Fürstenau, M. Pinkal, M. Spaniol, B. Taneva, S. Thater, G. Weikum, Robust disambiguation of named entities in text, in: EMNLP, ACL, 2011, pp. 782–792.
- [6] A. Mandalios, K. Tzamaloukas, A. Chortaras, G. Stamou, GEEK: incremental graph-based entity disambiguation, in: LDOW@WWW, volume 2073 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2018.
- [7] C. Möller, J. Lehmann, R. Usbeck, Survey on english entity linking on wikidata: Datasets and approaches, Semantic Web 13 (2022) 925–966.
- [8] A. Sakor, K. Singh, A. Patel, M. Vidal, Falcon 2.0: An entity and relation linking tool over wikidata, in: CIKM, ACM, 2020, pp. 3141–3148.
- [9] M. Doerr, The CIDOC CRM, an Ontological Approach to Schema Heterogeneity, in: Y. Kalfoglou, M. Schorlemmer, A. Sheth, S. Staab, M. Uschold (Eds.), Semantic Interoperability and Integration, volume 4391 of *Dagstuhl Seminar Proceedings (DagSemProc)*, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, 2005, pp. 1–5. URL: <https://drops.dagstuhl.de/opus/volltexte/2005/35>. doi:10.4230/DagSemProc.04391.22.
- [10] W. R. van Hage, V. Malaisé, R. Segers, L. Hollink, G. Schreiber, Design and use of the simple event model (SEM), J. Web Semant. 9 (2011) 128–136.
- [11] S. Hellmann, J. Lehmann, S. Auer, M. Brümmer, Integrating NLP using linked data, in: ISWC (2), volume 8219 of *Lecture Notes in Computer Science*, Springer, 2013, pp. 98–113.
- [12] J. Devlin, M. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, in: NAACL-HLT (1), Association for Computational Linguistics, 2019, pp. 4171–4186.

- [13] M. Mosbach, M. Andriushchenko, D. Klakow, On the stability of fine-tuning BERT: misconceptions, explanations, and strong baselines, in: ICLR, OpenReview.net, 2021.
- [14] L. Wu, F. Petroni, M. Josifoski, S. Riedel, L. Zettlemoyer, Scalable zero-shot entity linking with dense entity retrieval, in: EMNLP (1), Association for Computational Linguistics, 2020, pp. 6397–6407.
- [15] B. Z. Li, S. Min, S. Iyer, Y. Mehdad, W. Yih, Efficient one-pass end-to-end entity linking for questions, in: EMNLP (1), Association for Computational Linguistics, 2020, pp. 6433–6441.
- [16] J. Raiman, O. Raiman, Deeptype: Multilingual entity linking by neural type system evolution, in: AAAI, AAAI Press, 2018, pp. 5406–5413.
- [17] S. Broscheit, Investigating entity knowledge in BERT with simple neural end-to-end entity linking, in: CoNLL, Association for Computational Linguistics, 2019, pp. 677–685.
- [18] M. Clark, G. Barros, K. Stepanenko, Russian Offensive Campaign Assessment, February 28, 2022, <https://understandingwar.org/backgrounder/russian-offensive-campaign-assessment-february-28-2022>, 2022. Accessed: 2022-09-27.
- [19] K. Stepanenko, L. Philipson, K. Lawlor, F. W. Kagan, Russian Offensive Campaign Assessment, August 1, <https://understandingwar.org/backgrounder/russian-offensive-campaign-assessment-august-1>, 2022. Accessed: 2022-09-27.
- [20] K. Stepanenko, K. Hird, G. Barros, F. W. Kagan, Russian Offensive Campaign Assessment, August 31, <https://understandingwar.org/backgrounder/russian-offensive-campaign-assessment-august-31>, 2022. Accessed: 2022-09-27.
- [21] P. A. Schrod, Automated production of high-volume, real-time political event data, in: Apsa 2010 annual meeting paper, 2010.
- [22] C. Raleigh, R. Kishi, Updates to the Armed Conflict Location & Event Data Project, 2020. URL: https://acleddata.com/acleddatanew/wp-content/uploads/2020/10/ACLED_UpdatesOverview_2020.pdf, accessed: 2022-11-14.
- [23] EM-DAT, Disaster profile for floods. em-dat: International disaster database, 2006.
- [24] T. Ge, L. Cui, B. Chang, Z. Sui, F. Wei, M. Zhou, Eventwiki: A knowledge base of major events, in: LREC, European Language Resources Association (ELRA), 2018.
- [25] S. Gottschalk, E. Demidova, Eventkg - the hub of event knowledge on the web - and biographical timeline generation, CoRR abs/1905.08794 (2019).