# CStory: A Chinese Large-scale News Storyline Dataset

Kaijie Shi
1196479790@qq.com
Tsinghua University
Beijing, China

Xiaozhi Wang
wangxz20@mails.tsinghua.edu.cn
Tsinghua University
Beijing, China

Jifan Yu
yujf21@mails.tsinghua.edu.cn
Tsinghua University
Beijing, China

Yu Zhou
bryanzhjou008@ucla.edu
Department of Computer Science,
University of California, Los Angeles
Los Angeles, USA

Lei Hou*
houlei@tsinghua.edu.cn
Department of Computer Science
and Technology, BNRist, Tsinghua
University
Beijing, China

Juanzi Li
lijuanzi@tsinghua.edu.cn
Department of Computer Science
and Technology, BNRist, Tsinghua
University
Beijing, China

Jingtong Wu
wujingtong@huawei.com
Beijing Huawei Digital Technologies
Co., Ltd.
Beijing, China

Dingyu Yong
yongdingyu@huawei.com
Huawei Device Co., Ltd.
Beijing, China

Jinghui Xiao
xiaojinghui4@huawei.com
Huawei Noah's Ark Lab
Beijing, China

Qun Liu
qun.liu@huawei.com
Huawei Noah's Ark Lab
Beijing, China

## ABSTRACT

In today's massive news streams, storylines can help us discover related event pairs and understand the evolution of hot events. Hence many efforts have been devoted to automatically constructing news storylines. However, the development of these methods is strongly limited by the size and quality of existing storyline datasets since news storylines are expensive to annotate as they contain a myriad of unlabeled relationships growing quadratically with the number of news events. Working around these difficulties, we propose a sophisticated pre-processing method to filter candidate news pairs by entity co-occurrence and semantic similarity. With the filter reducing annotation overhead, we construct CStory, a large-scale Chinese news storyline dataset, which contains 11, 978 news articles, 112, 549 manually labeled storyline relation pairs, and 49, 832 evidence sentences for annotation judgment. We conduct extensive experiments on CStory using various algorithms and find that constructing news storylines is challenging even for pre-trained language models. Empirical analysis shows that the sample unbalance issue significantly influences model performance, which shall be the focus of future works. Our dataset is now publicly available at https://github.com/THU-KEG/CStory.

## CCS CONCEPTS

• **Information systems** → **Data stream mining**; **Web mining**; **Data extraction and integration**;

## KEYWORDS

Storyline Datasets, Event Evolution, Storyline Relation, Topic Detection and Tracking, Imbalanced Dataset

**ACM Reference Format:**
Kaijie Shi, Xiaozhi Wang, Jifan Yu, Yu Zhou, Lei Hou, Juanzi Li, Jingtong Wu, Dingyu Yong, Jinghui Xiao, and Qun Liu. 2022. CStory: A Chinese Large-scale News Storyline Dataset. In *Proceedings of the 31st ACM Int'l Conference on Information and Knowledge Management (CIKM '22), Oct. 17–21, 2022, Atlanta, GA, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3511808.3557573

## 1 INTRODUCTION

Curiosity in human nature propels us to learn about current events, infer their causes from past events, and use them to predict future events [19]. We naturally use a storyline, a tool that reveals the evolution of hot events, to help us understand event causes and consequences. Such causal knowledge can be further used in downstream tasks such as script event prediction [8, 12], causal inference [20], and event evolutionary graph construction [28]. The relationship between news articles in a storyline, i.e., news storyline relation, refers to two news articles that have a causal, dependency [17], or associative [1] relationship between their text

*Corresponding author.

*CIKM '22, October 17–21, 2022, Atlanta, GA, USA*

ACM ISBN 978-1-4503-9236-5/22/10.
https://doi.org/10.1145/3511808.3557573

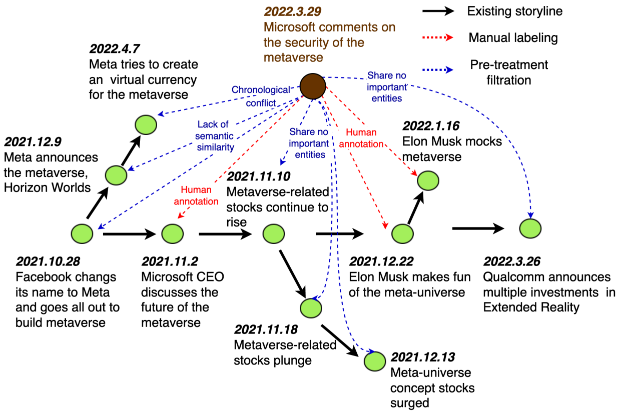

2022.3.29
Microsoft comments on the security of the metaverse
Existing storyline
Manual labeling
Pre-treatment filtration
2022.4.7
Meta tries to create an virtual currency for the metaverse
Chronological conflict
Share no important entities
Lack of semantic similarity
2021.12.9
Meta announces the metaverse, Horizon Worlds
Share no important entities
Human annotation
2022.1.16
Elon Musk mocks metaverse
Human annotation
2021.11.10
Metaverse-related stocks continue to rise
2021.10.28
Facebook changs its name to Meta and goes all out to build metaverse
2021.11.2
Microsoft CEO discusses the future of the metaverse
2021.12.22
Elon Musk makes fun of the meta-universe
2022.3.26
Qualcomm announces multiple investments in Extended Reality
2021.11.18
Metaverse-related stocks plunge
2021.12.13
Meta-universe concept stocks surged

**Figure 1: A storyline for events regarding metaverse.**

semantics. As shown in the fishbone diagram in Figure 1, storyline generation models can help to discover news pairs with dependencies and correlations [25], construct the rich structure between news articles [6, 17], and build a directed diagram of hot event development from massive news streams [14, 23].

Although news storyline generation has attracted a lot of attention, the quality of datasets in this field is unsatisfactory. **Fristly, the size of existing datasets is relatively small since it is expensive to annotate large numbers of relational news data**[6, 10, 31]. As shown in Figure 1, the number of relations to be annotated increases quadratically with the number of events because the new data need to be compared with all the existing news articles in the storyline. Some other news storyline datasets have a relatively large size[11, 22, 29, 30]. However, they are only labeled with storyline's structure similarity[22], time duration similarity[30] , keyword similarity[29] etc. Such datasets not only fail to precisely evaluate the quality of storylines, but also fail to evaluate the deep semantic relation between news articles. **Furthermore, existing storyline datasets lack evidential sentences for judging storyline relations.** Evidential sentences refer to the sentences in the original text that explicitly support the storyline relationship between the two news articles. Evidential sentences not only help storyline generation models to capture important semantic information, but also greatly improve the efficiency of annotation quality checking.

In order to address the shortcomings of existing datasets, we develop CStory, a large-scale news storyline dataset, which contains 11, 978 Chinese news articles, 71, 742, 231 news storyline relation pairs, and 49, 832 evidential sentences for judging storyline relation. To reduce the annotation overhead, we introduce a pre-processing step before manual labeling. Taking Figure 1 as an example, we use entity co-occurrence constraints to exclude news articles related to stocks and text similarity constraints to exclude news articles related to Meta Platforms, eventually only three remaining news pairs require to be manually labeled. CStory has the following advantages over previous works:

- **Large-scale and high-coverage**: Compared with the previous news storyline relation datasets [6, 10, 17, 27, 31], the number of news articles in our dataset has increased by order of magnitude, and the number of news relations has increased two orders of magnitude. We have successfully achieved hundreds of millions of judgments on news relations with only 0.01% of manually labeled data. The pre-processing method guarantees a positive sample recall close to 100% while sieving out 99.9% of the negative samples, making human annotation feasible.
- **Broad and general topics**: Our dataset resembles real-world news streams without selective filtering and covers topics including politics, military, economy, sports, culture, etc., making it a topic-complete data repository.
- **Evidential sentences for judging storyline relation**: Given the length and complexity of news articles, it is difficult to find the evidential sentences that directly support our judgment of the storyline relation. To the best of our knowledge, CStory is the first dataset to annotate evidential sentences of news storyline relations, which are potentially valuable for tasks such as semantic understanding and news summaries.

## 2 DATASET CREATION

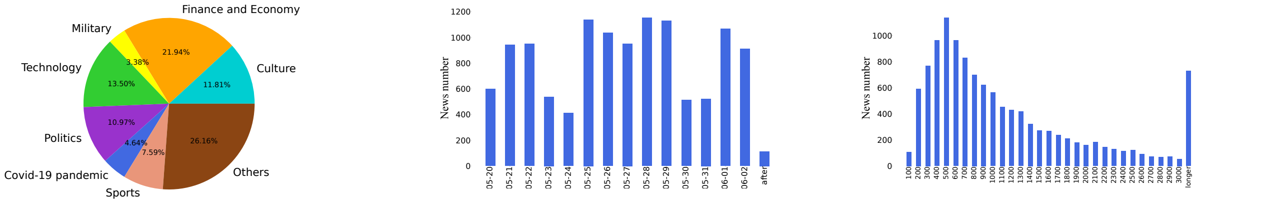
In the section, we describe the technical details of building CStory.

### 2.1 News Event Detection

News event detection is the basis for building a storyline dataset. It removes duplicate news articles and also locates the most popular ones. To ensure the quality of news articles, our news articles are obtained from mainstream media such as People's Daily, Global Times, and Xinhua News. After collecting the news articles, we concatenate the headline and body of the news article into a long text and then input the first 512 tokens of the text into the multilingual BERT model [21] to generate a 768-dimensional embedding vector. We use this vector as the semantic vector of the news article. Then we cluster news articles with the SinglePass [18] algorithm. We aggregated nearly 100, 000 news articles into 31, 352 events clusters. To control the size of CStory, we selected the top 11, 978 event clusters sorted by the number of news articles.

### 2.2 Pre-processing of News Pairs

A significant limitation to the size of the previous dataset is that the number of news pairs grows quadratically with the number of news articles; e.g., 10, 000 news articles require 50 million news pairs to be annotated. To solve the difficulties mentioned above, we designed a pre-processing method to select all candidate news pairs by constraining the co-occurring entities' number and semantic similarity between news articles. Based on our observation, associated news articles always have key entities in common, which serve as connecting bridges. Therefore, we require the two news articles to share at least one key entity in a news pair. Key entities can be people, organizations, institutions, etc. We use a Chinese NLP toolkit Jieba [24] to extract such entities in the news text. Furthermore, based on the assumption that associated news articles are semantically coherent and relevant, we constrain the text similarity of two news articles in a news pair to be higher than a predefined threshold, where text similarity is calculated using the cosine similarity of semantic vectors introduced in Section 2.1.



(a) Topic Distribution of CStory.

(b) News Release Time Distribution.

(c) News Length Distribution.

**Figure 2: CStory experimental results.**

Among the total 71, 742, 231 storyline relationships, 112, 549 relationships were manually labeled, and the other 71, 629, 682 relationships were automatically determined as negative cases by pre-processing. We obtained a total of 35, 204 positive cases out of the 112, 549 manually labeled storyline relationships. We randomly sampled 20 news articles from CStory to form a recall test set, then we manually found all news articles related to the recall test set. According to our experiment, our pre-processing method identified 250 of the 255 news storyline relations, achieving a 98% recall rate.

## 2.3 Quality Control of Annotation

We have two annotation tasks, storyline relationship annotation and key sentence annotation. Storyline relationship annotation needs annotators to determine whether two news articles have storyline relationships, and key sentence annotation needs annotators to find key sentences that support storyline relationships and mark their position in the news articles.

In the storyline relationship annotation process, each candidate news pair is assigned to three independent annotators. In cases of disagreement among their responses, a quality inspector is assigned to review and arbitrate the final result carefully. To further ensure high annotation quality, we periodically spot-check 5% of data to make sure annotation quality meets our criteria. In the key sentence annotation process, since there could be more than one key sentence, we only constrain the annotation accuracy to be more than 90%. Besides, we periodically spot-check 5% of data to ensure annotation quality.

## 2.4 Availability

Our dataset is now public available at https://github.com/THU-KEG/CStory. Besides, we also provide two toolkits as follows:

**CStory baseline tookit** contains all the code for the baseline models in this paper. We also provide a script that can run all experiments for researchers to reproduce all the results.

**CStory builder toolkit** is a dataset customization tool that controls keywords, positive and negative sample ratios and size of the dataset. Users can freely customize the dataset with positive and negative sample proportions, topics, and other characteristics.

| **Literature** | **#News Articles** | **#Relations** |
|---|---:|---:|
| Zhou et al. [31] | 228 | $10^4$ |
| Huang et al. [10] | 827 | $10^5$ |
| Nallapati et al. [17] | 1, 468 | $10^5$ |
| Yang et al. [27] | 782 | $10^5$ |
| Feng and Allan [6] | 280 | $10^4$ |
| Caselli and Vossen [3] | 984 | $10^5$ |
| **CStory** | **11,978** | $\mathbf{10^8}$ |

**Table 1: The scale of CStory compared with existing datasets.**

## 2.5 Data Analysis

In this section, we introduce statistics CStory to help readers better understand our dataset. As shown in Figure 2(a), CStory has a broad and balanced distribution of topics, including Technology, Politics, Finance and Economy etc. As shown in Figure 2(b), most of the news articles in CStory are published within half a month, and the daily article amount is maintained roughly between 600 and 1000. In Figure 2(c), we observe that the distribution of news articles over text length follows a Poisson distribution, with the peak 1000 occurring at the text length of 500 words. This distribution matches our expectations as our data is supposed to be a good representation of real-world news streams. As shown in Table 1 , compared with previous datasets, CStory is tens of times larger in terms of news article count and hundreds of times larger in terms of news relation count. Among the 71, 742, 23 news relations, the most challenging ones to classify are the 112, 549 manually labeled relation pairs. The similarity scores of these relation pairs are high, making it difficult for the model to distinguish their labels.

## 3 EXPERIMENT

Our comprehensive dataset can support numerous downstream tasks. We take two tasks, storyline relation classification and storyline generation, as examples. **Storyline relation classification** aims to identify storyline relations between given news pairs [6, 7, 14, 17, 31]. It works on the incremental scenario, where news articles are incrementally added to existing storylines. Moreover, **storyline generation** task requires to generate complete storylines for a number of news articles from scratch [14, 16, 17, 30]. It simulates constructing storylines from scratch.

## 3.1 Storyline Relation Classification

Storyline relation classification is a textual binary classification task. The input of the storyline relation classification is two news articles, and the output is a binary value indicating whether these two articles have a storyline relation.

**Data preparation.** The data we use for this part has been filtered through the pre-processing method in Section 2.2. To prevent information leakage, the data in our training set, validation set and test set are taken from different time periods with certain intervals. To investigate the effect of different positive and negative sample proportions, we designed two test sets: a balanced test set and an unbalanced test set. The specific information of each data set is shown in Table 2 where P/N means the ratio of positive pair numbers divided by negative pair numbers.

| Dataset | Number | #Positive | #Negative | P/N |
|---|---|---|---|---|
| Trainset | 30222 | 12488 | 17734 | 1:1.420 |
| Devset | 2983. | 1125 | 1858 | 1:1.652 |
| Balanced test set | 4405 | 1563 | 2842 | 1:1.818 |
| Unbalanced test set | 5565 | 751 | 4814 | 1:6.410 |

**Table 2: Storyline relation classification datasets.**

**Baselines.** We provide the results of 6 storyline relation classification methods which include: a random model, TF-IDF cosine similarity [26], Jaccard index [9], BERT [5], RoBERTa [4] and similarity combinations of location, participant and text (LPT) [16].

We use grid search to set the optimal parameters of TF-IDF cosine similarity, Jaccard index, and LPT. For BERT and RoBERTa, a special note of caution is that we only input the first 256 characters of a news article into the pre-trained models.

**Experiment results.** The experimental results of storyline relation classification are shown in Table 3. All models achieve a higher F1 score on the balanced dataset compared to the unbalanced dataset. Among the unsupervised algorithms, Jaccard index [9] has the best results on both test sets, which is due to the fact that storyline relationships and important entities are closely linked. Jaccard index can effectively measure the similarity between entities in two news articles.

The two pre-trained language models, BERT [2] and RoBERTa [15], have little difference in performance overall. On the balanced dataset, they both significantly outperform unsupervised models. However, their performance dropped below the unsupervised models on the unbalanced dataset. This experiment shows that the data unbalance issue has a more significant influence on the sophisticated large models. An interesting phenomenon is that the recall and F1 of BERT are greater than RoBERTa in both test sets, which may be due to BERT having a next-sentence-prediction pre-training task, which has a positive effect on news storyline relation classification. Furthermore, we explore the influence of our specially labeled evidential sentences under the setting of storyline classification. We add two special tokens, "keystart" and "keyend", to each side of the evidential sentences in the training data set to train RoBERTa. We observe that such information can significantly benefit the model performance and lift the F1-score to 0.948 (24.57% boost), indicating our evidential sentences of news articles are high-quality enough for inspiring relevant methods.

| Model | Dataset | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|---|
| random | Balanced | 0.499 | 0.354 | 0.500 | 0.414 |
| TF-IDF | Balanced | 0.535 | 0.579 | 0.593 | 0.584 |
| Jaccard | Balanced | 0.588 | 0.455 | 0.809 | 0.582 |
| LPT | Balanced | 0.573 | 0.437 | 0.703 | 0.539 |
| BERT | Balanced | 0.818 | 0.708 | **0.831** | **0.765** |
| RoBERTa | Balanced | **0.824** | **0.735** | 0.790 | 0.761 |
| random | Unbalanced | 0.501 | 0.136 | 0.504 | 0.214 |
| TF-IDF | Unbalanced | 0.725 | 0.226 | 0.426 | 0.295 |
| Jaccard | Unbalanced | **0.778** | **0.285** | 0.434 | **0.344** |
| LPT | Unbalanced | 0.671 | 0.210 | 0.521 | 0.299 |
| BERT | Unbalanced | 0.623 | 0.217 | **0.688** | 0.330 |
| RoBERTa | Unbalanced | 0.652 | 0.215 | 0.597 | 0.317 |

**Table 3: The results of Storyline relation classification.**

| Model | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|
| Timeline | 0.884 | 0.163 | 0.115 | 0.135 |
| Fully connected graph (TFIDF) | 0.894 | 0.299 | 0.278 | 0.288 |
| Fully connected graph (RoBERTa) | 0.909 | 0.400 | **0.874** | **0.549** |
| Maximum Spanning Tree (TFIDF) | 0.887 | 0.225 | 0.152 | 0.181 |
| Maximum Spanning Tree (RoBERTa) | **0.923** | **0.511** | 0.205 | 0.293 |

**Table 4: The results of Storyline generation.**

## 3.2 Storyline Generation

Storyline generation is a graph generation task. Its input is news articles, and the output is an adjacency matrix whose values represent the storyline relationships between all news in the news set.

**Dataset preparation.** We provide a test set that contains 20 storylines and 400 news articles, where each storyline contains roughly 20 news articles.

**Baselines.** For storyline generation, we provide the results of 3 methods, i.e., timeline [11], fully connected graph [27], and maximum spanning tree [13]. For fully connected graphs and maximum spanning trees, we use two models, TF-IDF [26] and RoBERTa [4], to calculate news similarities.

**Experiment results.** The experimental result of storyline generation is shown in Table 4. It can be seen that the storyline generation algorithms based on RoBERTa encoding significantly outperform the storyline generation algorithm based on TF-IDF encoding. In addition, the fully connected graph methods have a significant advantage on *Recall*, while the maximum spanning tree methods have a huge advantage on *Precision*.

## 4 IMPACT

In this section, we discuss the expected impact of CStory and its potential use cases.

**Impact on storyline evaluation standardization.** Currently, there is only a small number of publicly accessible annotated news event corpora to support storyline studies. However, they are small in size and not widely used among researchers. We hope that CStory, with its large scale, wide range of topics, and fully open access, will be widely accepted by researchers as a benchmark dataset for storyline construction.

**Potential use cases.** Researchers can use CStory for news topic detection and tracking, storyline generation, semantic news mining, and news event evolution analysis. It will also be useful in the industry for engineers who want to use automatic methods to track and analyze trending news events among large news streams. A storyline discovery demo powered by CStory is shown at http://166.111.68.66:15001/list.html.

## 5 CONCLUSION AND FUTURE WORK

We propose CStory, a large-scale Chinese news storyline dataset containing 11, 978 news articles, 71, 742, 231 storyline relation pairs, and 49, 832 evidenctial sentences for annotation judgement. CStory can provide sufficient training data for large-scale neural network models. We also conduct storyline relation classification and storyline generation experiments as example tasks to show the potential applications of CStory.

Promising future work directions include: (1) Study the effect of different positive and negative sample ratios on model training; (2) Conduct news evolution analysis, including news popularity development, evolution cycle analysis, evolution pattern analysis, etc.; (3) Employ more advanced backbone models.

## 6 ACKNOWLEDGEMENT

This work is supported by the National Key Research and Development Program of China (2020AAA0106501), grants from Huawei Inc. and a grant from the Institute for Guo Qiang, Tsinghua University (2019GQB0003).

## REFERENCES

[1] Jeffery Ansah, Lin Liu, Wei Kang, Selasie Kwashie, Jixue Li, and Jiuyong Li. 2019. A graph is worth a thousand words: Telling event stories using timeline summarization graphs. In *The World Wide Web Conference*. 2565–2571.

[2] Mieke Bal and Christine Van Boheemen. 2009. *Narratology: Introduction to the theory of narrative*. University of Toronto Press.

[3] Tommaso Caselli and Piek Vossen. 2016. The storyline annotation and representation scheme (star): A proposal. In *Proceedings of the 2nd Workshop on Computing News Storylines (CNS 2016)*. 67–72.

[4] Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, and Ziqing Yang. 2021. Pre-training with whole word masking for chinese bert. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 29 (2021), 3504–3514.

[5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[6] Ao Feng and James Allan. 2007. Finding and linking incidents in news. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*. 821–830.

[7] Ao Feng and James Allan. 2009. Incident threading for news passages. In *Proceedings of the 18th ACM conference on Information and knowledge management*. 1307–1316.

[8] Mark Granroth-Wilding and Stephen Clark. 2016. What happens next? event prediction using a compositional neural network model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 30.

[9] Lieve Hamers et al. 1989. Similarity measures in scientometric research: The Jaccard index versus Salton's cosine formula. *Information Processing and Management* 25, 3 (1989), 315–18.

[10] Dongping Huang, Shuyu Hu, Yi Cai, and Huaqing Min. 2014. Discovering event evolution graphs based on news articles relationships. In *2014 IEEE 11th International Conference on e-Business Engineering*. IEEE, 246–251.

[11] Lifu Huang et al. 2013. Optimized event storyline generation based on mixture-event-aspect model. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. 726–735.

[12] Zhongyang Li, Xiao Ding, and Ting Liu. 2018. Constructing narrative event evolutionary graph for script event prediction. *arXiv preprint arXiv:1805.05081* (2018).

[13] Fu-ren Lin and Chia-Hao Liang. 2008. Storyline-based summarization for news topic retrospection. *Decision Support Systems* 45, 3 (2008), 473–490.

[14] Bang Liu, Di Niu, Kunfeng Lai, Linglong Kong, and Yu Xu. 2017. Growing story forest online from massive breaking news. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. 777–785.

[15] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692* (2019).

[16] Zhongyu Lu, Weiren Yu, Richong Zhang, Jianxin Li, and Hua Wei. 2015. Discovering event evolution chain in microblog. In *2015 IEEE 17th International Conference on High Performance Computing and Communications, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security, and 2015 IEEE 12th International Conference on Embedded Software and Systems*. IEEE, 635–640.

[17] Ramesh Nallapati, Ao Feng, Fuchun Peng, and James Allan. 2004. Event threading within news topics. In *Proceedings of the thirteenth ACM international conference on Information and knowledge management*. 446–453.

[18] Ron Papka, James Allan, et al. 1998. On-line new event detection using single pass clustering. *University of Massachusetts, Amherst* 10, 290941.290954 (1998).

[19] Robert R Provine. 2012. Curious behavior. In *Curious Behavior*. Harvard University Press.

[20] Kira Radinsky, Sagie Davidovich, and Shaul Markovitch. 2012. Learning causality for news events prediction. In *Proceedings of the 21st international conference on World Wide Web*. 909–918.

[21] Nils Reimers and Iryna Gurevych. 2020. Making monolingual sentence embeddings multilingual using knowledge distillation. *arXiv preprint arXiv:2004.09813* (2020).

[22] Yeon Seonwoo, Alice Oh, and Sungjoon Park. 2018. Hierarchical dirichlet gaussian marked hawkes process for narrative reconstruction in continuous time domain. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 3316–3325.

[23] Dafna Shahaf, Jaewon Yang, Caroline Suen, Jeff Jacobs, Heidi Wang, and Jure Leskovec. 2013. Information cartography: creating zoomable, large-scale maps of information. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. 1097–1105.

[24] J Sun. 2012. Jieba chinese word segmentation tool.

[25] Piek Vossen, Tommaso Caselli, and Yiota Kontzopoulou. 2015. Storylines for structuring massive streams of news. In *Proceedings of the First Workshop on Computing News Storylines*. 40–49.

[26] Chunzi Wu, Bin Wu, and Bai Wang. 2016. Event evolution model based on random walk model with hot topic extraction. In *International Conference on Advanced Data Mining and Applications*. Springer, 591–603.

[27] Christopher C Yang, Xiaodong Shi, and Chih-Ping Wei. 2009. Discovering event evolution graphs from news corpora. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 39, 4 (2009), 850–863.

[28] Sendong Zhao, Quan Wang, Sean Massung, Bing Qin, Ting Liu, Bin Wang, and ChengXiang Zhai. 2017. Constructing and embedding abstract event causality networks from text snippets. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. 335–344.

[29] Deyu Zhou, Haiyang Xu, Xin-Yu Dai, and Yulan He. 2016. Unsupervised Storyline Extraction from News Articles.. In *IJCAI*. 3014–3021.

[30] Deyu Zhou, Haiyang Xu, and Yulan He. 2015. An unsupervised bayesian modelling approach for storyline detection on news articles. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. 1943–1948.

[31] Wubai Zhou, Chao Shen, Tao Li, Shu-Ching Chen, and Ning Xie. 2014. Generating textual storyline to improve situation awareness in disaster management. In *Proceedings of the 2014 IEEE 15th International Conference on Information Reuse and Integration (IEEE IRI 2014)*. IEEE, 585–592.