

Event Extraction Alone Is Not Enough

Junbo Huang^{1,*}, Longquan Jiang¹, Cedric Möller¹ and Ricardo Usbeck^{2,*}

¹University of Hamburg, Department of Computer Science, Vogt-Kölln-Straße 30, 22527 Hamburg, Germany

²Leuphana University Lüneburg, Universitätsallee 1, 21335 Lüneburg, Germany

Abstract

With the growing amount of online data, distinguishing between similar events and news about them poses a significant challenge for both companies and crisis reaction units. To discriminate event instances, we present *Eventist*, a silver-standard event instance dataset from news in English, containing 23,304 news headlines from 90 countries covering in total 113 storm-related events between 1st January 2021 and 1st September 2023. Sampled data is validated by two human raters. Additionally, we propose to adopt a sentence-level event representation for modeling media narrative discourse. Finally, we provide two pairwise comparison benchmarks on event deduplication and event temporal ordering, enabling the practicality of event extraction.

Keywords

Event Deduplication, Event Temporal Ordering, Media Narrative Discourse

1. Introduction

With climate change, a noticeable uptick in the frequency and severity of tropical storms and their associated impacts is observed [1, 2]. Notably, storm surges, including extreme floods and tsunamis, often result in more significant damage than the storms themselves. To automatically identify storm events holds the potential for fast crisis responses. Consequently, event extraction from media texts has emerged as a pivotal task in Natural Language Processing (NLP) [3, 4].

We view media texts as narratives describing specific events with their often implicit causal or temporal relations. We address news headlines as compressed narratives containing partial event information. An event, in this context, denotes the real-world occurrence of a particular incident, attributed to its constituent elements such as participants, temporal attributes, and geographical location, often resulting in a change of state of a set of geopolitical entities (GPEs) [5]. In this study, we represent events with English-language news headlines, with a focus on storm-related narratives. Specifically, storm-related narratives are texts with at least one storm-related event mention.

In NLP, events are referred to as word-based event mentions, typically represented as verbs or nouns [3, 4, 6]. For instance, in the headline “Latest from the Tropics: **Tropical Storm Bret**

In: R. Campos, A. Jorge, A. Jatowt, S. Bhatia, M. Litvak (eds.): *Proceedings of the Text2Story’24 Workshop, Glasgow (Scotland), 24-March-2024*

*Corresponding author.

✉ junbo.huang@uni-hamburg.de (J. Huang); longquan.jiang@uni-hamburg.de (L. Jiang);

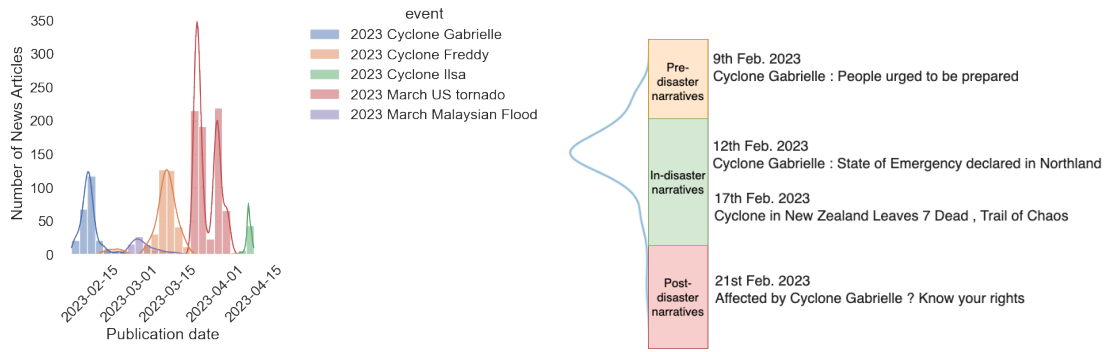
cedric.moeller@uni-hamburg.de (C. Möller); ricardo.usbeck@leuphana.de (R. Usbeck)

🆔 0000-0002-3192-5896 (J. Huang); 0000-0002-7333-2589 (L. Jiang); 0000-0001-6700-3482 (C. Möller);

0000-0002-0191-7211 (R. Usbeck)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



(a) Event distribution over temporally overlapping events. Note that multiple events can have a compound effect on a set of geopolitical entities. Such compound effect can not be captured by extracting only word-based event mentions. (b) Sentence-level event representation on 2023 Cyclone Gabrielle in different stages of event development. This representation captures event attributes, the effect of events and the telling of the events, known as narrative discourse.

Figure 1: Distribution of Events and Illustration on Different Media Narratives Overtime

weakens while **Cindy** strengthens”, the extracted word-based event mentions are “Tropical Storm Bret” (formed on June 19, 2023) and “Cindy” (formed on June 27, 2023). We argue for the significance of recognizing narrative discourses involves nuances in news storytelling. In this example, the narrative’s temporal focus shifts from the decline of tropical storm Bret to the emergence of tropical storm Cindy. We consider the self-referential change of state, of one being weakened and another being strengthened, as the effect of an event, represented by a single news headline containing two word-based event mentions.

A significant yet often overlooked challenge lies in determining whether different news articles pertain to the same event. Discrepancies between word-based event mentions and events present a formidable obstacle known as event instance discrimination [7, 8, 9]. Despite its critical importance, there exists no standardized benchmark or task formulation to address this issue. Prior attempts have involved entity matching based on factors such as location and time [7, 8, 10], or linking event mentions to entries in a knowledge graph (KG) like Wikidata [9]. However, heuristic-driven methods often suffer from poor generalization, while KG-driven approaches are limited by the inability to represent events absent from the KG. More importantly, extracting only word-based event mentions fails to represent the effect of events.

Furthermore, annotating mentions at the word level is a costly process, often resulting in low agreement among annotators [11, 12, 13]. Evidence suggests that word-based event mentions are not crucial for effective event detection [14, 15]. In our approach, we utilize sentence-level event representation, as illustrated in Figure 1b. Sentence-level event representation can better capture (1) event attributes, (2) the effect of events, and (3) the telling of the events, also known as narrative discourse [16]. Narrative discourse varies before, during, and after the event, which is practically important in many disciplines [3, 4, 17, 18].

To achieve this goal of distinguishing event instances in news streams (Figure 1a), drawing inspiration from the success of deep distance metric learning [19, 20], we suggest a baseline

system for two pairwise comparison tasks. Our contributions include:

- Introducing a large-scale silver-standard event instance dataset named *Eventist*. The dataset comprises 23,304 English news headlines, covering a total of 113 storm-related real-world events from 90 countries between January 1, 2021, and September 1, 2023.
- Providing two benchmarks on event deduplication and event temporal ordering. We have made the dataset, baseline models, and code openly accessible on <https://github.com/semantic-systems/paper-event-instance-discrimination>.

2. Related Work

Dataset To our knowledge, the most closely related dataset concerning crisis-related news articles linked to KGs is Crisisfacts [4], initially designed for extracting atomic facts for temporal summarization. Events are linked to Wikidata, and publication dates are available as meta-information. It is theoretically possible to re-formulate the temporal summarization task into event deduplication and event temporal ordering. However, as depicted in Table 1, Crisisfacts encompasses only 8 events spread across 31 unique dates (all confined to the United States), limiting its applicability. In contrast, *Eventist* contains news headlines describing 113 different events, covering 688 unique dates, and a significantly longer narrative duration.

Table 1
Dataset Descriptives.

	#Events	#Dates	#News	Mean Duration	Max Duration	Min Duration
Crisisfacts	8	31	41,147	3.88	7	2
Eventist	113	688	23,274	10.69	71	2

Methods for Event Deduplication Heuristic-based event deduplication methods first extract entities from texts, which are used as features for event deduplication. [10] deployed a graph-based event merging strategy. [8] considered multiple metrics measuring temporal similarity, string similarity, and entity similarity. [7] utilized external entities information in a KG. KG-based approaches aim to directly link word-based event mentions to Wikipedia [9]. While heuristic-based approaches can only identify texts of high similarity, KG-based approaches fail to capture events not included in the KG.

Representation of Event Temporal Relations Freksa’s cognitive perspective on time and temporal reasoning, as elucidated in his work [21], offers a simplified version of Allen’s interval-based temporal relations, comprising 13 distinct types [22]. Unlike Allen’s approach, which presents a compositionally complete framework for temporal reasoning, Freksa’s semi-interval-based representation acknowledges the uncertainty inherent in event boundaries and measures temporal relations based on the occurrence of events. While Allen’s model may be theoretically robust, it may not always align with the intricacies of narrative studies. In this work, the temporality of events refers to the narrated time of events rather than their actual occurrence. Viewing narratives as a point in time offers a coherent perspective, which recognizes that narratives encapsulate a specific temporal snapshot rather than a comprehensive depiction of events as they unfold in the real world.

3. Dataset Construction

Figure 2 shows the dataset construction pipeline. This includes news acquisition and denoising, headline pairs sampling and validation.

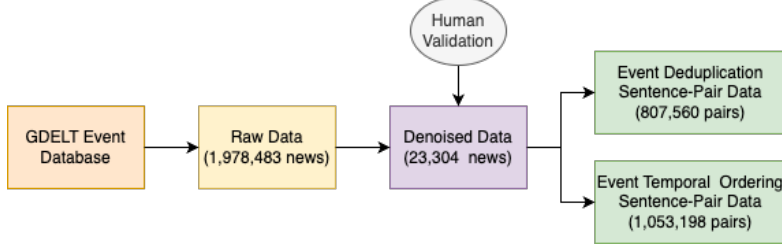


Figure 2: Dataset Construction Pipeline.

News Acquisition News articles are initially collected from the Global Database of Events, Language, and Tone (GDELT) [3], utilizing storm-related keywords such as storm, flood, hurricane, typhoon, and tornado. The search spans from January 1, 2021, to September 1, 2023, with English language restrictions, yielding 1,978,483 raw articles from 149 countries. After identifying and removing 50.28% of duplicate entries, 983,775 unique news articles remain. Our preliminary studies showed that the regular expressions-based GDELT search includes pages where at least one keyword appears in the title, body, image caption, or advertisements. To refine the selection to storm-related news, additional steps are taken, including cosine similarity-based clustering, event type annotation, entity recognition, and temporal clustering.

3.1. Denoising Pipeline

The steps in denoising data are designed to construct a dataset with high precision. This entails ensuring that each cluster uniquely refers to a specific event, and that all headlines within any cluster reference the same event, despite potential narrative variations.

Clustering on headlines To generate representations for event mentions, we employ a pre-trained sentence transformer (all-MiniLM-L6-v2¹). Subsequently, we calculate a distance matrix $d \in \mathbb{R}^{n \times n}$ based on cosine similarity for n sentence pairs. Clusters are established with two criteria: 1) each cluster must comprise a minimum of 50 instances, and 2) cosine similarities between every pair of instances must exceed 0.7. This process results in the retrieval of 786,944 news instances, organized into 264 clusters.

Annotation on Event Type and GPE We used the Text REtrieval Conference Incident Stream (TREC-IS) dataset to train an event type classifier for soft labeling clusters. TREC-IS is a collection of microblog posts about pre-disaster, in-disaster, and post-disaster discussions [3]. It contains gold-standard annotation of 118 events. Clusters are chosen if the majority of headlines are predicted as tropical storms, hurricanes, floods, typhoons, or tornadoes. Additionally, spaCy’s NER (en_core_web_md²) is used to annotate GPEs. Given the anticipated variations in

¹<https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2>

²https://spacy.io/models/en#en_core_web_md

narrative discourse within each cluster, differences in GPEs among headlines belonging to the same cluster are expected. The annotated GPEs play a crucial role in the subsequent merging of clusters.

Temporal Clustering For each cluster, a one-dimensional temporal clustering algorithm (DBSCAN) is employed to eliminate temporal outliers, with temporal information represented by the news publish date. We adopted a temporal granularity of one day, setting $min_samples = 3$ and $\epsilon = 1$, resulting in the removal of 752,429 headlines.

Merging Clusters The remaining 34,515 continuous mentions form 278 clusters representing storm-related disaster instances. These clusters encompass various narratives, including pre-disaster, in-disaster and post-disaster information. The maximum overlapping ratio, denoted as r , between the GPEs in any pair of clusters (C_i and C_j) is computed using Equation 1.

$$r_{ij} = \frac{|C_i \cap C_j|}{\max(|C_i|, |C_j|)} \quad (1)$$

To capture long events such as 2023 Cyclone Freddy,³ we allow a temporal gap of 10 days between clusters. Two clusters are merged if (1) GPEs overlap ratio $r_{ij} \geq 0.5$, and (2) minimum between-cluster temporal distance $d_{ij} = \min_{c \in C_i, c' \in C_j} d(c, c') \leq 10$. As a result, 263 unique clusters are returned. To further examine the clustering quality, one domain expert manually linked each cluster to a Wikidata entity of type “occurrence”⁴ and checked for geospatial and temporal consistency between mentions and the actual event occurrences, concluding 113 unique clusters.

3.2. Dataset Validation

We randomly selected 5 headlines from each cluster, resulting in a total of 565 sampled headlines for assessing cluster coherence and uniqueness. All headlines are shown at once. Cluster coherence was evaluated using a 5-point Likert scale, with two human raters indicating the extent to which media narratives described the same event, ranging from “Strongly Disagree” (1) to “Strongly Agree” (5). The mean Likert response score for coherence was $s_{mean} = 4.03$, Krippendorff’s alpha $\alpha = 0.736$. The uniqueness of events within each cluster was assessed by determining whether each event was distinct within the set. Cohen’s kappa statistic for uniqueness was $\kappa = 1$.

4. Benchmarks

Event deduplication is a binary classification task⁵ and event temporal ordering is a multiclass classification task⁶. We use a bi-encoder (siamese network structure) comparing DistilRoBERTa-base, DistilBERT-base-cased, RoBERTa-base and BERT-base-cased. Consistent with [19], a vector concatenation considering element-wise difference between both embedded vectors,

³The longest recorded tropical cyclone

⁴<https://www.wikidata.org/wiki/Q1190554>

⁵Two labels for event deduplication: *same_event_instance*, *different_event_instance*

⁶Three labels for temporal ordering: *before*, *equal* and *after*

$u \oplus v \oplus |u - v|$, is later fed into a softmax classification layer. We used Adam optimizer with learning rate $2e-5$, and a linear learning rate warm-up over 10% of the training data. Number of training epoch is 5. The dataset is partitioned to include clusters of varying sizes in each split (Table 2), consisting of 9 large clusters ($N_c \geq 500$), 48 medium-sized clusters ($500 > N_c \geq 200$), and 58 small clusters ($N_c < 200$).

Table 2

Number of instances of different splits in the final dataset. Parentheses in #Events indicate the number of events from a large/medium/small cluster.

Split	#Events	#Headlines	#Sampled headline pairs (Event Deduplication)	#Sampled headline pairs (Event Temporal Ordering)
Train	74(7/33/34)	17,449	610,714	872,448
Valid	17(1/7/9)	2,455	73,650	73,650
Test	24(1/8/15)	3,570	107,100	107,100

Within each split, stratified sampling is applied to sample headline pairs, balancing date distribution, label distribution, and event instance distribution in headline pairs. Each model is run with three seeds, and all experiments are run with an NVIDIA RTX A6000 48 GB graphics card. Table 3 shows the benchmark result over 4 seeds (0, 1, 2, 3).

Table 3

Benchmark Results.

	Event Deduplication	Event Temporal Ordering
DistilRoBERTa-base	81.90±3.03	55.07±0.24
DistilBERT-base-cased	89.02±4.98	56.59±0.47
RoBERTa-base	77.33±2.37	56.76±0.67
BERT-base-cased	86.14±6.46	55.55±0.29

The baseline result under the pairwise comparison task formulation reveals two significant implications. Firstly, it demonstrates the feasibility of discriminating event instances using only news headlines, as evidenced by the result in event deduplication. This suggests that the denoising pipeline and event type detection mechanisms are effective in distinguishing how different storm-related events are mentioned in the headlines. Secondly, the result highlights the challenge of identifying the temporal ordering of headlines mentioning storm-related events. This task proves to be much more complex compared to discriminating between individual event instances. The nuanced temporal relationships and contextual dependencies within narratives pose a greater difficulty in determining the chronological sequence of events discussed in news articles.

5. Limitations and Discussions

5.1. Generalizability of the Proposed Denoising Pipeline

One noticeable ad-hoc component in the denoising pipeline that hinders generalizability is the domain-specific event type detector, trained in a supervised learning fashion. We used

an event type detector on a gold-standard human-annotated dataset (TREC-IS) to annotate event types to select news headlines classified as being storm-related. While this approach prioritizes the accuracy of event type annotations, it limits the generalizability to different event types. We highlight the importance of showcasing the practical implications of the relatively under-explored task of event instance discrimination, exemplified by storm-related events. However, advancements in zero-shot learning techniques, such as prompt-based approaches leveraging Large Language Models (LLMs), offer promising avenues to replace the ad-hoc event type detector, potentially enhancing the pipeline’s generalizability across diverse event types. It is crucial to conduct a thorough evaluation of zero-shot event type detectors to ensure the quality of annotations.

5.2. Event Temporal Ordering Beyond News Publication Date

We used news publication dates as features to identify the temporal order of any pair of news headlines. This representation ignores the temporal lag between news publication and the actual occurrence of events. We acknowledge the limitations of this approach, particularly in extracting precise absolute temporal attributes of events. A more accurate method would involve linking event mentions to well-structured data sources such as Knowledge Graphs (KGs) to extract strictly precise temporal attributes. However, the task of event linking remains relatively unexplored, with existing works primarily linking event mentions to community-driven sources like Wikipedia, which may contain inaccurate, or false information [9].

Despite the inherent limitations, leveraging news publication dates for temporal ordering provides an approximate timeline of event developments within media narratives. This approach allows for the construction of a detailed narrative timeline, facilitating the identification of in-disaster news reports and further extraction, normalization, and analysis of temporal expressions.

6. Conclusion

In conclusion, we introduced the *Eventist* dataset, a comprehensive silver-standard dataset that serves as a benchmark for two essential pairwise comparison tasks: event deduplication and event temporal ordering. While we validated the dataset with human raters, it’s important to note that it may contain noise introduced by biases inherent in the denoising pipeline. However, our focus on event instance discrimination underscores the dataset’s significance and opens avenues for further research in refining and enhancing these critical NLP tasks.

Acknowledgments

The authors acknowledge the financial support by the Federal Ministry for Economic Affairs and Energy of Germany in the project CoyPu (project number 01MK21007G).

References

- [1] R. Méndez-Tejeda, J. J. Hernández-Ayala, Links between climate change and hurricanes in the North Atlantic, *PLOS Climate* 2 (2023) e0000186. URL: <https://doi.org/10.1371/journal.pclm.0000186>. doi:10.1371/journal.pclm.0000186.
- [2] D. Xi, N. Lin, A. Gori, Increasing sequential tropical cyclone hazards along the US East and Gulf coasts, *Nature Climate Change* 13 (2023) 258–265. URL: <https://doi.org/10.1038/s41558-023-01595-7>. doi:10.1038/s41558-023-01595-7.
- [3] R. McCreadie, C. Buntain, I. Soboroff, TREC incident streams: Finding actionable information on social media, in: Z. Franco, J. J. González, J. H. Canós (Eds.), *Proceedings of the 16th International Conference on Information Systems for Crisis Response and Management*, València, Spain, May 19-22, 2019, ISCRAM Association, 2019. URL: http://idl.iscrum.org/files/richardmccreadie/2019/1867_RichardMcCreadie_etal2019.pdf.
- [4] R. McCreadie, C. Buntain, Crisisfacts: Building and evaluating crisis timelines, in: *20th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2023)*, Omaha, NE, USA, 2023, pp. 320–339. URL: <https://doi.org/10.59297/JVQZ9405>. doi:10.59297/JVQZ9405.
- [5] G. R. Doddington, A. Mitchell, M. A. Przybocki, L. A. Ramshaw, S. M. Strassel, R. M. Weischedel, The automatic content extraction (ACE) program - Tasks, Data, and Evaluation, in: *Proceedings of the Fourth International Conference on Language Resources and Evaluation, LREC 2004*, May 26-28, 2004, Lisbon, Portugal, European Language Resources Association, 2004. URL: <http://www.lrec-conf.org/proceedings/lrec2004/summaries/5.htm>.
- [6] X. Wang, Z. Wang, X. Han, W. Jiang, R. Han, Z. Liu, J. Li, P. Li, Y. Lin, J. Zhou, MAVEN: A massive general domain event detection dataset, in: B. Webber, T. Cohn, Y. He, Y. Liu (Eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020*, Online, November 16-20, 2020, Association for Computational Linguistics, 2020, pp. 1652–1671. URL: <https://doi.org/10.18653/v1/2020.emnlp-main.129>. doi:10.18653/v1/2020.EMNLP-MAIN.129.
- [7] F. Rollo, L. Po, Crime event localization and deduplication, in: J. Z. Pan, V. A. M. Tamma, C. d’Amato, K. Janowicz, B. Fu, A. Polleres, O. Seneviratne, L. Kagal (Eds.), *The Semantic Web - ISWC 2020 - 19th International Semantic Web Conference*, Athens, Greece, November 2-6, 2020, *Proceedings, Part II*, volume 12507 of *Lecture Notes in Computer Science*, Springer, 2020, pp. 361–377. URL: https://doi.org/10.1007/978-3-030-62466-8_23. doi:10.1007/978-3-030-62466-8_23.
- [8] V. Zavarella, J. Piskorski, C. Ignat, H. Tanev, M. Atkinson, Mastering the media hype: Methods for deduplication of conflict events from news reports, in: A. M. Jorge, R. Campos, A. Jatowt, A. Aizawa (Eds.), *Proceedings of AI4Narratives - Workshop on Artificial Intelligence for Narratives in conjunction with the 29th International Joint Conference on Artificial Intelligence and the 17th Pacific Rim International Conference on Artificial Intelligence (IJCAI 2020)*, Yokohama, Japan, January 7th and 8th, 2021 (online event due to Covid-19 outbreak), volume 2794 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 29–34. URL: <https://ceur-ws.org/Vol-2794/paper6.pdf>.
- [9] X. Yu, W. Yin, N. Gupta, D. Roth, Event linking: Grounding event mentions to wikipedia, in: A. Vlachos, I. Augenstein (Eds.), *Proceedings of the 17th Conference of the European*

- Chapter of the Association for Computational Linguistics, EACL 2023, Dubrovnik, Croatia, May 2-6, 2023, Association for Computational Linguistics, 2023, pp. 2671–2680. URL: <https://doi.org/10.18653/v1/2023.eacl-main.196>. doi:10.18653/v1/2023.eacl-main.196.
- [10] W. Ai, J. Xu, H. Shao, Z. Wang, T. Meng, An entity event deduplication method based on connected subgraph, in: J. Yang, K. Li, W. Tu, Z. Xiao, L. Wang (Eds.), 7th International Conference on Systems and Informatics, ICSAI 2021, Chongqing, China, November 13-15, 2021, IEEE, 2021, pp. 1–6. URL: <https://doi.org/10.1109/ICSAI53574.2021.9664040>. doi:10.1109/ICSAI53574.2021.9664040.
- [11] O. Inel, L. Aroyo, Validation methodology for expert-annotated datasets: Event annotation case study, in: M. Eskevich, G. de Melo, C. Fäth, J. P. McCrae, P. Buitelaar, C. Chiarcos, B. Klimek, M. Dojchinovski (Eds.), 2nd Conference on Language, Data and Knowledge, LDK 2019, May 20-23, 2019, Leipzig, Germany, volume 70 of *OASICS*, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019, pp. 12:1–12:15. URL: <https://doi.org/10.4230/OASICS.LDK.2019.12>. doi:10.4230/OASICS.LDK.2019.12.
- [12] Z. Song, A. Bies, S. M. Strassel, J. Ellis, T. Mitamura, H. T. Dang, Y. Yamakawa, S. Holm, Event nugget and event coreference annotation, in: M. Palmer, E. H. Hovy, T. Mitamura, T. O’Gorman (Eds.), Proceedings of the Fourth Workshop on Events, EVENTS@HLT-NAACL 2016, San Diego, California, USA, June 17, 2016, Association for Computational Linguistics, 2016, pp. 37–45. URL: <https://doi.org/10.18653/v1/W16-1005>. doi:10.18653/v1/W16-1005.
- [13] C. Colruyt, O. D. Clercq, T. Desot, V. Hoste, EventDNA: A dataset for dutch news event extraction as a basis for news diversification, *Lang. Resour. Evaluation* 57 (2023) 189–221. URL: <https://doi.org/10.1007/s10579-022-09623-2>. doi:10.1007/s10579-022-09623-2.
- [14] S. Liu, Y. Li, F. Zhang, T. Yang, X. Zhou, Event detection without triggers, in: J. Burstein, C. Doran, T. Solorio (Eds.), Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers), Association for Computational Linguistics, 2019, pp. 735–744. URL: <https://doi.org/10.18653/v1/n19-1080>. doi:10.18653/v1/n19-1080.
- [15] T. Ling, L. Chen, H. Sheng, Z. Cai, H. Liu, Sentence-level event detection without triggers via prompt learning and machine reading comprehension, *CoRR* abs/2306.14176 (2023). URL: <https://doi.org/10.48550/arXiv.2306.14176>. doi:10.48550/ARXIV.2306.14176. arXiv:2306.14176.
- [16] A. Piper, R. J. So, D. Bamman, Narrative theory for computational narrative understanding, in: M. Moens, X. Huang, L. Specia, S. W. Yih (Eds.), Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021, Association for Computational Linguistics, 2021, pp. 298–311. URL: <https://doi.org/10.18653/v1/2021.emnlp-main.26>. doi:10.18653/v1/2021.emnlp-main.26.
- [17] S. Diaf, J. Döpke, U. Fritsche, I. Rockenbach, Sharks and minnows in a shoal of words: Measuring latent ideological positions based on text mining techniques, *European Journal of Political Economy* 75 (2022) 102179. URL: <https://doi.org/10.1016/j.ejpoleco.2022.102179>. doi:10.1016/j.ejpoleco.2022.102179.
- [18] T. Zhang, A. M. Schoene, S. Ji, S. Ananiadou, Natural language processing applied to

- mental illness detection: A narrative review, *npj Digit. Medicine* 5 (2022). URL: <https://doi.org/10.1038/s41746-022-00589-7>. doi:10.1038/s41746-022-00589-7.
- [19] N. Reimers, I. Gurevych, Sentence-BERT: Sentence embeddings using siamese BERT-networks, in: K. Inui, J. Jiang, V. Ng, X. Wan (Eds.), *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, Association for Computational Linguistics, 2019, pp. 3980–3990. URL: <https://doi.org/10.18653/v1/D19-1410>. doi:10.18653/v1/D19-1410.
- [20] S. Nishikawa, R. Ri, I. Yamada, Y. Tsuruoka, I. Echizen, EASE: Entity-aware contrastive learning of sentence embedding, in: M. Carpuat, M. de Marneffe, I. V. M. Ruiz (Eds.), *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL 2022, Seattle, WA, United States, July 10-15, 2022*, Association for Computational Linguistics, 2022, pp. 3870–3885. URL: <https://doi.org/10.18653/v1/2022.naacl-main.284>. doi:10.18653/v1/2022.naacl-main.284.
- [21] C. Freksa, Temporal reasoning based on semi-intervals, *Artif. Intell.* 54 (1992) 199–227. URL: [https://doi.org/10.1016/0004-3702\(92\)90090-K](https://doi.org/10.1016/0004-3702(92)90090-K). doi:10.1016/0004-3702(92)90090-K.
- [22] J. F. Allen, Maintaining knowledge about temporal intervals, *Commun. ACM* 26 (1983) 832–843. URL: <https://doi.org/10.1145/182.358434>. doi:10.1145/182.358434.