



Application of natural language processing algorithms for extracting information from news articles in event-based surveillance

Victoria Ng^{1*}, Erin E Rees¹, Jingcheng Niu², Abdelhamid Zaghlool³, Homeira Ghiasbeglou³, Adrian Verster⁴

Abstract

The focus of this article is the application of natural language processing (NLP) for information extraction in event-based surveillance (EBS) systems. We describe common information extraction applications from open-source news articles and media sources in EBS systems, methods, value in public health, challenges and emerging developments.

Suggested citation: Ng V, Rees EE, Niu J, Zaghlool A, Ghiasbeglou H, Verster A. Application of natural language processing algorithms for extracting information from news articles in event-based surveillance. *Can Commun Dis Rep* 2020;46(6):186–91. <https://doi.org/10.14745/ccdr.v46i06a06>

Keywords: natural language processing, NLP, event-based surveillance, algorithms, information extraction, open-source data

Background

Natural language processing (NLP) methods enable computers to analyse, process and derive meaning from human discourse. Although the field of NLP has been around since the 1950s (1), progress in technology and methods in recent years have made NLP applications easier to implement, with some tasks outperforming human performance (2). There are many day-to-day applications of NLP including machine translation, spam recognition and speech recognition. NLP is a powerful tool in health care because of the large volumes of text data, for example, electronic health records, being produced. Indeed electronic health records have already been the focus of NLP applications, including detecting melanocytic proliferations (3,4), the risk of dementia (5) and neurological phenotypes (6). But NLP applications in health care extend beyond electronic health records, for example, it is possible to identify people with Alzheimer's disease based on their speech patterns (7).

The focus of this article is the application of NLP for information extraction in event-based surveillance (EBS) systems. We describe common information extraction applications from open-source news articles and media sources in EBS systems, methods, value in public health, challenges and emerging developments.

EBS systems mine the Internet for open-source data, relying on informal sources (e.g. social media activity) and formal sources (e.g. media or epidemiological reports from individuals, media

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



Affiliations

¹ National Microbiology Laboratory, Public Health Agency of Canada

² Department of Computer Science, University of Toronto, Toronto, ON

³ Centre for Emergency Preparedness and Response, Public Health Agency of Canada

⁴ Food Directorate, Health Canada, Ottawa, ON

*Correspondence:

victoria.ng@canada.ca

outlets and/or health organizations) to help detect emerging threats (8). Operational systems include the Public Health Agency of Canada's Global Public Health Intelligence Network (9), HealthMap (10) and the World Health Organization's Epidemic Intelligence from Open Sources (11). Due to the growing volume, variety and velocity of digital information, a wealth of unstructured open-source data is generated daily, mainly as spoken or written communication (9). Unstructured open-source data contains pertinent information about emerging threats that can be processed to extract structured data from the background noise to aid in early threat detection (12). For EBS systems, this includes information about what happened (threat classification; number of cases), where it happened (geolocation) and when it happened (temporal information). The ability to identify this information allows governments and researchers to monitor and respond to emerging infectious disease threats.

One of the challenges in infectious disease surveillance, such as COVID-19, is that there is an immense amount of text data continuously being generated, and in an ongoing pandemic, this amount can be far more than humans are capable of processing. NLP algorithms can help in these efforts by automating the



filtering of large volumes of text data to triage articles into levels of importance and to identify and extract key pieces of information.

In this article, we discuss some important NLP algorithms and how they can be applied to public health. For a glossary of common technical terminology in NLP, see **Table 1**.

Table 1: Glossary of common technical terminology in natural language processing

Term	Definition
(Linguistic) annotation	The association of descriptive or analytic notations with language data, generally performed to generate a corpus for algorithm training
Artificial intelligence (AI)	A branch of computer science dealing with the simulation of human intelligence by machines
Computational linguistics (CL)	The branch of computer science trying to model human language (including various linguistic phenomena and language related applications) using computational algorithms
Corpora (singular – corpus)	A set of articles where the unstructured text has been annotated (labelled) to identify different types of named entity. Corpus are developed for different domains to train ML algorithms to identify named entities (e.g. WikToR corpus of Wikipedia articles for geographic locations, TimeBank corpus of new report documents for temporal information)
F1 score	A performance measure used to evaluate the ability of NLP to correctly identify NEs by calculating the harmonic mean of precision and recall: $F1 = 2 * \text{Precision} * \text{Recall} / (\text{Recall} + \text{Precision})$. The F1 score privileges balanced algorithms because the score tends toward the least number, minimizing the impact of large outliers and maximizing the impact of small ones
Geocoding	Also known as georesolution, assigns geographic coordinates to toponyms
Geoparsing	The combined process of geotagging and geocoding
Geotagging	A subset of named NER that identifies geographic entities in unstructured text
Machine learning (ML)	The study of computer algorithms that learn patterns from experience. ML approaches may be supervised (the algorithm learn from labelled training samples), unsupervised (the algorithm retrieve patterns from unlabeled data), or semi-supervised (the algorithm perform learning with a small set of labelled data and a large set of unlabeled data)
Named entity (NE)	A word or phrase that identifies an item with particular attributes that make it stand apart from other items with similar attributes (e.g. person, organization, location)
Natural language processing (NLP)	A subfield of AI to process human (natural) language inputs for various applications, including automatic speech recognition, natural language understanding, natural language generation and machine translation
Named entity recognition (NER)	The process of identifying a word or phrase that represents a NE within the text. NER formerly appeared in the Sixth Message Understanding Conference (MUC-6), from which NEs were categorized into three labels: ENAMEX (person, organization, location), TIMEX (date, time) and NUMEX (money, percentage, quantity)
Polysemy	The association of a word or phrase with two or more distinct meanings (e.g. a mouse is a small rodent or a pointing device for a computer)

Table 1: Glossary of common technical terminology in natural language processing (continued)

Term	Definition
Precision (also known as positive predictive value)	Percentage of named entities found by the algorithm that are correct: (true positives) / (true positives + false positives)
Recall	Fraction of the total amount of relevant instances that were actually retrieved (true positives) / (true positives + false negatives)
Semi-supervised	Due to the high cost of creating annotated data, semi-supervised learning algorithms combine the learning from a small set of labelled data (supervised) and a large set of unlabeled data (unsupervised) to achieve the tradeoff between cost and performance
RSS feed	RSS stands for Really Simple Syndication or Rich Site Summary, it is a type of web feed that allows users and applications to receive regular and automated updates from a website of their choice without having to visit websites manually for updates
Supervised learning	Supervised learning algorithms is the type of ML algorithms that learn from labelled input-output pairs. Features of the input data are extracted automatically through learning, and patterns are generalized from those features to make predictions of the output. Common algorithms include hidden Markov models (HMM), decision trees, maximum entropy estimation models, support vector machines (SVM) and conditional random fields (CRF)
Synonyms	Words of the same language that have the same or nearly the same meaning as another
Toponym	A NE of the place name for a geographic location such as a country, province and city
Unsupervised learning	A type of ML method that does not use labelled data, but instead, typically uses clustering and principal component analytical approaches so that the algorithm can find shared attributes to group the data into different outcomes

NLP algorithms and their application to public health

The simplest way to extract information from unstructured text data is by keyword search. Though effective, this ignores the issue of synonyms and related concepts (e.g. nausea and vomiting are related to stomach sickness); it also ignores the context of the sentence (e.g. Apple can be either a fruit or a company). The problem with identifying and classifying important words (entities) based on the structure of the sentence is known as named entity recognition (NER) (13). The most common entities are persons, organizations and locations. Many early NER methods were rule-based, identifying and classifying words with dictionaries (e.g. dictionary of pathogen names) and rules (e.g. using “H#N#” to classify a new influenza strain not found in the dictionary) (14). Synonyms and related concepts can be resolved using databases that organize the structure of words in the language (e.g. WordNet (15)). Newer NER methods use classifications and relationships predefined in corpora to

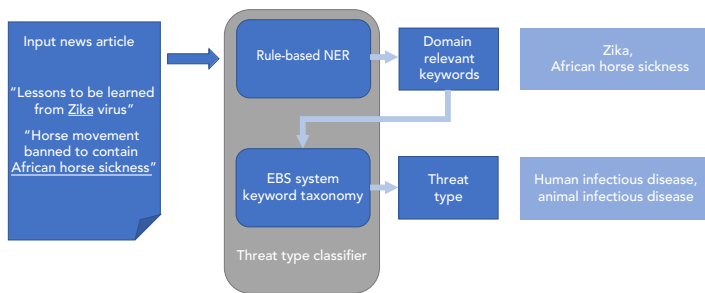


develop machine learning (ML) algorithms to identify and classify entities (13). For NER, terms are annotated to categories and the algorithm learns how to recognize other examples of the category from the term and surrounding sentence structure. Because language data are converted to word tokens as part of the analysis, NLP algorithms are not limited to languages using the Latin alphabet; they can also be used with character-based languages such as Chinese.

1. Article classification (threat type)

Classifying articles by taxonomy keywords into threat types allows EBS system users to prioritize emerging threats. For example, analysts monitoring an event can filter out articles to focus on a specific threat category. Rule-based NER identifies keywords to assign each article to different categories of health threats (e.g. disease type). Keywords are then organized into a predetermined, multilingual taxonomy (e.g. "Zika virus" is a human infectious disease, "African horse sickness" is an animal infectious disease, etc.) that can be updated as new threats are discovered. The taxonomy takes advantage of the structure of the language similar to WordNet (16). This mitigates part of the problem with keyword matching because it allows synonyms and related concepts to stand in for one another (Figure 1).

Figure 1: Article classification



Abbreviations: EBS, event-based surveillance; NER, named entity recognition

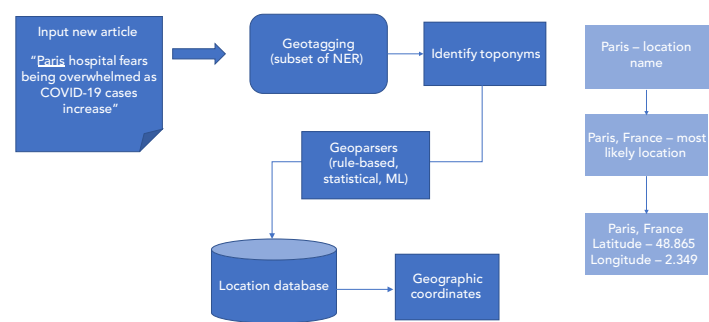
2. Geoparsing

Identifying places where health-related events are reported from articles can help locate susceptible populations. Geoparsing is the task of assigning geographic coordinates to location entities (i.e. toponyms such as city, country) identified in unstructured text. The process starts with geotagging, a subset of NER for identifying the toponyms, and then geocoding to assign geographic coordinates from a dictionary such as from GeoNames (17). Geoparsers use computational methods that are rule-based, statistical and based on ML. The general approach of geoparsing is to characterize toponyms by a set of features (e.g. toponym name, first and last character position in text, character length). Feature information is then processed by computational methods to link each toponym to a geographic name in a location database (e.g. GeoNames (17)) and then assign the corresponding coordinates (18).

Advancements in geoparsing, like other NLP applications, focus on increasing leverage from unstructured text to resolve

ambiguities. One advancement is using semi-supervised learning techniques that utilize programmatically generated corpora to train ML algorithms from larger datasets of annotated examples. Using code to annotate articles is faster and results in larger and more consistent corpora than from human annotation (19). Leveraging more context is also resulting from extending feature information to be topological (spatial relationships among toponyms, e.g. distance to closest neighbouring toponym) (20). A toponym from a phrase like "There are new cases of influenza in London" can be difficult to resolve because there are multiple potential locations. Toponym coordinates can be resolved by assigning a bias towards more populated areas because they are typically mentioned more often in discourse; however, emerging diseases do not always favour highly populated areas (Figure 2).

Figure 2: Geoparsing



Abbreviations: ML, machine learning; NER, named entity recognition

3. Temporal information extraction and temporal reasoning

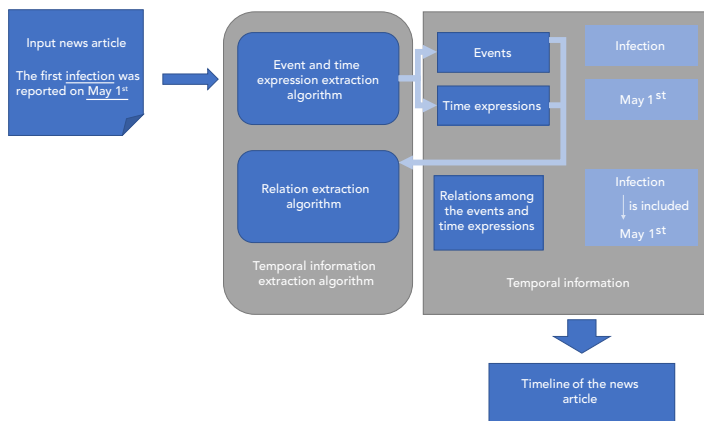
Identifying the timing of events described in articles is necessary for coherent temporal ordering of those events. It is important to be able to differentiate an article reporting on a new event from an article reporting on a previous known event. The most common temporal identifiers in EBS systems are the article publication date and the received/import date (the timestamp for receiving the article into the EBS system). Neither of these dates extract the reported timing of event described in the articles. A subset of NLP—temporal information extraction—has been developed to extract this information. Temporal information extraction is used to identify tokens in text that contain temporal information of relevant events.

Two subtasks of temporal information extraction help resolve ambiguities arising from complicated narratives reporting on multiple events. First, temporal relation extraction focuses on classifying temporal relationships between the extracted events and temporal expressions. Using those relationships, EBS systems can anchor events to time (e.g. in the sentence "the first infection was reported on May 1st," the relation between the event "infection" and the date "May 1st" is used to timestamp the first infection). Second, temporal reasoning (21) focuses on chronological ordering of events through inference.



Multiple temporal information extraction systems have been developed including TimeML (developed for temporal extraction of news articles in finance) (22); ISO-TimeML (a revised version of TimeML) (23); and THYME (developed for temporal extraction in patient records) (24). Results have reached near-human performance (25–28). Based on these annotation standards, an annotation standard for news articles in the public health domain, Temporal Histories of Epidemic Events (THEE), was recently developed for EBS systems by the authors of this article (29) (Figure 3).

Figure 3: Temporal information extraction and temporal reasoning

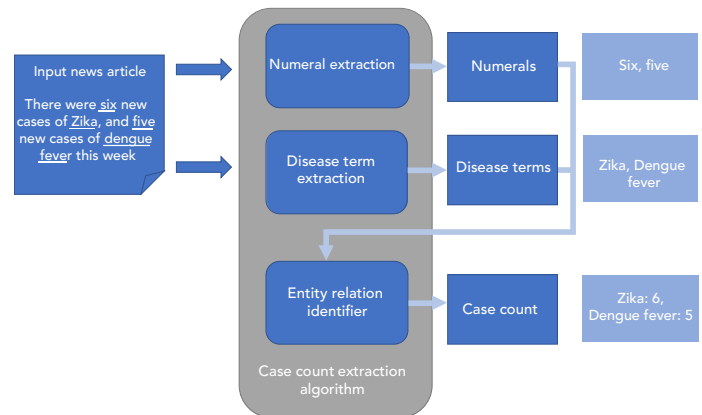


4. Case count extraction

Extracting the number of disease cases reported in articles would help EBS system users to monitor and forecast disease progression. Currently, there is no NLP algorithm incorporated into EBS systems capable of this task, however, there are algorithms capable of tackling related tasks that can be leveraged to develop a case count algorithm. News articles in epidemiology frequently mention the occurrence of disease cases (e.g. “There were six new cases of Zika this week”) so that identifying cases requires identifying the relationships between a quantitative reference in the text (six new cases) and a disease term (of Zika). Many algorithms already identify relationships between entities in diverse fields, for example, the ReEx algorithm identifies relations between genes that are recorded in MEDLINE abstracts and performs with an F1 of 0.80 (30). Based on the ReEx algorithm, an algorithm has been developed to identify sentences in news articles that report on case counts of foodborne illnesses (31).

The authors of this article are developing and refining this algorithm to extract case count information from sentences that have been identified to contain case count information (Figure 4).

Figure 4: Case count extraction

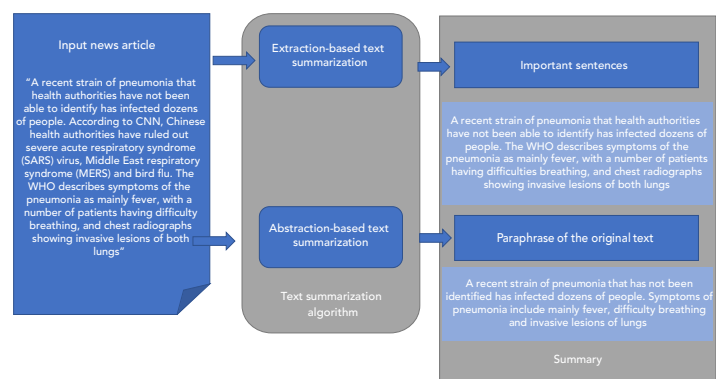


5. Automatic text summarization

The goal of text summarization is to quickly and accurately create a concise summary that retains the essential information in the original text. Text summarization in EBS systems would increase the number of articles that can be scanned for threat detection by reducing the volume of text that needs to be read. There are two main types of text summarization: extraction-based and abstraction-based. Extraction-based summarization involves identifying the most important key words and phrases from the text and combining them verbatim to produce a summary. Abstraction-based summarization uses a more sophisticated technique that involves paraphrasing the original text to write new text, thus mimicking human text summarization.

Text summarization in NLP is normally developed using supervised ML models trained on corpora. For both extraction-based and abstraction-based summarization, key phrases are extracted from the source document using methods including part-of-speech tagging, word sequences or other linguistic pattern recognition (32). Abstraction-based summarization goes a step further and attempts to create new phrases and sentences from the extracted key phrases. A number of techniques are used to improve the level of abstraction including deep learning techniques and pre-trained language models (33) (Figure 5).

Figure 5: Automatic text summarization



Abbreviation: WHO, World Health Organization



Discussion

NLP has a huge number of potential applications in health care because of the omnipresence of text data. Electronic health records are an obvious source of data for NLP application, but text relevant to health care extends far beyond health records; it includes traditional and social media sources, which are the main sources of data for EBS systems, in addition to official government reports and documents.

As NLP algorithms can interpret text and extract critical information from such diverse sources of data, they will continue to play a growing role in the monitoring and detection of emerging infectious diseases. The current COVID-19 pandemic is an example of where NLP algorithms could be used for the surveillance of public health crises. (This is, in fact, something several co-authors of this article are currently developing).

While NLP algorithms are powerful, they are not perfect. Current key challenges involve grouping multiple sources referring to the same event together and dealing with imperfections in the accuracy of information extraction due to nuances in human languages. Next-generation information extraction NLP research that can improve these challenges include event resolution (deduplication and linkage of the same events together) (34) and advancements in neural NLP approaches such as transformers networks (35), attention mechanism (36) and large-scale language models such as ELMo (37), BERT (38) and XLNet (39) to improve on the current performance of algorithms.

Conclusion

We have discussed several common NLP extraction algorithms for EBS systems: article classification, which can identify articles that contain crucial information about the spread of infectious diseases; geolocation, which identifies where a new case of the disease has occurred; temporal extraction, which identifies when a new case occurred; case count extraction, which identifies how many cases occurred; and article summarization, which can greatly reduce the amount of text for a human to read.

Although the field of NLP for information extraction is well established, there are many existing and emerging developments relevant to public health surveillance on the horizon. If capitalized, these developments could translate to earlier detection of emerging health threats with an immense impact on Canadians and the world.

Conflict of interest

None.

Funding

EE Rees and V Ng are currently co-Principal Investigators for a Canadian Safety and Security Program (CSSP), a federally-funded

program from the Department of National Defence, the grant is a three-year grant titled 'Incorporating Advanced Data Analytics into a Health Intelligence Surveillance System' – CSSP-2018-CP-2334.

References

1. Bates M. Models of natural language understanding. *Proc Natl Acad Sci USA* 1995 Oct;92(22):9977–82. [DOI PubMed](#)
2. Wang A, Pruksachatkun Y, Nangia N, Singh A, Michael J, Hill F, Levy O, Bowman S. SuperGLUE: a stickier benchmark for general-purpose language understanding systems. Ithaca (NY): arXiv; 2019 (revised 2020-02-13; accessed 2020-02-24). <https://arxiv.org/abs/1905.00537>
3. Lott JP, Boudreau DM, Barnhill RL, Weinstock MA, Knopp E, Piepkorn MW, Elder DE, Knezevich SR, Baer A, Tosteson AN, Elmore JG. Population-based analysis of histologically confirmed melanocytic proliferations using natural language processing. *JAMA Dermatol* 2018 Jan;154(1):24–9. [DOI PubMed](#)
4. Nguyen AN, Truran D, Kemp M, Koopman B, Conlan D, O'Dwyer J, Zhang M, Karimi S, Hassanzadeh H, Lawley MJ, Green D. Computer-Assisted Diagnostic Coding: effectiveness of an NLP-based approach using SNOMED CT to ICD-10 mappings. *AMIA Annu Symp Proc* 2018 Dec;2018:807–16. [PubMed](#)
5. McCoy TH Jr, Han L, Pellegrini AM, Tanzi RE, Berretta S, Perlis RH. Stratifying risk for dementia onset using large-scale electronic health record data: A retrospective cohort study. *Alzheimers Dement* 2020 Mar;16(3):531–40. [DOI PubMed](#)
6. Wheeler E, Mair G, Sudlow C, Alex B, Grover C, Whiteley W. A validated natural language processing algorithm for brain imaging phenotypes from radiology reports in UK electronic health records. *BMC Med Inform Decis Mak* 2019 Sep;19(1):184. [DOI PubMed](#)
7. Karlekar S, Niu T, Bansal M. Detecting linguistic characteristics of Alzheimer's Dementia by interpreting neural models. In: *Proceedings of NAACL-HLT 2018*. New Orleans (LA). June 1–6, 2018.
8. World Health Organization. A guide to establishing event-based surveillance. Manila (PH): WHO Regional Office for the Western Pacific; 2008.
9. Dion M, AbdelMalik P, Mawudeku A. Big Data and the Global Public Health Intelligence Network (GPHIN). *Can Commun Dis Rep* 2015 Sep;41(9):209–14. [DOI PubMed](#)
10. Freifeld CC, Mandl KD, Reis BY, Brownstein JS. HealthMap: global infectious disease monitoring through automated classification and visualization of Internet media reports. *J Am Med Inform Assoc* 2008 Mar-Apr;15(2):150–7. [DOI PubMed](#)
11. World Health Organization. Epidemic Intelligence from Open Sources (EIOS): saving lives through early detection. Geneva (CH): World Health Organization; 2020 (accessed 2020-01-24). <https://www.who.int/eios>
12. Barboza P, Vaillant L, Mawudeku A, Nelson NP, Hartley DM, Madoff LC, Linge JP, Collier N, Brownstein JS, Yangarber R, Astagneau P. Early Alerting Reporting Project Of The Global Health Security Initiative. Evaluation of epidemic intelligence systems integrated in



- the early alerting and reporting project for the detection of A/H5N1 influenza events. *PLoS One* 2013;8(3):e57252. [DOI PubMed](#)
13. Nadeau D, Sekine S. A survey of named entity recognition and classification. In: Nadeau D, Sekine S, editors. *Named entities: recognition, classification and use*. *Linguisticae Investigationes* 2007;30(1):3–26.
 14. Sekine S, Nabota C. Definition, dictionaries and tagger for extended named entity hierarchy. In: *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004)*. Lisbon (PT). 26–28 May 2004.
 15. Princeton University. *WordNet: an electronic lexical database*. Princeton (NJ): Princeton University; 2010 (accessed 2020-01-24). <https://wordnet.princeton.edu/>
 16. Miller GA. *WordNet: a lexical database for English*. *Commun ACM* 1995;38(11):39–41.
 17. GeoNames [database]. 2020. <https://www.geonames.org/>
 18. Santos J, Anastácio I, Martins B. Using machine learning methods for disambiguating place references in textual documents. *GeoJournal* 2015;80(3):375–92. [DOI](#)
 19. Gritta M, Pilehvar MT, Limsopatham N, Collier N. What's missing in geographical parsing? *Lang Resour Eval* 2018;52(2):603–23. [DOI PubMed](#)
 20. DeLozier G, Baldrige J, London L. Gazetteer-independent toponym resolution using geographic word profiles. *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*. Austin (TX): AAAI Press; 2015. p. 2382–8.
 21. Allen JF. Maintaining knowledge about temporal intervals. In: Weld DS, De Kleer J. *Readings in qualitative reasoning about physical systems*. 1990, Elsevier; pp. 361–72. [DOI](#)
 22. Pustejovsky J, Ingria R, Sauri R, Castano JM, Littman J, Gaizauskas RJ, Setzer A, Katz G, Mani I. The specification language TimeML. In: Mani I, Pustejovsky J, Gaizauskas R, editors. *The language of time: a reader*. 2005. p. 545–58.
 23. Pustejovsky J, Kiyong L, Bunt H, Romary L. ISO-TimeML: an international standard for semantic annotation. In: *Proceeding from the International Conference on Language Resources and Evaluation 2010*. La Valette (MT). 2010 May.
 24. Styler WF 4th, Bethard S, Finan S, Palmer M, Pradhan S, de Groen PC, Erickson B, Miller T, Lin C, Savova G, Pustejovsky J. Temporal annotation in the clinical domain. *Trans Assoc Comput Linguist* 2014 Apr;2:143–54. [DOI PubMed](#)
 25. Chambers N. Navytime: event and time ordering from raw text. In: *Second Joint Conference on Lexical and Computational Semantics (*SEM)*. Volume 2: *Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*. Annapolis (MD): Naval Academy; 2013.
 26. Lee HJ, Xu H, Wang J, Zhang Y, Moon S, Xu J, Wu Y. UHealth at SemEval-2016 Task 12: an end-to-end system for temporal information extraction from clinical notes. In: *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. San Diego (CA): Association for Computational Linguistics; 2016. [DOI](#)
 27. Strötgen J, Zell J, Gertz M. HeidelTime: tuning English and developing Spanish resources for TempEval-3. In: *Proceedings of the Second Joint Conference on Lexical and Computational Semantics (*SEM)*. Volume 2: *Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*. Atlanta (GA);2013.
 28. Lin C, Miller T, Dligach D, Bethard S, Savova G. A BERT-based universal model for both within- and cross-sentence clinical temporal relation extraction. In: *Proceedings of the 2nd Clinical Natural Language Processing Workshop*. Minneapolis (MN);2019.
 29. Niu J, Ng V, Penn G, Rees E. Temporal histories of epidemic events (THEE): a case study in temporal annotation for public health. In: *Proceedings of the International Conference on Language Resources and Evaluation*. Marseille (FR);2020.
 30. Fundel K, Küffner R, Zimmer R. RelEx--relation extraction using dependency parse trees. *Bioinformatics* 2007 Feb;23(3):365–71. [DOI PubMed](#)
 31. Nasheri N, Vester A, Petronella N. Foodborne viral outbreaks associated with frozen produce. *Epidemiol Infect* 2019 Oct;147:e291. [DOI PubMed](#)
 32. Aries A, Eddine ZD, Hidouci WK. Automatic text summarization: what has been done and what has to be done. [arXiv:1904.00688](https://arxiv.org/abs/1904.00688). <https://arxiv.org/abs/1904.00688>
 33. Kryscinski W, Pauls R, Xiong C, Socher R. Improving abstraction in text summarization. In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Brussels (BE): Association for Computational Linguistics; 2018. [DOI](#)
 34. Petroni F, Raman N, Nugent T, Nourbakhsh A, Panic Z, Shah S, Leidner J. An extensive event extraction system with cross-media event resolution. In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2018.
 35. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A, Kaiser Ł, Polosukhin I. Attention is all you need. In: *Proceedings from the 31st Conference on Neural Information Processing Systems (NIPS 2017)*. Long Beach (CA);2017.
 36. Liu B, Lane I. Attention-based recurrent neural network models for joint intent detection and slot filling. *Proc Interspeech* 2016;685–9. [DOI](#)
 37. Peters M, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, Zettlemoyer L. Deep contextualized word representations. In: *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT) 2018*. New Orleans (LA). [DOI](#)
 38. Devlin J, Chang M, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR* 2018;abs/1810.04805. <https://arxiv.org/abs/1810.04805>
 39. Yang Z, Dai Z, Yang Y, Carbonell J, Salakhutdinov R, Le Q. XLNet: generalized autoregressive pretraining for language understanding. *arXiv* 2019;1906.08237. <https://arxiv.org/abs/1906.08237>