



# Survey on Event Timeline Generation

Fahma Bakkar K A<sup>1</sup>, Rafeeqe P C<sup>2</sup>

PG Student, Department of Computer Science & Engineering., Government Engineering College,  
Palakkad, Kerala, India<sup>1</sup>

Head of the Department, Department of Computer Science and Engineering, Government Engineering College,  
Palakkad, Kerala, India<sup>2</sup>

**Abstract:** Event is an important thing that happened or took place at a certain point of time. Important part for the event planning is having an event timeline. This event timeline is a guiding company through tout in the planning process. Lack of timeline leads to missing of important deadlines. This paper focused on event timeline generation in different domains like medicine, education, law enforcement etc. The methods used for extracting events are syntactic, algorithm based and summarization technique. This paper thus discusses respective motivations of these approaches and their advantages and limitations.

**Keywords:** Co-reference Resolution, Entity Extraction, Event, Natural Language Processing

## I. INTRODUCTION

Planning events primarily depends on time. Management of events gets its finishing while it is completed within in time. Connecting with time things are easy to be recollected. Time is the base of proper planning management and evaluation. Making time lines will beneficiary in all these scenarios. Time-line generation for History, Today's work planning and tomorrow's event all are important, varies in different domains like medicine, education and law enforcement. When given a text, a timeline provides a concise summary for the article in the variety of a temporal index that indicates the written record extent of the article, as well the number and distribution of events across this extent. E-learning enables the advantage of learners to do "any time, any place, any pace" learning. For that timeliness like short notes gets helped [1].

Automatically generating timelines for events invite competitions in various areas differently. To be specific [2], consider creating a timeline for the American actor Robert Downey Jr. There Hundreds of possible candidate events are there and it is infeasible to display all of them. Robert Downey Jr. is best known for his starring roles in the movie Iron Man and Avengers, but even for a single movie there are dozens of related events to display (production, release dates, opening, and award ceremonies). And in most historical periods covered in history articles (e.g., Antiquity or the Middle Ages), such cues are not commonly available, as no news articles from these eras exist. While using social media data as input summarize the stream of tweets, because tweets information will be in different forms. This information should be gathered and segregation into related groups. And is refined using mining techniques. The inherent nature of historical events along with its narration poses some specific challenges [3] from NLP viewpoint. They comprise temporal co-reference, Entity co-reference resolution, Event co-reference resolution, Normalization of entity names (Title resolution, Multiple names to same person, Location standardization), Location-Actor ambiguity etc.

Timelines are often utilized in education to help students and researchers to understand the order or chronology of historical events and trends for a subject. When showing a ascending order sequence of events along a drawn line that enables a viewer to understand temporal relationships quickly, and the simultaneity or overlap of spans and events. Timelines play an important role in understanding a historical phenomenon, as they convey a sense of change over time. Social movements and wars [1] are often shown as timelines. Also used in the natural world, sciences and for subjects such as astronomy, biology, and geology. Project management can also be done using timelines, help team members to know what milestones need to be achieved and under what time the project want to complete. Freelancers maintain a timeline before they contact their clients.

Timeline generation precede with syntactic [4], Algorithm Based and summarization techniques [5]. Summarization techniques like multi-document summarization, single-document summarisation, Temporal Summarization, novel



continuous summarization are used. Sequence of time stamped summaries may produced by Summarization system. Without reading through the entire tweet stream, learn major news/discussion.

This paper is organized as follows: Section II gives a formal definition of the event time-line generation and an example. Section III discusses various future research directions of intention detection. Section IV gives brief concluding comments.

## II. EVENT TIMELINE GENERATION

Event is an important thing that happened or took place at a certain point of time [3]. This time reference is important for generating time lines. Therefore extracting events and corresponding time period is the challenging task. Extracting information from text includes syntactic, algorithm based and summarization techniques. Timeline generation spreads in different domains. Input changes in accordance with the domain. If timeline is generating for history of a person, input will be Wikipedia profile. If it is news progress, news articles will be the inputs. Fig. 1. shows the basic idea of Event Timeline Generation system from History textbooks.

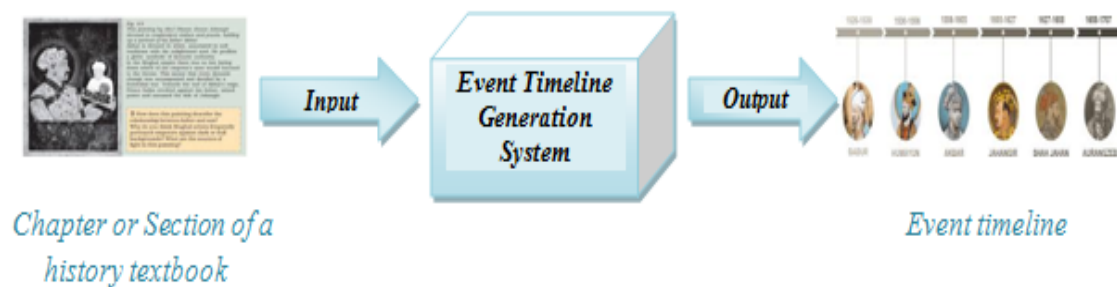


Fig. 1. Basic Idea of Event Timeline Generation System.

Not only in the case of history, most of the source for extracting events and corresponding time period will be Wikipedia. While creating an event time-line for history of popular social media platforms. Here the events are taken as names of social media. Wikipedia data about social media, names of the social media sites can be extracted.

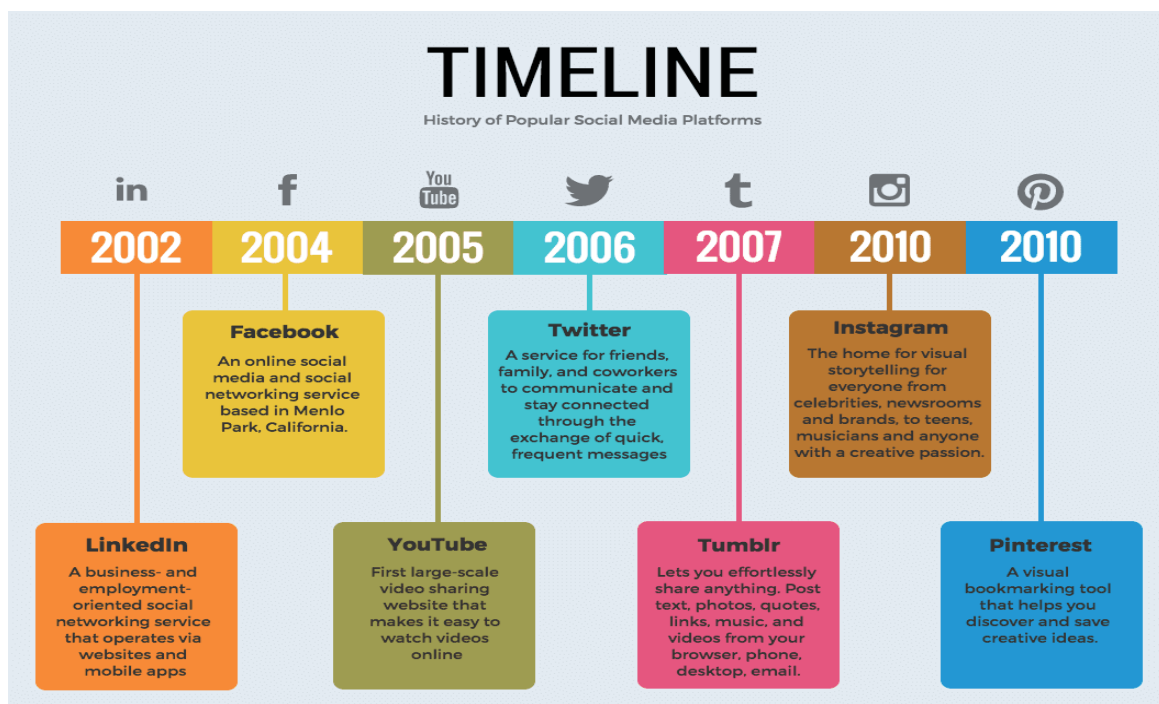


Fig. 2. Event Timelines of History of popular social media platforms [16].



Once the system identifies the events and time span visualization is another major task. According with the nature of the subject visualization techniques also want to be changed. Then only the whole advantages of making timelines will get to the user.

#### A. *Algorithm Based Approaches*

Harsimran Bed et. al. [3], the authors proposes an automatic event time line generation system. This system is considered as an algorithm. The input to the system is chapter or section of a history textbook. Event is an important thing that happened or took place at a certain point of time [1]. They consider events as verb. Verbs like die, consider, defeat are absolute physical action verbs and psychological actions respectively. For getting Gold standard data set annotated schema is used. Event Timeline Generation system have three main steps. First step is Entity extraction, Time-Expression Extraction, Timeline Generation. First, extract the named entities in the text using Stanford CoreNLP. The PERSON and ORGANIZATION type named entities form the set of actors. The LOCATION entities give us spatial information about the events. Then we resolve the co-references of these entities. In the second step uses SUTime temporal expression tagger to extract the temporal expressions from the sentences having the mentions of actors or locations.

System [3] considers only those sentences for creating a time line which contain at least one temporal expression. They name these sentences as time line sentences. Also provide facility to create an actor specific time-line generation. Given an actor, filter those time-line sentences that mention the actor or have co-reference to the actor. In addition to named entities, identify relations mentioned in a sentence using OpenIE component of Stanford CoreNLP. They select the relation which has mentions of the maximum number of named entities as the title of the sentence.

Omar Alonso et. al. [6] suggests a technique that uses social information as relevance surrogates to generate an informative timeline. A core component is a variation of pseudo-relevance feedback automatically generated using social data without external evidence. Finally, describes the implementation of such technique after that presenting evaluation result using a real-world data set. Take advantage of all these behavioural data in conjunction with relevant tweets that contain links that point to web pages with good titles from trusted domains as the basis for constructing a timeline. To the presents, an algorithm based on pseudo-relevance feedback demonstrating this is possible. Proposed method in this paper, social pseudo relevance feedback (sprf), combines user feed- back with query expansion using the contextual vectors presented earlier. The contextual vectors are derived from user-generated content (i.e., tweets) and can think of those n-grams as explicit terms selected by users as votes in aggregate.

Similar to pseudo-relevance feedback [6], ranking those document links extracted from tweets by counters (i.e., behavioural data) indicates that the top-k links are relevant to a given hash tag on a specific time stamp. For query expansion, use the contextual vector to re rank the links based on how similar the links titles are to the terms in the contextual vector. Because all these links have been tagged by users with the relevant and frequent hash tag. The hypothesis is that they belong to the same topic. Conducted an on-line evaluation that shows that technique produce smaller size time lines with much more relevant content. A data analysis reveals that sprf performs well when the hash tag describes an event that covers multiple days, like political or world events. Sports hash tags have a more specific temporal pattern, mostly due to specific games, so sprf is comparable to other methods. Sentences which contain dates, this leads to the wrong selection of events.

#### B. *Syntactic Based Approaches*

Sandro Bauer et. al. [4], a generic approach to content selection for creating timelines from individual history articles. For which no external information about the same topic is available. The presence of a large corpus of news articles requires the contrast to the current scenario is into existing works on timeline generation. To identify salient events in a given history articles exploit lexical cues about the articles subject area, as well as time expressions that are syntactically attached to an event word. The problem is that of finding an optimal sequence of events (of a given maximum length) in a given input article. Follow the literature on event extraction and use TimeML events [7].

Most TimeML events are verbs, but some are nominalizations such as invasion or other event-like words such as war. The use of TimeML events, aside from the practical advantage that commonly-available event extraction algorithms exist. This allows us to evaluate content selection at the event rather than at the sentence level. Assume that there are both local and global factors that determine which events should be contained in the timeline. Local factors reflect how important an event is in its own right. Global factors represent intuitions about which combinations of events should or should not be selected. The approach, which is unsupervised, takes into account the factors described in what follows. The system also outperforms the improved supervised baseline by a large margin.



The fact that a completely unsupervised system performs best is encouraging, as training data for this task is very expensive to obtain. Results suggest that it might be worth investigating other types of prior knowledge about the semantics of an input text in further research. The crucial advantage of such generic methods is that no texts on exactly the same topic are needed, which is a requirement with texts about niche topics.

C. *Summarization Based Approach*

Giang Tran et. al. [5] is all about Automatic timeline summarization (TLS). TLS generates precise, dated overviews over events such as wars or economic crises. Selecting important dates is one of the sub tasks of TLS. For an event within a certain time frame. Date selection is done by using supervised machine learning approaches that estimate the importance of each date separately, using features like frequency of date mentions in news corpora. This approach neglects interactions between different dates, occurs due to connections between sub events. They use supervised machine learning methods that score dates independently of each other. Features are extracted from a corpus of event-related newspaper articles, including frequency-based features (such as how often the date is referred to in the corpus), temporal distance features (such as how long into the future a date keeps being referred to) and topical features (such as whether the date mention is associated with the most significant keywords of the event).

However, in a graphical model score dates jointly makes use of interactions between dates. This improves sub substantially over prior approaches. Also, propose unsupervised variations that perform competitively to prior supervised models. This approach neglects interactions between different dates. This occurs due to connections between sub events. Therefore a joint graphical model for date selection is suggested. Even unsupervised versions of this model perform as well as supervised state-of-the-art approaches. Training data with parameter tuning, it outperforms prior supervised models by a considerable margin. However, score dates jointly making use of interactions between dates in a graphical model. This improves substantially over prior approaches. Build a date reference graph, which is a fully directed graph  $G = (V, E)$ , where  $V$  is the set of dates mentioned in any text in corpus  $C$ , including publication dates. The edges  $E = e(d_i, d_j)$  indicate that at least one text published on  $d_i$  refers to the date  $d_j$ . Then introduce a random walk model that uses these perspectives to rank the collection of dates.

D. *Discussion on approaches for event timeline generation*

Harsimran Bed et. al. [3] presents a system for automatic event timeline generation from history textbooks.

Table I: Overview of Compared Approaches

Type of Approach	Proposed Approaches	
	Model	Highlight
Summarization Based Approach	Giang Tran et. al. [5]	Joint graphical model for date selection
Syntactic Based Approaches	Sandro Bauer et. al.[4]	Unsupervised method
Algorithm Based Approaches	Harsimran Bed et. al.[3]	Timeline Generation Algorithm
	Omar Alonso et. al.[6]	Pseudo relevance feedback

Additionally, proposes Message Sequence Chart (MSC) and timemap based visualization techniques to visualize an event timeline. In this paper, particularly focus on creating event timeline (or chronology) from history textbooks. Event timelines play an important role in understanding a historical phenomenon. It enables a student to situate her knowledge of history in relation to a spatio-temporal context. Now this paper is only focusing on generating a timeline specific to a human actor.

Omar Alonso et. al. [6] proposes an automatic timeline generation using contextual vectors a pseudo relevance feedback. It actually represents a ranked list of n-grams for a set of tweets related to a hashtag or entity. Important source for getting real time updates from different parts of the world is social media, because of its phenomenal growth. The intuition is that links from trusted news-like domains (e.g., cnn.com, bbc.co.uk, etc.) are likely to be authoritative, contain a title that is well-written in English, and they are easy to read. By relying on a set of worldwide news domains, constructed by mining Twitter accounts and web domains, the timeline has the potential to be diverse. They only using sentences which contains dates, this leads to the wrong selection of events.

Sandro Bauer et. al. [4] content selection for creating timelines from individual history articles is available. However, for most historical periods covered in history articles (e.g., Antiquity or the Middle Ages), such cues are not commonly available, as no news articles from these eras exist. Generating event timelines for arbitrary historical periods is therefore a much harder problem, which requires methods that rely less heavily on the types of rich, parallel and dense



information contained in news clusters. In contrast to existing works on timeline generation, this paper requires the presence of a large corpus of news articles. Exploit lexical cues about the articles subject area, as well as time expressions that are syntactically attached to an event word are used to identify salient events in a given history article.

Also test different methods of ensuring timeline coverage of the entire historical time span described. This gives lights to importance of this paper.

Giang Tran et. al. [5] is about Automatic timeline summarization. Mainly they focusing on supervise methods. They select frequency of time as feature here. Approach leverages the interactions between dates via a joint model based on a date reference graph, improving on individual scoring of dates. This improves the quality of the method. Date selection is challenging, as normally only a small set of the available dates is chosen for inclusion in the timeline. Date selection may be partially subjective: different journalists might include different dates. This makes the system more complex.

### III. CONCLUSION AND FUTURE WORK

In above discussion, wide varieties of methodologies are there for timeline generation from different domains. For extracting events from the given text information extraction technique and natural language processing are used. Syntactic, Algorithm Based and summarization techniques are proposed. Outperforming extraction technique is based on natural language processing. Co- reference resolution technique helps to get the better result in that case. In future, timeline for a non-human actor can be generated, e.g., a timeline of art or science in the Renaissance. Also aim to define annotation guidelines. Also aim to define annotation guidelines for annotation of historical events and release a much larger annotated dataset that can be used for various tasks such as entity/event extraction and segmentation, co-reference resolution of named entities as well as events.

### REFERENCES

- [1]. M. Fullan, *Change forces: Probing the depths of educational reform*. Routledge, 2012.
- [2]. A.Chisholm, W. Radford, and B. Hachey, "Learning to generate onesentence biographies from wikidata," arXiv preprint arXiv:1702.06235, 2017.
- [3]. H. Bedi, S. Patil, S. Hingmire, and G. Palshikar, "Event timeline generation from history textbooks," in *Proceedings of the 4th Workshop on Natural Language Processing Techniques for Educational Applications (NLPTEA 2017)*, pp. 69–77, 2017.
- [4]. S. Bauer and S. Teufel, "Unsupervised timeline generation for wikipedia history articles," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 2343–2349, 2016.
- [5]. R. Agrawal, S. Gollapudi, A. Kannan, and K. Kenthapadi, "Studying from electronic textbooks," in *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pp. 1715–1720, ACM, 2013.
- [6]. O. Alonso, S.-E. Tremblay, and F. Diaz, "Automatic generation of event timelines from social data," in *Proceedings of the 2017 ACM on Web Science Conference*, pp. 207–211, ACM, 2017.
- [7]. J. Pustejovsky, P. Hanks, R. Sauri, A. See, R. Gaizauskas, A. Setzer, D. Radev, B. Sundheim, D. Day, L. Ferro, et al., "The timebank corpus," in *Corpus linguistics*, vol. 2003, p. 40, Lancaster, UK., 2003.
- [8]. G. Angeli, M. J. J. Premkumar, and C. D. Manning, "Leveraging linguistic structure for open domain information extraction," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, vol. 1, pp. 344–354, 2015.
- [9]. S. Chandra, *Medieval India: From Sultanat to the Mughals Part-II*. HarAnand Publications, 2005.
- [10]. R. De Keyser and P. Vandepitte, "Historical formation. design of vision," Brussels, Belgium: Flemish Board for Catholic Secondary Education and Huijgen, I, Van Boxtel, C, Van de Grift, W. and Holthuis, P.(2014) Towards historical perspective taking: students' way of reasoning when imagining the actions of people in the past'(submitted manuscript), 1998.
- [11]. T. Ge, W. Pei, H. Ji, S. Li, B. Chang, and Z. Sui, "Bring you to the past: Automatic generation of topically relevant event chronicles," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, vol. 1, pp. 575–585, 2015.
- [12]. J. Li and C. Cardie, "Timeline generation: Tracking individuals on twitter," in *Proceedings of the 23rd international conference on World wide web*, pp. 643–652, ACM, 2014.
- [13]. C. Yogi, P. Solanki, A. Zurunge, G. Patel, et al., "A survey on continuous summarization and timeline generation for topic evolutionary tweet streams," *Imperial Journal of Interdisciplinary Research*, vol. 3, no. 5, 2017.
- [14]. K. Bontcheva and D. Rout, "Making sense of social media streams through semantics: a survey," *Semantic Web*, vol. 5, no. 5, pp. 373–403, 2014.
- [15]. V. L. Narasamma and M. Sreedevi, "Modeling of tweet summarization systems using data mining techniques: A review report," *Indian Journal of Science and Technology*, vol. 9, no. 44, 2016.
- [16]. <https://venngage.com/blog/timeline->