# StoryNetworks: An Annotated Dataset of Event Dependencies from Short Descriptions

Daisuke Machizawa[1⋆], Naoki Sawahata[1⋆],
Ryohei Ikejiri[2], and Yasunobu Sumikawa[1]

[1] Takushoku University, Tokyo, Japan sumikawa.lab@gmail.com
[2] Hiroshima University, Hiroshima, Japan
rikejiri@hiroshima-u.ac.jp

**Abstract.** Modeling real-world events as structured graphs is essential for advancing research in information retrieval, digital history, and narrative analysis. In this paper, we propose StoryNetworks, a novel dataset that transforms short event texts into annotated event networks. We curated 5,204 events from the Wikipedia Current Events Portal spanning 2016 and 2017, and manually annotated 2,494 directed dependencies between them. By bridging unstructured textual data with graph-based event modeling, StoryNetworks offers a valuable resource for computational social science and digital humanities. In addition, as creating dependency graph from short texts is a challenging task, this dataset should be useful for designing new models in event evolution modeling, narrative structure analysis, and information diffusion to obtain better accuracy. The dataset is publicly available at `https://github.com/sumilab/dataset`

**Keywords:** Short event texts · event network · Wikipedia · structured dataset · timelines

## 1 Introduction

Events are fundamental units of information in digital libraries, news archives, and knowledge bases. Analyzing them in a structured manner, particularly by extracting dependencies between events, offers significant benefits. One such benefit is the enhanced ability to analogically apply historical knowledge to current societal issues. This ability is widely regarded as essential in education across many countries [2–4]. Structuring events into coherent datasets supports a range of applications, including temporal information retrieval [7, 13, 18, 19, 21], event detection and tracking [19], and event evolution graph creation [20,23]. However, the automatic organization of events remains a challenging task, especially when the source texts are short and lack clear narrative context.

This paper presents a dataset of events extracted from the Wikipedia Current Events Portal (WCEP)[3]. The WCEP is a community-maintained resource that

⋆ co-first authors

[3] `https://en.wikipedia.org/wiki/Portal:Current_events`

captures notable daily events worldwide, presenting them as short summaries grouped by topic and date. Each entry is written concisely, often resembling a news headline, which makes it an ideal yet underutilized source for studying the structure and dynamics of short-form event texts.

The primary contribution of this study is the creation of StoryNetworks, a manually created dataset that transforms short real-world event texts into structured event networks based on dependency relationships. Unlike existing event graph datasets that often rely on long documents or narrative-rich content, StoryNetworks focuses on short texts, which are common in digital libraries and web-based collections but challenging for event understanding. By linking these minimal texts through dependencies, the dataset supports improved model creation for event detection, timeline generation [5], and event evolution graph creation from short descriptions. It also enables the development of information retrieval methods and educational tools that help users explore historical developments through structured, interpretable event relationships.

**Related Work.** Prior work has introduced event datasets with diverse domains and structures. Narrative-focused resources such as GLUCOSE [15] and CaTeRS [14] capture causal and temporal reasoning in fictional settings, while real-world datasets like DocRED [24], EventKG [6], GDELT [12], and TimeBank [17] provide factual relations, large-scale annotations, or temporal information, often derived through automated methods. However, these typically lack explicit, human-annotated dependency structures between events. Several datasets have been proposed for timeline summarization, including Timeline17 [1], Social Timeline [22], and TLS-Covid19 [16]. Timeline17 consists of 4,650 news articles retrieved via Google Search, aligned with 17 manually curated timelines. Social Timeline includes 5,788 articles from CNN, BBC, and The New York Times, with six timelines covering four major events. TLS-Covid19 contains 100,399 news articles in English and Portuguese, defining 178 timelines per language with detailed annotations of COVID-19-related developments. Our proposed dataset, StoryNetworks, differs in two key aspects. First, whereas existing datasets primarily consist of long-form, narrative-rich texts, StoryNetworks is constructed from short-form texts, specifically headline-like entries derived from the WCEP. Second, unlike prior datasets that organize events linearly, StoryNetworks models events as networks with dependency relations, enabling richer representations such as event threading and merging [23]. The W2E dataset [8] is closely related to our work, as it also collectes events from the WCEP and organizes them into topics for timeline construction and topic tracking. Yet, the topics in W2E are fundamentally presented as timelines and are not defined as network structures, as in the present study.

StoryNetworks introduces a directed event graph over 5,204 real-world events from 2016 and 2017, annotated with 2,494 dependencies. This enables structured analysis of event evolution, narrative flow, and information diffusion, offering a richer foundation for modeling how events relate and unfold over time.

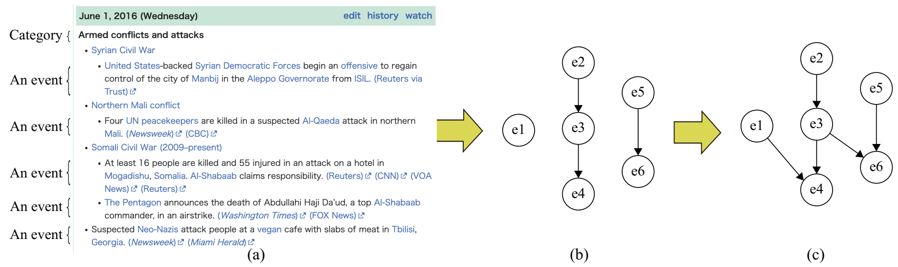


**Fig. 1.** Example events stored in the Wikipedia Current Events Portal. (a) Events of Wikipedia Current Events Portal. (b) Generated timeline example. (c) Timeline connection example

## 2 Data Collection & Creation

Figure 1 illustrates the process of creating our dataset. We began by collecting event data from the WCEP. As shown in Fig. 1(a), WCEP provides a set of short textual descriptions for events, grouped by category. From these collected events, we then constructed timelines, as depicted in Fig. 1(b). Finally, as shown in Fig. 1(c), we connected the timelines to form a network structure.

**Data Collection.** We collected event descriptions as individual events, along with their allocated categories. We collected the following ten categories, which were manually assigned by Wikipedia editors and subsequently organized by [10]: `Law and Crime` (`LC`), `Politics and Election` (`PE`), `Armed Conflict and Attack` (`AA`), `Art and Culture` (`AC`), `International Reations` (`IR`), `Disaster and Accident` (`DA`), `Business and Economy` (`BE`), `Sport` (`S`), `Health and Medicine` (`HM`), and `Science and Environment` (`SE`).

Wikipedia's extensive repository of events, systematically organized by year and date, makes it a well-suited foundation for constructing the event networks central to our research. However, comprehensively covering all available data and manually constructing the network is a prohibitively labor-intensive task and therefore infeasible. Consequently, we decided that we focus on analyzing the period from January 1, 2016, to December 31, 2017 in this study. The rationale for selecting this specific timespan is twofold. First, regarding data quality: event records tend to decrease in number for earlier periods, while the most recent entries may suffer from inconsistent or incomplete documentation. In contrast, the data from this period is both well-documented and relatively stable in quality. Second, the chosen timespan encompasses a diverse range of global events, making it particularly suitable for our research objectives.

**Timeline Creation.** Timelines were manually constructed by four annotators. First, two annotators independently created timelines from the event data. A third validator then reviewed and revised them through discussion with the original annotators. Finally, a fourth inspector reviewed all revisions and removed any entries of questionable validity. We created 2,889 timelines as a result.


**Fig. 2.** Named entities extracted from all event texts

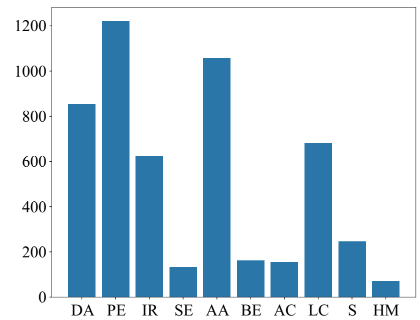


**Fig. 3.** Distribution of the number of occurrences per category

**Network Creation.** The same four annotators followed a similar procedure. Two annotators independently identified dependencies to connect timelines, which were reviewed and revised by a third annotator through discussion. The final inspector then checked all connections and removed any that were questionable. This entire process resulted in the construction of 2,494 dependency relationships.

## 3 Dataset Analysis

This section provides statistical insights into the events, timelines, and networks to characterize the properties of our dataset.

### 3.1 Event Analysis

We began by analyzing the entire event data to highlight two key characteristics: the brevity of the event texts and the diversity of event types. These analyses were conducted on 5,204 event descriptions stored in our dataset.

**Text Analysis.** We calculated the average number of tokens per description and found it to be just 12.5, highlighting the concise nature of the entries. To explore the range of subjects mentioned in these short texts, we analyzed the named entities appearing in the event descriptions using spaCy [9]. This analysis identified 26,815 entities, with an average of 5.1 entities per description. Figure 2 shows the word cloud of the collected entities. As larger words are more frequently mentioned in the dataset, we can see that most of events are related to countries such as the United States, Russia, Syria, China, North Korea, and Turkey. These findings demonstrate that the dataset covers a broad geographical range and includes a wide variety of global actors.

**Category Analysis.** Next, we analyze the types of events included in the dataset. Figure 3 shows the number of occurrences per category. The results indicate that political and conflict-related events constitute the majority of the

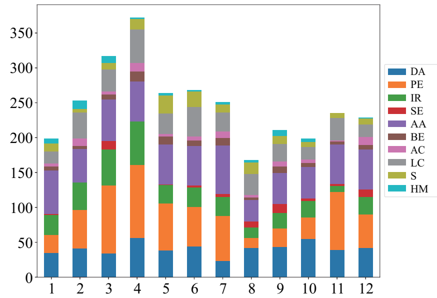


**Fig. 4.** Monthly distribution of event category occurrences in 2016

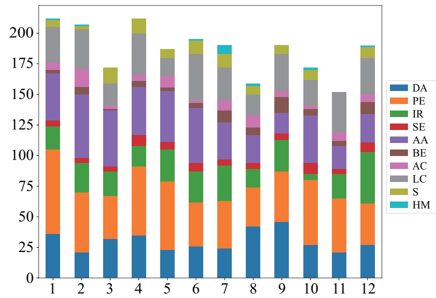


**Fig. 5.** Monthly distribution of event category occurrences in 2017

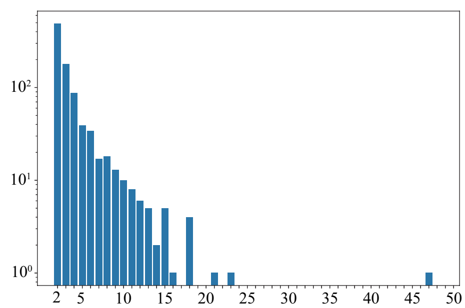

**Fig. 6.** Distribution of the number of events per timeline.

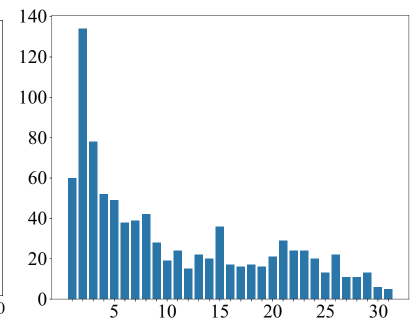

**Fig. 7.** Distribution of durations per timeline.

dataset, reflecting its primary thematic focus. Finally, we examined the temporal consistency of these category distributions. Figures 4 and 5 present stacked bar charts illustrating the number of categories per month for 2016 and 2017, respectively. These results confirm that political and conflict-related events consistently remain among the top categories; however, events from other categories are also recorded in each month.

### 3.2 Timeline Analysis

We performed quantitative analyses on the timelines in our dataset to investigate their sizes and the degree of similarity between events within each timeline.

**Size Analysis.** Figure 6 illustrates the distribution of timeline sizes. A timeline’ s size is defined as the number of events it contains. The horizontal axis represents timeline size, while the vertical axis shows the count on a logarithmic scale. As shown in the figure, the most frequent timeline size is 2, and the number of timelines decreases as size increases. The largest timeline contains 47 events. We conducted a detailed examination of the results and found that timelines related to natural disasters and sports tend to be smaller in size, whereas those

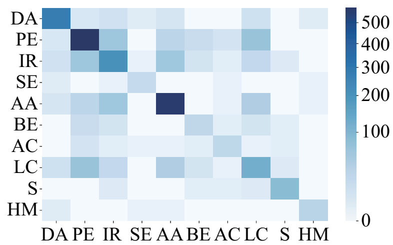


**Fig. 8.** Frequencies of category combinations within timelines

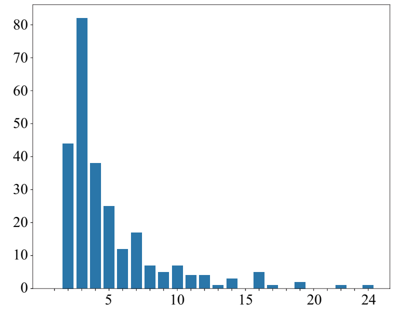


**Fig. 9.** Distribution of the number of events per network

concerning conflicts and elections, especially, U.S. presidential election, tend to be larger.

We next analyzed the timespans of all timelines. For each timeline, we extracted the timestamps of the first and last events, then calculated the duration by subtracting the date of the first event from that of the last and adding one day. Figure 7 illustrates the distribution of timeline durations, with the horizontal axis indicating duration and the vertical axis showing the count. The results show that a substantial number of timelines span only a few days. Moreover, the number of timelines tends to decrease as the duration increases.

**Similarity Analysis.** Next, we analyzed the degree of similarity among events grouped within each timeline. This analysis focused on two aspects: the extent of shared vocabularies in the event texts within a timeline, and the combinations of event categories present in each timeline. In particular, the latter analysis aimed to examine whether events within the same timeline tend to share similar categories, reflecting the assumption that timelines typically group events with related content.

We computed two types of text similarity measures: (1) TF-IDF with cosine similarity based on all tokens, and (2) the Jaccard coefficient based on extracted entities. For every pair of events within each timeline, these similarity scores were calculated. The average values were 0.101 and 0.00044, respectively. These low scores indicate the difficulty of constructing timelines from short event texts, as shared tokens and entities are relatively uncommon. Consequently, simple text-based similarity models are insufficient for identifying coherent timelines. Incorporating contextual information from external sources is therefore essential for effective timeline construction.

We then counted the number of category combinations. For all defined dependencies within each timeline, we collected the associated events and their categories. Figure 8 plots the frequency of category combinations within timelines. We can see that for many categories, the most frequent combination was with the same category; in other words, many dependencies were defined between events of the same type.


**Fig. 10.** Combination of event categories theta connect timelines

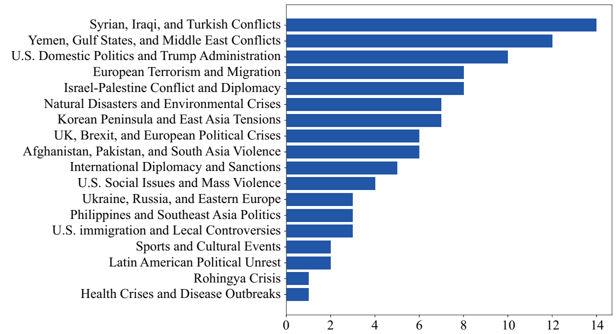


**Fig. 11.** The number of networks

### 3.3 Network Analysis

We conducted a final analysis focusing on the following three aspects of the networks: (1) the number of events contained in each network; (2) the number of combinations of event categories that serve as connection points between timelines; and (3) the contents represented across all networks.

Figure 9 shows the distribution of the number of events included in each network. We observe that the number of networks peaks at three events, with a subsequent decline as the number of events increases. This is a natural outcome, as the networks in this study are constructed by merging timelines, many of which are relatively small. Consequently, the number of events grouped into a single network tends not to be excessively large.

Next, Fig. 10 shows frequencies of category pairs for events connected between timelines. This result indicates that events within the same category often served as connection points across many categories. However, it is observed that the four categories `IR`, `PE`, `AA`, and `LC` exhibit high inter-connectivity with other categories. This suggests that, during the period from 2016 to 2017, numerous conflicts and policy developments involving multiple countries occurred, and events classified under `AA` and `LC` frequently acted as triggers for those in `IR` and `PE`, or vice versa. In contrast, `S` events had dependencies only on events within the same category; this is a distinctive characteristic of this category.

Finally, we analyzed the contents of the networks included in the dataset. To conduct this analysis, we assigned a name to each network and grouped networks with similar names. We then counted the number of networks belonging to each group. The initial naming and grouping were drafted by GPT-4, and the resulting assignments were reviewed and validated by two human annotators. Figure 11 presents the results of this analysis. The results indicate that many of the networks pertain to conflicts involving multiple countries, the policies of countries involved, and the 2016 U.S. presidential election. In addition, natural disasters constitute the sixth most frequent topic, while sports-related events, primarily associated with the 2016 Olympics, rank fifteenth in frequency.

## 4 Potential Use Cases

StoryNetworks enables structured analysis of over 5,000 short-form event descriptions by organizing them into dependency-based networks. This is especially valuable for studying short texts, which are prevalent in digital libraries and web archives but often difficult to interpret due to limited context and sparse information.

A key use case is the development of machine learning models for tasks such as event detection, linking, and timeline construction from short texts, where systems must determine whether multiple brief descriptions refer to the same underlying event. These tasks are particularly challenging when crucial details are missing or only indirectly referenced. StoryNetworks provides a realistic benchmark for advancing model performance in such low-context scenarios.

Beyond computational tasks, StoryNetworks also supports novel information retrieval methods, particularly in educational and social science contexts. Its network structure enables retrieval based on causal, temporal, or thematic relationships rather than simple keyword matching. In history education, this makes possible intelligent search interfaces that allow students and educators to trace the development of historical situations, understand cause-and-effect dynamics, and discover connections between past and present events [11]. Retrieval systems built on StoryNetworks can facilitate exploratory learning and narrative reconstruction by surfacing relevant event paths or clusters based on meaningful relationships [2, 3]. Social scientists can also leverage the dataset to analyze the progression of political events, policy decisions, or international developments through structured dependencies that illustrate how actions lead to outcomes.

These use cases demonstrate how StoryNetworks serves both computational and interdisciplinary goals. By centering on short-form, real-world texts and explicitly modeling their interdependencies, the dataset supports robust event understanding and semantic access in information-sparse environments. This enables both automated systems and human users to reason effectively with minimal but richly connected data.

## 5 Conclusion

We presented StoryNetworks, a novel dataset of real-world events derived from the WCEP, which organizes over 5,000 short event texts into structured networks based on manually annotated dependencies. By focusing on short-form content, our dataset addresses a critical gap in event understanding and information retrieval research, where minimal context and textual sparsity pose unique challenges. StoryNetworks enables a range of applications, including event detection, timeline construction, and semantic enrichment of digital library collections. It also opens new opportunities for developing retrieval algorithms and educational tools that support exploration of historical narratives through structured event relationships. In future work, we plan to expand the dataset to additional years and investigate automatic methods for dependency annotation to support larger-scale analysis.

## References


1. Binh Tran, G., Alrifai, M., Quoc Nguyen, D.: Predicting relevant news events for timeline summaries. In: Proceedings of the 22nd International Conference on World Wide Web. pp. 91–92. WWW '13 Companion, Association for Computing Machinery, New York, NY, USA (2013). https://doi.org/10.1145/2487788.2487829, https://doi.org/10.1145/2487788.2487829
2. Boix-Mansilla, V.: Historical understanding: Beyond the past and into the present. Knowing, Teaching, and Learning History: National and International Perspectives pp. 390–418 (2000)
3. van Drie, J., van Boxtel, C.: Historical reasoning: Towards a framework for analyzing students' reasoning about the past. Educational Psychology Review **20**(2), 87–110 (2008)
4. Ministry of Education, Culture, S.S., (MEXT), T.: Course of study for upper secondary schools (public notice of heisei 30) (2018), https://www.mext.go.jp/component/a_menu/education/micro_detail/__icsFiles/afieldfile/2019/11/22/1407073_03_2_2.pdf, [In Japanese]
5. Gholipour Ghalandari, D., Ifrim, G.: Examining the state-of-the-art in news timeline summarization. In: Jurafsky, D., Chai, J., Schluter, N., Tetreault, J. (eds.) Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 1322–1334. Association for Computational Linguistics, Online (Jul 2020). https://doi.org/10.18653/v1/2020.acl-main.122, https://aclanthology.org/2020.acl-main.122
6. Gottschalk, S., Demidova, E.: Eventkg: A multilingual event-centric temporal knowledge graph. In: The Semantic Web: 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3 – 7, 2018, Proceedings. pp. 272–287. Springer-Verlag, Berlin, Heidelberg (2018). https://doi.org/10.1007/978-3-319-93417-4_18, https://doi.org/10.1007/978-3-319-93417-4_18
7. Hayashi, K., Maehara, T., Toyoda, M., Kawarabayashi, K.i.: Real-time top-r topic detection on twitter with topic hijack filtering. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 417–426. KDD '15, Association for Computing Machinery, New York, NY, USA (2015). https://doi.org/10.1145/2783258.2783402, https://doi.org/10.1145/2783258.2783402
8. Hoang, T.A., Vo, K.D., Nejdl, W.: W2e: A worldwide-event benchmark dataset for topic detection and tracking. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management. pp. 1847–1850. CIKM '18, Association for Computing Machinery, New York, NY, USA (2018). https://doi.org/10.1145/3269206.3269309, https://doi.org/10.1145/3269206.3269309
9. Honnibal, M., Montani, I.: spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing (2017)
10. Košmerlj, A., Belyaeva, E., Leban, G., Grobelnik, M., Fortuna, B.: Towards a complete event type taxonomy. In: Proceedings of the 24th International Conference on World Wide Web. pp. 899–902. WWW '15 Companion, Association for Computing Machinery, New York, NY, USA (2015). https://doi.org/10.1145/2740908.2742005, https://doi.org/10.1145/2740908.2742005
11. Lee, P.: Historical literacy: Theory and research. International Journal of Historical Learning, Teaching and Research **5**(1), 25–40 (2005)

12. Leetaru, K., Schrodt, P.A.: Gdelt: Global data on events, location, and tone. ISA Annual Convention (2013), http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.686.6605
13. Liu, W., Wang, H., Wang, J., Guo, H., Sun, Y., Hou, M., Yu, B., Wang, H., Peng, Q., Zhang, C., Liu, C.: A popular topic detection method based on microblog images and short text information. Journal of Web Semantics **81**, 100820 (2024). https://doi.org/https://doi.org/10.1016/j.websem.2024.100820, https://www.sciencedirect.com/science/article/pii/S1570826824000064
14. Mostafazadeh, N., Grealish, A., Chambers, N., Allen, J., Vanderwende, L.: CaTeRS: Causal and temporal relation scheme for semantic annotation of event structures. In: Palmer, M., Hovy, E., Mitamura, T., O'Gorman, T. (eds.) Proceedings of the Fourth Workshop on Events. pp. 51–61. Association for Computational Linguistics, San Diego, California (Jun 2016). https://doi.org/10.18653/v1/W16-1007, https://aclanthology.org/W16-1007/
15. Mostafazadeh, N., Kalyanpur, A., Moon, L., Buchanan, D., Berkowitz, L., Biran, O., Chu-Carroll, J.: GLUCOSE: GeneraLized and COntextualized story explanations. In: Webber, B., Cohn, T., He, Y., Liu, Y. (eds.) Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 4569–4586. Association for Computational Linguistics, Online (Nov 2020). https://doi.org/10.18653/v1/2020.emnlp-main.370, https://aclanthology.org/2020.emnlp-main.370/
16. Pasquali, A., Campos, R., Ribeiro, A., Santana, B., Jorge, A., Jatowt, A.: Tls-covid19: A new annotated corpus for timeline summarization. In: Advances in Information Retrieval: 43rd European Conference on IR Research, ECIR 2021, Virtual Event, March 28 – April 1, 2021, Proceedings, Part I. pp. 497–512. Springer-Verlag, Berlin, Heidelberg (2021). https://doi.org/10.1007/978-3-030-72113-8_33, https://doi.org/10.1007/978-3-030-72113-8_33
17. Pustejovsky, J., Castano, J., Ingria, R., Sauri, R., Gaizauskas, R., Setzer, A., Katz, G., Radev, D.: Timebank: A corpus for annotating temporal information. In: Proceedings of the Corpus Linguistics Conference. pp. 647–656. Lancaster, UK (2003)
18. Qi, Y., Zhou, L., Si, H., Wan, J., Jin, T.: An approach to news event detection and tracking based on stream of online news. In: 2017 9th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC). vol. 2, pp. 193–196 (2017). https://doi.org/10.1109/IHMSC.2017.158
19. Radinsky, K., Horvitz, E.: Mining the web to predict future events. In: Proceedings of the Sixth ACM International Conference on Web Search and Data Mining. pp. 255–264. WSDM '13, Association for Computing Machinery, New York, NY, USA (2013). https://doi.org/10.1145/2433396.2433431, https://doi.org/10.1145/2433396.2433431
20. Sawahata, N., Machizawa, D., Ikejiri, R., Sumikawa, Y.: Llm-based dependency tracking for short event descriptions. In: The 29th International Conference on Theory and Practice of Digital Libraries. TPDL '25, Springer-Verlag, Berlin, Heidelberg (2025), https://sumilab.github.io/web/pdf/2025/tpdl2025_sawahata.pdf, accepted
21. Tan, Z., Zhang, P., Tan, J., Guo, L.: A multi-layer event detection algorithm for detecting global and local hot events in social networks. Procedia Computer Science **29**, 2080–2089 (2014). https://doi.org/https://doi.org/10.1016/j.procs.2014.05.192, https://www.sciencedirect.com/science/article/pii/S187705091400369X, 2014 International Conference on Computational Science

22. Wang, L., Cardie, C., Marchetti, G.: Socially-informed timeline generation for complex events. In: Mihalcea, R., Chai, J., Sarkar, A. (eds.) Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. pp. 1055–1065. Association for Computational Linguistics, Denver, Colorado (2015). https://doi.org/10.3115/v1/N15-1112, https://aclanthology.org/N15-1112/
23. Yang, C.C., Shi, X., Wei, C.P.: Discovering event evolution graphs from news corpora. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans **39**(4), 850–863 (2009). https://doi.org/10.1109/TSMCA.2009.2015885
24. Yao, Y., Ye, D., Li, P., Han, X., Lin, Y., Liu, Z., Liu, Z., Huang, L., Zhou, J., Sun, M.: DocRED: A large-scale document-level relation extraction dataset. In: Korhonen, A., Traum, D., Màrquez, L. (eds.) Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. pp. 764–777. Association for Computational Linguistics, Florence, Italy (Jul 2019). https://doi.org/10.18653/v1/P19-1074, https://aclanthology.org/P19-1074/