StoryNetworks: An Annotated Dataset of Event Dependencies from Short Descriptions

Daisuke Machizawa[1], Yasunobu Sumikawa[1], Ryohei Ikejiri[2], and Naoki Sawahata[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 introduce StoryNetworks, a novel dataset that transforms short event texts into richly annotated event networks. Drawing from the 2016 and 2017 Wikipedia Current Events Portal, we curated 5,204 events and manually annotated 2,494 directed dependencies that represent temporal, causal, or contextual relationships between events. This structured representation enables new research directions in event evolution modeling, narrative structure analysis, and information diffusion. By bridging unstructured textual data with graph-based event modeling, StoryNetworks offers a valuable resource for computational social science and digital humanities. The dataset is publicly available at `https://https://shorturl.at/eplnx`

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. Structuring events into coherent datasets supports a range of applications, including temporal information retrieval, event detection and tracking, and narrative construction. However, the automatic organization of events remains a challenging task, especially when the source texts are short and lack clear narrative context.

This paper introduces a curated dataset of events extracted from the Wikipedia Current Events Portal (WCEP)[3]. The WCEP is a community-maintained resource that captures notable daily events worldwide, presenting them as short

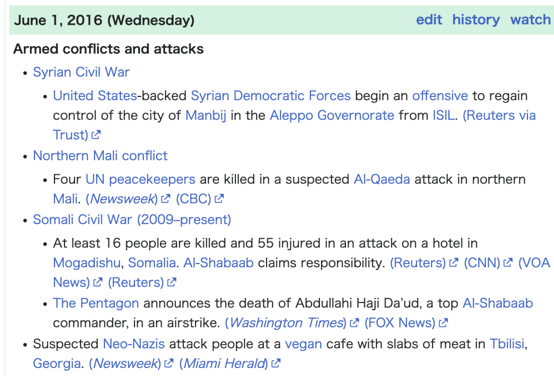
[3] `https://en.wikipedia.org/wiki/Portal:Current_events`

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 curated 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 research in event detection, timeline generation, and entity tracking in low-context settings. 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 [8] and CaTeRS [7] capture causal and temporal reasoning in fictional settings, while real-world datasets like DocRED [12], EventKG [2], GDELT [6], and TimeBank [10] 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 [11], and TLS-Covid19 [9]. 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, while existing datasets focus on long-form, narrative-rich texts, StoryNetworks is built from short-form texts—specifically, headline-like entries from the WCEP. This supports event analysis in low-context settings, common in digital libraries and web archives. 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. The W2E dataset [3] is closely related to our work, as it also draws from the WCEP and organizes events into topics for timeline construction and topic tracking. Yet, topics of the W2E may contain only single events, and its structure remains limited to flat clusters.

In contrast, 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.


June 1, 2016 (Wednesday) edit history watch

Armed conflicts and attacks

- Syrian Civil War
  - United States-backed Syrian Democratic Forces begin an offensive to regain control of the city of Manbij in the Aleppo Governorate from ISIL. (Reuters via Trust)
- Northern Mali conflict
  - Four UN peacekeepers are killed in a suspected Al-Qaeda attack in northern Mali. (*Newsweek*) (CBC)
- Somali Civil War (2009–present)
  - At least 16 people are killed and 55 injured in an attack on a hotel in Mogadishu, Somalia. Al-Shabaab claims responsibility. (Reuters) (CNN) (VOA News) (Reuters)
  - The Pentagon announces the death of Abdullahi Haji Da'ud, a top Al-Shabaab commander, in an airstrike. (*Washington Times*) (FOX News)
- Suspected Neo-Nazis attack people at a vegan cafe with slabs of meat in Tbilisi, Georgia. (*Newsweek*) (*Miami Herald*)

Fig. 1. An example of the Wikipedia Current Events Portal. This portal served as the data source for our short event texts.

2 Data Collection & Creation

Data Collection. This study utilized WCEP as the primary source for data collection. Figure 1 shows example events recorded in WCEP. WCEP enumerates events in the form of short textual descriptions for each event category. We collected these descriptions as individual events, along with their allocated categories. As shown in Fig. 1, some events are accompanied by a title. We appended this title to the beginning of the event description. As a result, in this example, we collected five events under the `Armed Conflicts and Attacks` category. We collected following ten categories, as defined by [5]: `Law and Crime` (`LC`), `Politics and Election` (`PE`), `Armed Conflict and Attack` (`AA`), `Art and Culture` (`AC`), `International Relations` (`IR`), `Disaster and Accident` (`DA`), `Business and Economy` (`BE`), `Sport` (`S`), `Health and Medicine` (`HM`), and `Science and Environment` (`SE`). These categories were manually assigned by editors on Wikipedia, the source of our data. 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 confined our scope of analysis to the period from January 1, 2016, to December 31, 2017. The rationale for selecting this specific timespan is twofold. First, data quality: while older entries frequently suffer from broken links and the most recent entries can be poorly structured, the data from this period is comprehensively documented with relatively stable quality. Second, this period witnessed a diverse range of global events, which we deemed highly suitable for our research objectives.

Timeline Creation. The construction of timelines was a manual process conducted by four individuals. First, two annotators independently created timelines from the event data. Next, to validate the resulting timelines, a third independent validator reviewed each one. During this validation process, the validator and the original annotator reached a consensus through discussion and applied the necessary revisions. A fourth inspector then reviewed all revised data, and any entries with questionable validity were removed. This process resulted in the construction of 921 timelines. **Network Creation.** Four annotators manually performed the construction of connections between timelines. Initially, two annotators independently identified timelines with dependency relationships and established the connections. Subsequently, for the validation of these created connections, a third annotator reviewed all connections. Through discussions between the third annotator and the original creators, necessary revisions were implemented at that time. Following this, a fourth annotator reviewed the revised data and conducted a validity check. This entire process resulted in the construction of 138 dependency relationships.

3 Dataset Analysis

This section presents an analysis of event characteristics, timeline structure, and network structure, conducted to elucidate the properties of the constructed dataset.

3.1 Event Characteristic Analysis

The dataset comprises 5,204 event descriptions, containing a total of 65,422 tokens. Each event is composed of a concise sentence averaging 12.5 tokens.

To assess the comprehensiveness of the collected events, named entities were extracted using spaCy [4]. Of these, 6,542 unique entities were identified. This observation suggests that event descriptions within the dataset encompass a broad and diverse range of proper nouns, rather than being confined to a limited set of specific personal names, organizational names, or geographical locations.

To further illustrate the specific entities included, Figure 2 presents a word cloud of the proper nouns. This visualization reveals the frequent occurrence of various countries, including the United States, China, Russia, and Syria. This finding strongly corroborates that the dataset encompasses diverse events across a broad geographical range.

Moreover, to provide a detailed understanding of the event subjects, Figure 3 illustrates the distribution of categories. Figure 3 indicates that political and conflict-related topics, such as `PE` and `AA`, constitute the majority of the dataset and represent its central themes. However, the dataset also includes a diverse range of other categories, such as `DA`, `IR`, `BE`, `AC`, `LC`, `S`, `HM`, and `SE`, which demonstrates its coverage of events pertaining to various facets of society.

Finally, the temporal stability of the category distribution was examined. Figures 5 and 4 present the total number of events and the categorical breakdown for each month in 2016 and 2017, respectively. These figures confirm that the aforementioned category composition exhibits a relatively stable proportion throughout the year, with no significant monthly fluctuations. This consistent

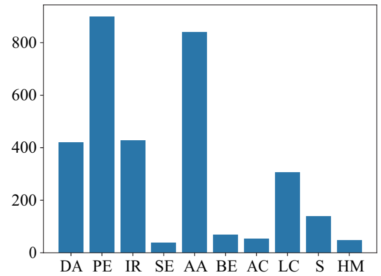


Fig. 3. Distribution of event categories

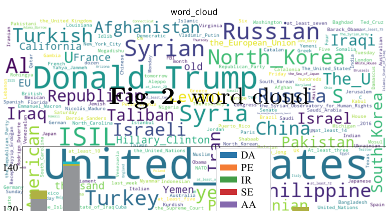


Fig. 2. word cloud

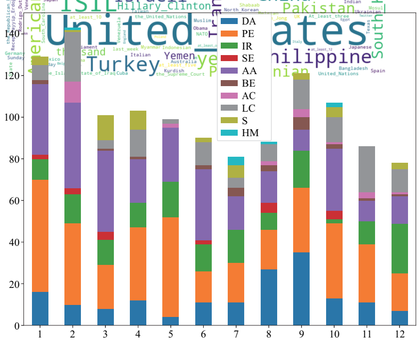


Fig. 4. Distribution of events and categories by month 2017

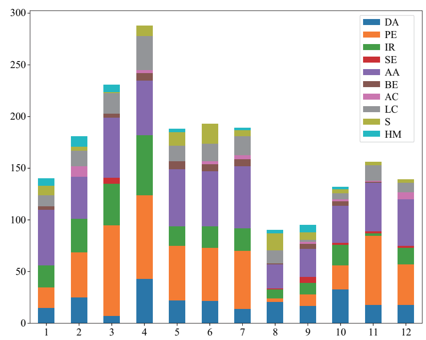


Fig. 5. Distribution of events and categories by month 2016

distribution suggests that, in addition to primary topics, the dataset consistently captures social events across various other categories throughout the year.

3.2 Analysis of Timeline Structure

3.3 タイムライン構造の分析

本データセットを構成するタイムラインの構造的特性を明らかにするため、その複雑性、時間的スケール、および内部構造について定量的な分析を行った。**複雑性と時間的スケール。**タイムラインにおけるイベントの規模を評価するために、各タイムラインを構成するイベントの長さの分布を分析した。図 6 に示すように、タイムラインあたりのイベント数で最も頻繁なのは 2 個であり、最大は 47 個であった。この結果は、比較的少数のイベントで構成される単純なタイムラインが多い一方で、数十のイベントの規模が大きいタイムラインも少数ながら存在し、イベント数において幅広い分布を示す。タイムラインが捕捉する事象の時間的スケールを評価するため、図 7 に示す期間の分布を示す。タイムライン内の最初と最後のイベント間の最も一般的な時間差が 2 日間であるのに対し、平均は 15.0 日間、最大は 30 日間であることを示している。これらの結果は、本データセットが、イベント数および期間の両面で幅広い分布を示しており、少数のイベントか

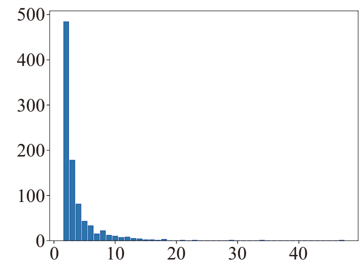


Fig. 6. Distribution of length of timelines

らなる単純なタイムラインから、比較的多数のイベントを含み、数週間にわたるものまで、多様な構造的特性を有していることを示唆する。**タイムラインの内部構造。**タイムラインを構成するイベント間の関係性を評価するため、イベント間の依存関係を調査した。図 8 は、隣接するイベントカテゴリ間の集計された遷移を示している。データは、遷移が主に同じカテゴリ内で発生し、異なるカテゴリ間の遷移はまれであることを示している。このことは、タイムラインが複数のトピックを横断するのではなく、単一のトピックを時系列的に追跡する傾向があることを示唆している。さらに、イベント間の接続がどのような性質を持つかを評価する。接続にはテキストから判断してタイムラインを作成したため、タイムライン内の各イベント同士のテキスト類似度スコアとエンティティ類似度スコアを計算した。TF-IDF とコサイン類似度で測定された平均テキスト類似度は 0.101 であり、Jaccard インデックスで測定された平均エンティティ類似度は 0.00044 であった。類似度スコアの低さは、タイムラインの形成において、共有語や固有表現といった表層的手がかりが接続の基準とはなっていない可能性を示している。これは、本データセットが人手によるアノテーションを用いて構築されており、イベント間の文脈的・因果的関係に基づく接続が考慮されていることを示している。To elucidate the structural characteristics of the timelines that constitute our dataset, we conducted a quantitative analysis of their complexity, temporal scale, and internal structure.

Complexity and Temporal Scale. To evaluate the scale of each timeline, we analyzed the number of constituent events and their duration. As shown in Fig. 6, the most frequent number of events per timeline is two, with an average of 3.5 and a maximum of 47. Furthermore, the distribution of durations, shown in Fig. 7, indicates that the most common time difference between the first and last event in a timeline is two days, while the average is 15.0 days and the maximum extends to 30 days. This dataset encompasses a diverse range of complexity and temporal scales, from simple timelines with few events to complex, long-term phenomena spanning up to a month.

Internal Structure of Timelines. To evaluate the internal structure of the timelines, the connections between events were investigated. Figure 8 shows the aggregated transitions between the categories of adjacent events. The data indicate that transitions predominantly occur within the same category, while transitions between different categories are infrequent. This suggests that time-

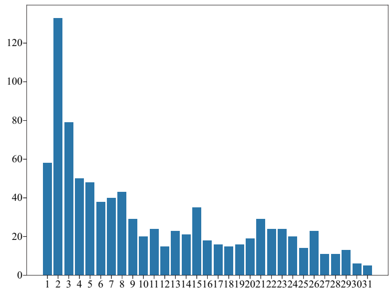


Fig. 7. Time distribution of timeline

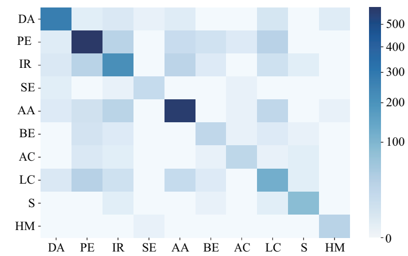


Fig. 8. Categories of events summarised on the timeline

Table 1. Top 10 prominent international political topics.

Rank	Topic	Rank	Topic
1	U.S. politics and elections	6	Yemeni Civil War
2	Syrian Civil War	7	North Korean nuclear and missile program
3	Brexit and UK domestic politics	8	War in Afghanistan
4	Conflict with ISIL and politics in Iraq	9	Disasters and accidents in the U.S.
5	Politics of the Philippines	10	French politics and elections

lines tend to track a single topic chronologically, rather than traversing multiple topics. Furthermore, to examine the basis for cohesion, text and entity similarity scores were calculated. The average text similarity, measured with TF-IDF and cosine similarity, was 0.101, and the average entity similarity, measured with the Jaccard index, was 0.00044. These low scores suggest that the formation of timelines is not reliant on superficial cues, such as shared words or named entities. Instead, the evidence implies that they are formed by semantic connections, including contextual relevance and causal relationships.

Topics of timeline To grasp the specific content covered by the entire dataset, we extracted its central entities and topics. Furthermore, Table 1 shows that major topics involving these entities, such as "U.S. politics and elections" and the "Syrian Civil War," appear frequently. These findings confirm that our dataset primarily covers international politics, with a particular focus on U.S. affairs and Middle Eastern situations.

3.4 Analysis of Network Structure

To understand the structural characteristics of the network, we analyzed the length distribution of the timelines that form the network. Figure 9 reveals that the network is primarily composed of connected timelines that themselves consist of a small number of events. Furthermore, the average degree of connectivity is 1.21, indicating that the dataset is characterized by linear chain structures limited to specific themes or periods.

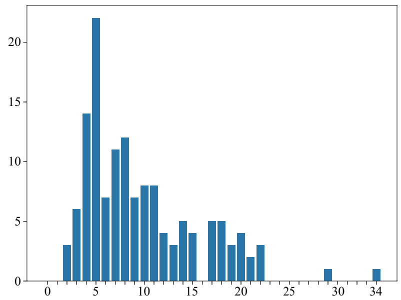


Fig. 9. Distribution of events within the network

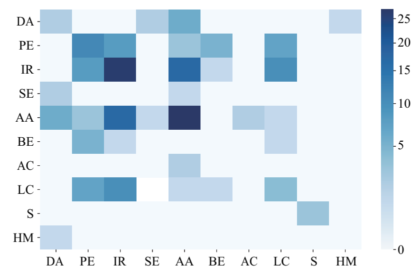


Fig. 10. Combination of event categorise theta connect timelines

To evaluate the thematic coherence of the network, we analyzed the combination of categories for connected events, with the results shown in Fig. 10. The strong tendency for events of similar categories to have dependencies on each other indicates that, much like the individual timelines, the networks are also predominantly constructed around a single theme.

4 Potential Use Cases

The value of our dataset lies not only in its structured representation of real-world events but also in its explicit focus on short-form texts. These short texts, such as news headlines, summaries, and brief bullet points, are increasingly prevalent in digital libraries, news archives, and web collections. However, they pose significant challenges for event understanding due to their limited context, ambiguity, and information sparsity. StoryNetworks addresses these challenges by organizing over 5,000 short event descriptions into dependency-based networks, enabling structured analysis of how events relate to and evolve from one another.

A prominent use case is event detection and linking from short texts, where systems must determine whether multiple brief descriptions refer to the same underlying event. This task is particularly difficult when key details are missing or indirectly referenced. Our dataset provides manually curated groupings and dependencies among event texts, offering a realistic benchmark for clustering algorithms, event detection, temporal reasoning, or modeling of event causality and evolution in low-context environments. It also supports timeline construction from sparse data sources such as headline feeds or summary portals. Unlike traditional approaches that rely on long documents rich in temporal cues, StoryNetworks enables research on how coherent event timelines can be derived from minimal textual evidence. This supports the development of scalable timeline generation tools suitable for collections containing only metadata or summaries.

StoryNetworks also presents a strong opportunity for developing novel information retrieval methods, especially in educational and social science contexts.

The event network structure allows for retrieval based on causal, temporal, or thematic relationships rather than simple keyword matching. In history education, this opens the door to intelligent search interfaces where students and educators can trace the development of historical situations, understand cause-and-effect patterns, or discover connections between present and past events. Retrieval systems built on StoryNetworks can support exploratory learning and narrative reconstruction by surfacing relevant event paths or clusters based on meaningful event relationships. Social scientists can use the dataset to analyze the progression of political events, policy outcomes, or international developments across time, supported by structured dependencies that illustrate how actions led to specific results.

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 WCEPortal, 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. 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

3. 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
4. Honnibal, M., Montani, I.: spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing (2017)
5. 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
6. 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
7. 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/
8. 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/
9. 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
10. 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)
11. 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/
12. 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/