

Temporal expression identification in natural language processing: A survey

Anupama Sharma, Madan Lal
Department of Computer Science and Engineering,
Punjabi university, Patiala, Punjab, India

Abstract

One of the most important measures of our lives is undoubtedly time. Temporal information is crucial in various fields of natural language processing like in information extraction, question answering, text summarization etc. There has been fascinating work in the past few years on the analysis and exploitation of temporal information. The included literature on temporal expression recognition is classified into number of categories. Different temporal expression recognition approaches, machine learning algorithms, categories of temporal expressions are reported. Future researchers investigating temporal expressions will find this review to be a useful and concise resource.

Keywords

Temporal expression, Temporal information extraction, Temporal information systems, Machine translation.

1. Introduction

Information space revolves around time, and this has been examined in various fields that include information extraction, question answering, text summarization, machine translation, searching in Time, temporal clustering. Temporal expressions can be used to describe the chronological context of the document or to recreate a specific historical era [34]. In order to deal with the growing volume of data available today, information extraction methods and methodologies have been extensively deployed. One of the many types of information that can be derived from such information sources, such as text documents, is temporal information. The ability to extract time data from text is essential for understanding a language and a crucial component of many language processing programs [8].

1.1 Classification of temporal expressions

- Absolute: When a certain moment, like Sunday, is completely clear and explicit.
- Deictic: When determining the moment to which an expression refers requires using the moment of expression e.g. one week ago
- Anaphoric: When the reference is made at a time other than the moment of expression for instance, in the morning

1.2 There are two main categories into which temporal expression processing can be divided:

- I. Identifying temporal expressions that relate to durations, metrics, and periods or points of time.
- II. Reinterpreting these time-based references in a standard manner that clarifies dates and times from various time zones, accurately captures the meaning of expressions, and facilitates deducing the order of events as they are described in the texts.

1.3 Temporal information characteristics:

- Well-defined:

The fact that we can determine the relationship between two points in time indicates that temporal information is well defined [38].

- Normalization:

Time-related data can be normalized.

- Organization:

Information about time can be organized in hierarchy [4].

1.4. Natural language presents a number of challenges for temporal representation and reasoning, including:

- Wide range of temporal expressions: their wide diversity makes them challenging to recognize, the variations they have may make it more difficult to compare them as well, their meaning may be unclear or challenging to ascertain
- Establishing time relations between events is complex.
- Handling granularity of temporal expressions is complex.
- Complexities due to conjunction, anaphora and ambiguity [6].

1.5 Paper Organization

Section 1 provides the introduction to temporal expression, characteristics of temporal expression information, also, the challenges faced for temporal representation and reasoning are provided. Section 2 describe categories of temporal expressions along with the applications of temporal expression. For enhancing the objectivity of the review process, three research questions are framed that are presented in the section 3.2 Electronic database search is core element in a systematic review. Section 3.3 describes the electronic database searched for the study. Inclusion and exclusion metric (Section 3.5) are vital in systematic review as they specify the studies to be included and excluded ensuring that review is centered and pertinent. Section 4 discuss the results i.e. Answers to the framed questions.

2. Time in documents

A set of documents contains a lot of temporal information. Different categories of temporal information are described in the section 2.1. The realization of the temporal expressions in natural language is explained in Section 2.2.

2.1 Types of Temporal Information

TimeML, a standard markup language for temporal information, divides temporal expressions into four types: date, time, duration, and set [35].

Expressions for date and time indicate distinct times that correspond to specific dates and times, respectively. Duration expressions give information about the duration of interval. Set expressions provide details about an event's periodic nature.

2.2 Temporal expressions and their possible forms:

Temporal expressions can be expressed in a number of ways. They are as follows:

Explicit temporal expressions represent a particular point in time. The term itself provides all of the necessary information for normalization and is thus completely described. Relative temporal expressions cannot be normalized without considering some contextual information. They need some context for complete information. Implicit temporal expressions represent some events [4].

2.3 Temporal Tagging

Extraction and standardization of the temporal expressions to a standard format are the objectives of temporal tagging. TIMEX2 and TIMEX3 are commonly used standard formats. TIMEX2 tags comprises pre as well as post modifiers of temporal expressions. This standard also allows the nested form of temporal expressions.

2.4 Applications of Temporal expressions:

- In information extraction: In information extraction, the temporal expressions are used to obtain information about an event. That event is related with time, date, frequency, that is given by temporal expression.
- In question answering: The temporal information obtained from documents may be used to let the user give query in a temporal manner [4]. In question answering, the temporal expressions are used to get answers of “when” questions. It is also used to answer which event happened prior to particular event.
- For text summarization: Time sensitive summaries can be generated which can be utilized as temporal extract or snippet [26]. A snippet's primary objective is to serve as a representation of a document so that the user can scan the text without having to read the entire document [4].
- In machine translation: In machine translation, the temporal expressions are used for translating the temporal references.
- For viewing events in chronological order [16].
- Search: Time entities may be used to find in documents the past events for example for digital forensics, historical analysis or linguistic analysis [4].
- Spatio-temporal Information recognition: For exploration tasks, some research areas combine spatial and temporal information extracted from text. Spatial information can also be normalized as temporal information can be normalized [17][26]. Temporal information is combined with the geo information and it will be used to explore the documents according to the events. Time Trails is such a system which uses spatio-temporal information [28].
- Temporal Clustering: Temporal clustering is used to dissect a document depending upon

collection by time [4].

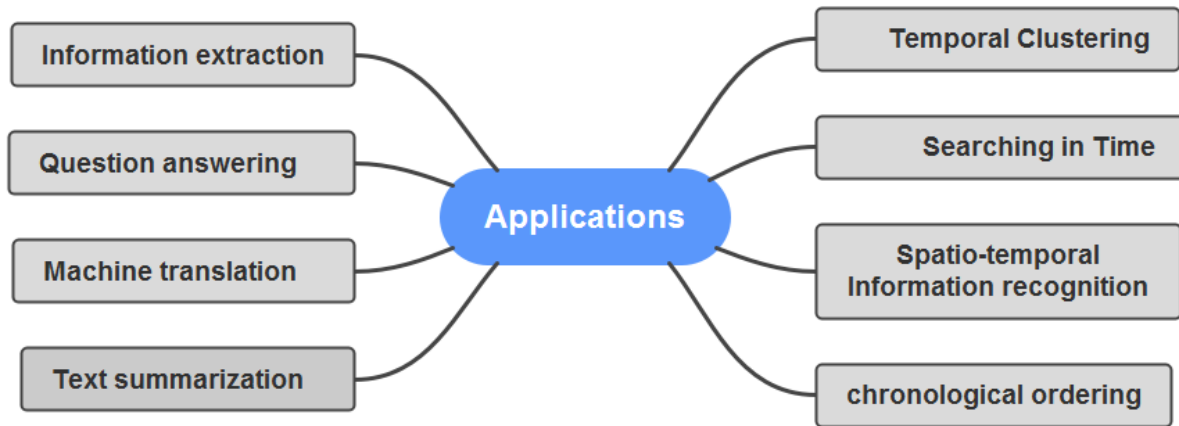


Figure 1. Applications of temporal expressions.

3. Review method

The review includes the steps shown in figure 2 which includes developing the review protocol, conducting the review and analyzing the results.



Figure 2. Review method.

3.1 Planning the review

The review process includes:

- Framing the research questions.
- Database searching.
- Identification of relevant data from the results of a search.
- Extraction and analysis of related data.
- Data specification in appropriate formats, such as tables and graphs.
- Reporting the findings.

3.2 Research questions

Following research questions are framed for carrying out systematic review.

RQ1. What is current status of work done in Temporal expression recognition system?

1.1 What are the approaches used in temporal expressions recognition?

1.2 What are the languages for which the different systems are developed and what are the subject systems?

1.3 What are the temporal expressions systems developed so far?

RQ2. What are the machine learning approaches used for developing different temporal expressions detection systems?

RQ3. What are the different categories of temporal information?

3.3. Sources of information

Finding the pertinent literature to review is an important step in the process [37]. In order to maximize the likelihood of discovering relevant articles, it is necessary to choose the relevant databases prior to beginning the review process [38, 39]. The electronic databases that we searched are mentioned below.

_ ACM Digital Library (www.acm.org)

_ IEEE eXplore (<https://ieeexplore.ieee.org/>)

_ Springer (<https://link.springer.com/>)

3.4. Search criteria

With the objective of covering all articles in this field, the research papers are extracted from all three sources. Table 1 shows the e resources, search string used content type and total number of articles.

Table1. Search string

Sr no.	E-resources	Search string	Product/content type	#
1.	https://ieeexplore.ieee.org/	Temporal AND expressions	Conferences publications, Journals and magazines	7
2.	https://link.springer.com/	Temporal AND expressions AND "Natural language processing"	Conferences publications, Journals, chapters and articles	18
3.	acm.org	[All: temporal] AND [All: expressions] AND [All: nlp]	Conferences publications, Journals	74

3.5 Inclusion and Exclusion

The selection of research papers is the main step in a review as it pertains to the database that will be used as the basis for the review [36]. In our case, there are lot of irrelevant publications. These publications come from the fields of geography, medical and NLP domain other than temporal expression recognition. This review included both quantitative and qualitative research articles published up to and including 2023, beginning from the date of the digital library's inception. Only studies written in English were taken into account. Figure 3 shows inclusion and exclusion criteria.

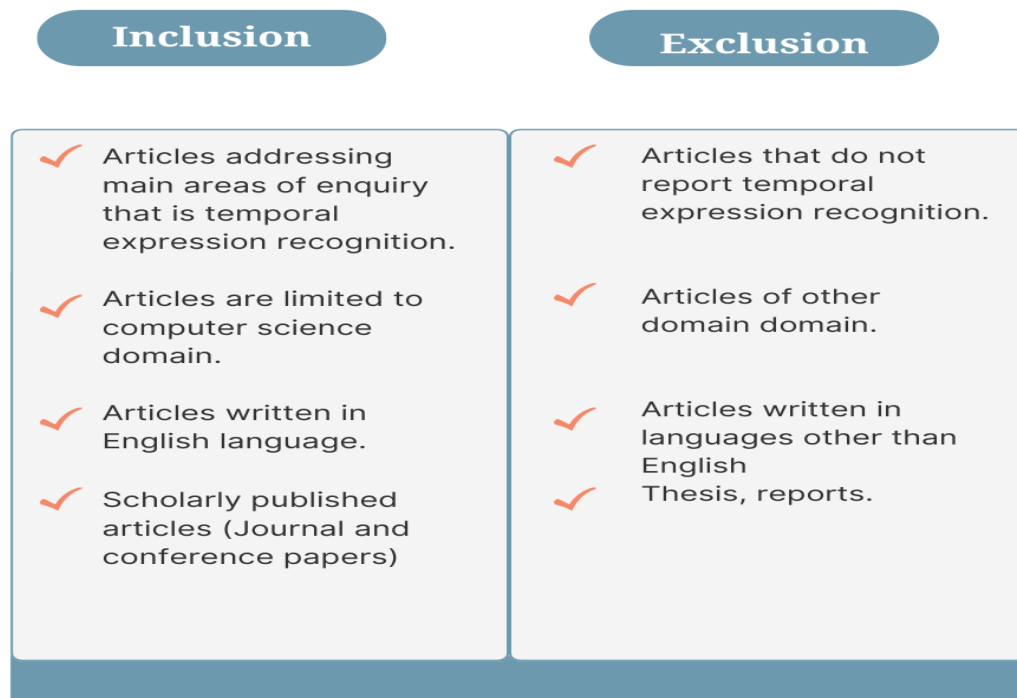


Figure 3. Inclusion and exclusion criteria.

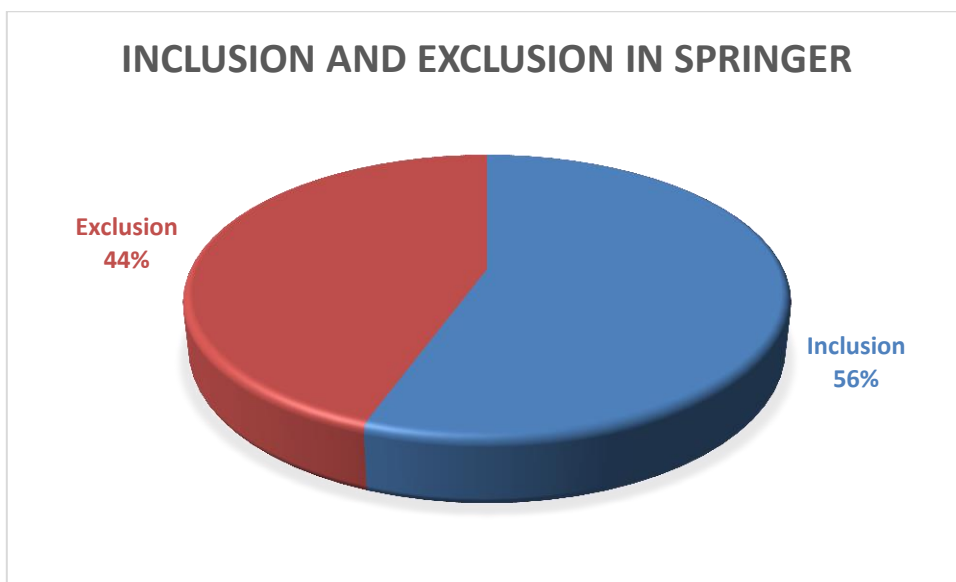


Figure 4. Inclusion and exclusion in springer.

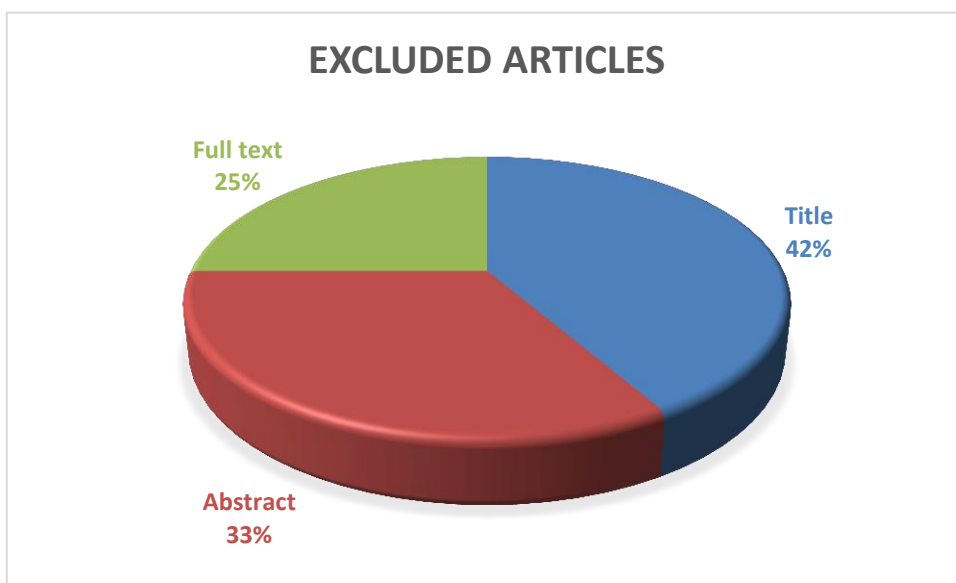


Figure 5. Excluded articles in springer.

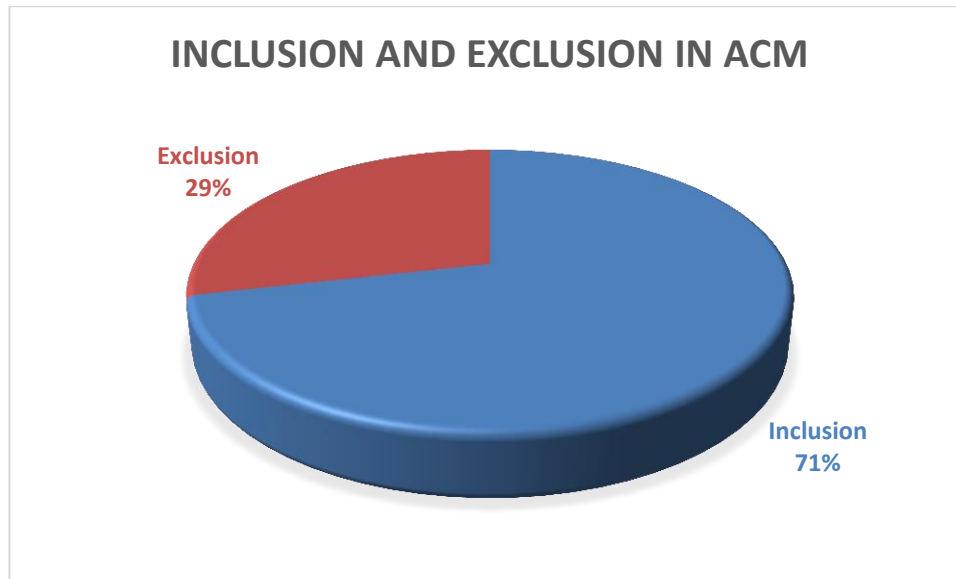


Figure 6. Inclusion and exclusion in ACM.

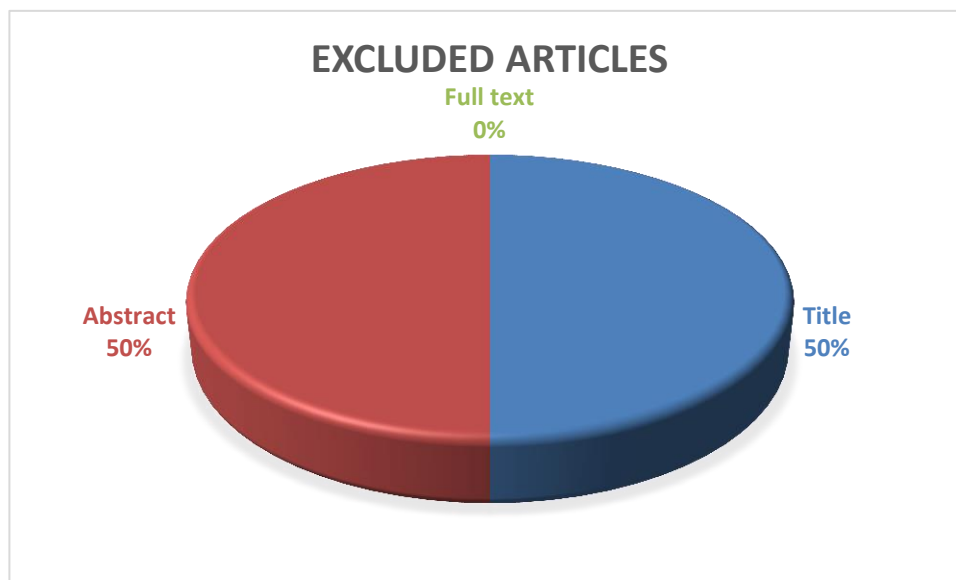


Figure 7. Exclusion in ACM.

4. Results

In order to address the research questions described in the section 3.2, the objective of the study is to review the pertinent literature that has already been published.

4.1 Current status of work done in Temporal expression recognition system.

Table 2 shows temporal expression recognition approaches identified in the selected studies. Temporal expression recognition systems for different languages and corresponding subject systems are depicted in the table 3. Various timex systems are identified; the names along with language for which the system is developed are mentioned in table number 4.

Vichayakitti and Jaruskulchai [1] discussed rule-based approach for temporal event recognition from

Thai text news. Point of time and time interval are the two classifications that are used for temporal expressions. The task is divided into three groups; word segmentation, part of speech tagging and implementation of the algorithm. To extract the temporal information, first, the enormous news text has been organized in one file for every title news. Following that, the findings of the preceding phase were examined by a language specialist for correctness. At last, the resultant text was annotated with different five types of tags: number, date, end of sentence, vocabulary related to time and the relevant events. The accuracy of the system is 68%.

Tissot et al [8] discussed the categorization of imprecise temporal expressions. By using the manual interpretation of temporal expressions, an innovative approach is developed to get membership functions of temporal expressions. A metric called the F13D is presented that can be used to compare membership functions across all three planes.

Llorens et al [10] discussed their system TIRSemZH (Temporal Information Recognition in Chinese) which utilizes morphosyntactic information along with semantic roles as characteristics. The study's primary focus is on timex and event recognition. The results show the accuracy as 85%. Furthermore the findings reveal that applying semantic roles improves a baseline's performance based solely on morphosyntax. Dale et al [11] discussed three steps that are recognition of temporal expressions, interpretation of temporal expressions and temporal expression normalization. DANTE architecture, a system that reflects the methodology for temporal expression interpretation is presented. Craveiro [12] discussed approach for detection of temporal named entity expressions by utilizing the timex patterns obtained from time references having high probability and low probability expressions are simply ignored. The experiment is split up into two phases: that are identification and classification. The identification section includes getting comprehensive temporal expressions. While assigning the type and subtype specification was the notion behind the classification task. The result shows 80% recall.

Blamey et al [13] discussed distributed approach for temporal expression parsing. A method for generating distributed description of temporal expressions by statistically modeling the definitions derived from an extensive data set is also presented. The distributed model is integrated into a temporal algebraic software system, enabling the implementation of some temporal algebraic operators as algebraic operators.

Nunes [14] discussed Temporal Expression Utilization in Web Search. Two sets of web search queries are used to explore the usage of temporal expressions. The classification of dataset is done on the basis of Home & Garden, Finances, Places, Travel, Personal, Research, Shopping, URL and Misspellings.

Tianyong Hao [15] discussed TEXer, an algorithm used to extract and normalize temporal expressions from clinical data. A triplet, ti , ai , vi is used to represent each temporal expression. In this triplet, ti stands for the expression, ai for the expression's type attribute, and vi for normalized value. As per the TimeML standard, four temporal types are employed that are Time, Date, Set and Duration. TEXer obtained 0.858 recall and a 0.945 precision. Hartrumpf [16] discussed InSicht, A German question answering system It expands upon complete sentence parses, Inferences made from semantic representations, comparing representations of semantics obtained from questioning as well as document sentences. To divide the documents into individual sentences, a tokenizer as well as expression boundary detectors are applied. Mirza [17] discussed temporal recognition and normalization of Indonesian text. A preexisting timex normalizer, TimeNorm for the English language is adapted. To identify temporal expression a deterministic finite state transducer is built. The system achieves F1 score of 92.87% for recognition and 85.26% for normalization.

Mansouri [18] discussed ParsTime, which is a tagger to recognize temporal expressions in Persian

language. The steps of the temporal expression recognizer are pre-processing, temporal annotation, extraction, normalizing. The system achieves 0.89 F1-score. Haffar [19] provided the first publicly available Arabic event and timex corpus with annotations built around the Timeml prerequisites. The steps followed by the system are preprocessing, event extraction, TIMEX3 Extraction and Manual Markup and Verification. Schwitter [33] discussed approach to integrate temporal expressions into controlled natural language (simplified version of natural languages created by limiting the size of vocabulary and grammar). PENG controlled language is used. For constructing temporal specifications, event calculus reasoning is combined with answer set programming. Najafabadipour et al [5] presented temporal tagger to identify as well as normalize temporal expressions in clinical text written in Spanish language. The four tasks of the Temporal Tagger include (1) extracting different temporal expressions; (2) filtering of temporal expressions; (3) resolving of temporal expressions with reference date (4) normalization temporal expressions to a standardized date format. The initial use case temporal tagger results in 0.927 ± 0.021 precision, 0.932 ± 0.021 recall of 0.932 and 0.93 F1 score. The precision, recall, and F1 score for the second use case are 0.831 ± 0.033 , 0.766 ± 0.036 , and 0.797 respectively.

P. Martinez-Barco [7] discussed their system TIRSemZH (Temporal Information Recognition in Chinese) which utilizes morphosyntactic information along with semantic roles as characteristics. The study's primary focus is on timex and event recognition. The results show the accuracy as 85%. Furthermore the findings reveal that applying semantic roles improves a baseline's performance based solely on morphosyntax. Dale et al [11] discussed three steps that are recognition of temporal expressions, interpretation of temporal expressions and temporal expression normalization. DANTE architecture, a system that reflects the methodology for temporal expression interpretation is presented. Craveiro [12] discussed approach for detection of temporal named entity expressions by utilizing the timex patterns obtained from time references having high probability and low probability expressions are simply ignored. The experiment is split up into two phases: that are identification and classification. The identification section includes getting comprehensive temporal expressions. While assigning the type and subtype specification was the notion behind the classification task. The result shows 80% recall. Blamey et al [13] discussed distributed approach for temporal expression parsing. A method for generating distributed description of temporal expressions by statistically modeling the definitions derived from an extensive data set is also presented. The distributed model is integrated into a temporal algebraic software system, enabling the implementation of some temporal algebraic operators as algebraic operators.

Nunes [14] discussed Temporal Expression Utilization in Web Search. Two sets of web search queries are used to explore the usage of temporal expressions. The classification of dataset is done on the basis of business, Games, Health, News, Sports, Autos, Computing, Entertainment, Holidays, Organizations, discussed applications of temporal expression identification which includes question answering, event relation, analysis, and recording, event visualization in time, phrase production, and translation. In order to identify these expressions and annotating the corpus, time expressions are examined that contain a time noun that is defined by a time adverb. Li and Wong [21] developed a system Temporal Information extraction from Chinese Sources (TICS) for Chinese language. The system takes as input the Chinese text, examines each sentence in order to obtain the temporal information, puts this information into a frame, connects all the frames chronologically based on relationships between or within events, and then uses this connected knowledge to answer users' questions. The system deals with two kinds of time relations; absolute relations and relative relations. Kim and Myaeng [22]

discussed temporal information processing as important area of topic detection and tracking. Finite state automata and lexicon comprising time vocabulary is used to derive temporal information. A canonical depiction of a temporal point is generated from the extracted data. The results of the study revealed a precision of 94% and recall of 90%. Ahmed [23] discussed an algorithm to automatically labeled temporal data generation. In order to extract significant information, the system focuses on precisely recognizing temporal information in implicit parts of documents. The discovered representation is utilized to train and categorize the input text via an attention-based approach. Strotgen et al [24] discussed the challenges involved in temporally tagging various languages, specifically Arabic and Vietnamese. Corpus for Arabic and Vietnamese language is annotated. Data is analyzed on recently created and publicly accessible corpora. provided the resources for each of the four languages along with HeidelTime, making HeidelTime the first temporal tagger for Arabic and Vietnamese that is accessible publicly. Ahmed et al [26] presented a lexicon expansion approach built around Wordnet input. By combining words with identical meanings, this technique broadens the vocabulary and improves the precision of event detection methods. Semantic meanings for words and phrases in the text are discovered after the text has been analyzed. Subsequently, the phrases are categorized according to their semantic resemblance, and the most appropriate phrase to depict the cluster is chosen. Finally, the labelled dataset is used to train a classifier to detect new terms for the dictionary. For temporal events, the F1 score is 0.74; for non-temporal events is 0.63.

Table 2. Temporal expression recognition approaches

Author	Approach	Citation
Vichayakitti and Jaruskulchai	Rule based approach	1
Shekhar and Sharma	Rule based approach, Conditional Random field, Support vector machine, Voting technique, LSTM	2
Yin and Jin	Character embedding feature method, neural network	3
Wu, Li, Chen, Lu	Rule based approach and ML classification technique	4
Najafabadipour et al	Rule based approach	5
Sofia and Haro	Finite State Automata	6
P. Martinez-Barco	Grammar rules	7
Tissot et al	Human interpretation-based membership functions	8
Kolomiyets et al	Rule based approach and machine learning approach	9
Llorens et al	Data-driven approach	10
Dale et al	Finite state pattern	11
Craveiro et al	Probabilistic based technique	12
Ramrakhiani and Majumder	Rule based approach, CRF-based classifier, SVM-based classifier	20
Mani et al.	Rule based approach	23
Negri and Marseglia	Rule based approach	39
Mazur and Dale	Rule based approach	42
Filannino and Nenadic	NorMA, CRF and post-processing pipeline, SVM	38

Hagege et al.	Rule based approach	39
Ahn et al	LIBSVM	40
Smith	Rule based approach	41
Ben Blamey	Distributed temporal semantic, probability density function	13
Tianyong Hao	Pattern learning and rule-based approach	15
Hartrumpf	Rule based approach	16
Mirza	Rule based approach	17
Mansouri et al	Rule based approach	18
Haffar	Rule based approach	19
Li and Wong	Rule based approach	21
Kim and Myaeng	Rule based approach, Finite state automata	22
Kamila	Probabilistic expansion	25
Ahmed	Lexicon expansion method	26
Haffar	Neural network architecture	27
Ahmed et al	Semi-supervised learning	34

Table 3. Temporal expression recognition systems for different languages and corresponding subject systems

Citation	Language	Subject system/Resource
1	Thai	300 news reports, linguistic books, ORCHID corpus
2	English and Hindi in transliterated form	Social media text (Facebook, Twitter, and WhatsApp) Labels- ICON(2016), MSIR(2015), MSIR (2016)
3	Chinese	n/a
4	Chinese	n/a
5	Spanish	199,835 EHRs of 749 lung cancer patients
6	Spanish	Four daily Mexican newspapers (from 1998-2002) that publish online containing sections as culture, politics, economy, sport etc.
7	Spanish	Online newspapers
8	Portuguese and English	n/a
9.	English	Training data contains 862 documents and test data contains 192 documents. (News articles)
10	Chinese	Xinhua News Agency (Training data contains 44 documents and test contain 8 documents)
11	English	n/a
12	Portuguese	The Second HAREM Collection (2ndHC) the text includes journalistic, FAQ, Blog, literary.
23	Hindi	ILTIMEX2012 corpus
24	English	Telephone dialog corpus, 'Enthusiast' corpus
25	English	English news (50K words)

26	English	ACE 2007 Corpus
27	English	n/a
28	English	TempEval-3 silver corpus
29	English	TERN 2004 data set
30	English	n/a
31	English	Online social network
32	English	Internet search terms gathered using AOL search engine
33	English	400 text queries uploaded to web-based system for Columbia University Medical Center Clinical Data Warehouse
34	German	N/A
35	Indonesia	75 news articles from www.kompas.com
36	Persian	Hamshahri, a news dataset (1996 to 2006) that includes politics, sports and entertainment.
37	Arabic	N/A
38	Hindi	FIRE 2011 Hindi corpus (Articles of news from the Hindi daily Navbharat Times)
39	Chinese	Financial news on the web
40	Korean	News articles
41	English	3,500 news documents from Google News.

Table 4. Temporal expression recognition systems identified for different languages

Sr. No.	System Name	Language
1	TimeTrails	English
2	Chronos	English, Italian
3	TempEx	English
4	GUTime	English
5	DANTE	English
6	TERSEO	English, Spanish
7	TimexTag	English
8	TIPSem	English, Spanish
9	HeidelTime	English, German, Dutch
10	MedTime	Clinical narratives
11	ParsTime	Persian
12	TICS	Chinese
13	InSicht	German

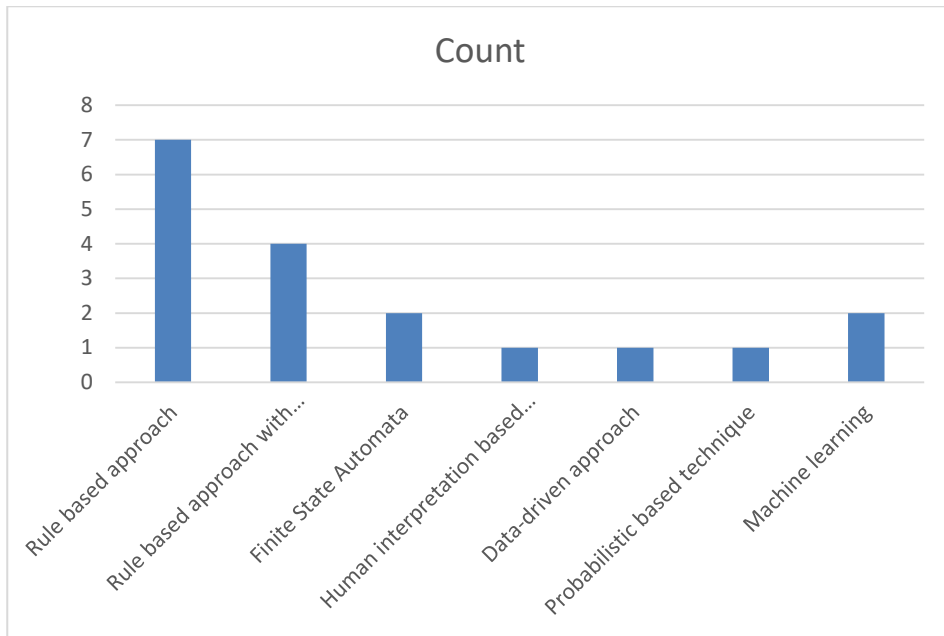


Figure 8. Approaches of temporal expression recognition systems.

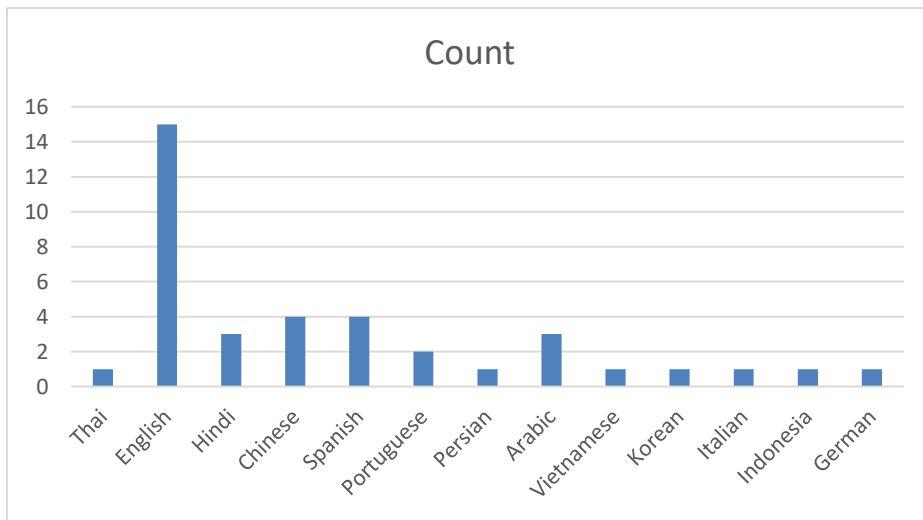


Figure 9. Temporal expression recognition systems developed for different languages.

4.2 Various machine learning approaches used for developing different temporal expressions detection systems.

Machine Learning is a branch of AI that enables computers to generate knowledge from raw data. It consists of statistical and mathematical algorithms to train computers to do defined tasks. Two important components of machine learning are training and statistical tools. In training module, the system is trained to perform specific task. Computer works on probabilistic math for performing the tasks. Table 5 depicts different Machine Learning algorithms used in the selected studies.

Shekhar and Sharma [2] discussed ruled based approach for identifying the temporal expressions in mixed social media text. Other statistical approaches like Conditional random field and Support Vector Machine are used. By using output of these techniques, a voting technique is used which chooses the optimum tag for the term on the basis of maximum vote. When comparing with standard classifiers to

identify temporal words, the generated HLSTM model outperforms them. Normalizing transliteration variances of the term is discussed as one of the key challenges. The results show F-score of 57.7 for English text and 20.4 for Hindi text.

Yin and Jin [3] presented character embedding features approach that utilizes distributed representation of words and a neural network-based multi-label classification method for Chinese language. Temporal expressions were divided into seven classes.

Wu et al [4] defined temporal expression normalization as retrieval of underlying temporal meaning from the given expression as well as depiction of the meaning as per the standard. The authors proposed an approach to temporal expression normalization that combined rule-driven mapping and machine learning techniques (ML). Two models, referred to as independent binary classification (IBC) model along with compared binary classification model (CBC) model are assessed, contrasted, and analyzed. 16 categories are used to classify Chinese temporal expressions. The features for categorizing temporal expressions are discussed.

Kolomiyets [9] discussed two methods discussed rule-based along with machine-learning approach as two approaches for solving temporal expression recognition problem. Machine learning approaches are categorized into token-by-token classifiers and binary constituent-based classifiers (Chunk based). The results depict that the classification based on constituents performs better than token-based classification.

Ramrakhiyani and Majumder [20] developed ILTIMEX2012 corpus. The approaches used are rule based approach, CRF, SVM, Voting based approach and another Voting-based approach is used to select the most suitable class from the three approaches' outputs for every token. The system achieves 0.88 F1 score. Various challenges for temporal expression identification for Hindi language are also discussed.

Kamila et al [25] discussed a semi-supervised learning approach in which every Hindi WordNet synset is categorized into one of five groups: present, past, future, atemporal and neutral. Starting with a limited number of seed instances, the algorithm loops up until accuracy on gold test data set fails to improve in four successive repetitions and cross-validation accuracy decreases. Three learning algorithms SVM, Naive Bayes, and decision tree have been employed.

Table 5. Machine learning techniques identified

Machine Learning technique	Citation	Category
Conditional Random field	2, 10, 23,27	Semi-supervised
Support vector machine	2	Supervised
LSTM	2	Semi-Supervised
Logistic regression	2	Supervised
Random forest	2	Supervised
KNN	2	Supervised
Neural networks	3	Supervised
SVM	23,27	Supervised
LIBSVM	4,29	Supervised
Multilayer perceptron (MLP)	8	Supervised
Linear regression	8	Supervised
Maximum Entropy Markov Model	9	Supervised
k-means clustering	13	Semi-Supervised
Decision tree	25	Semi-Supervised

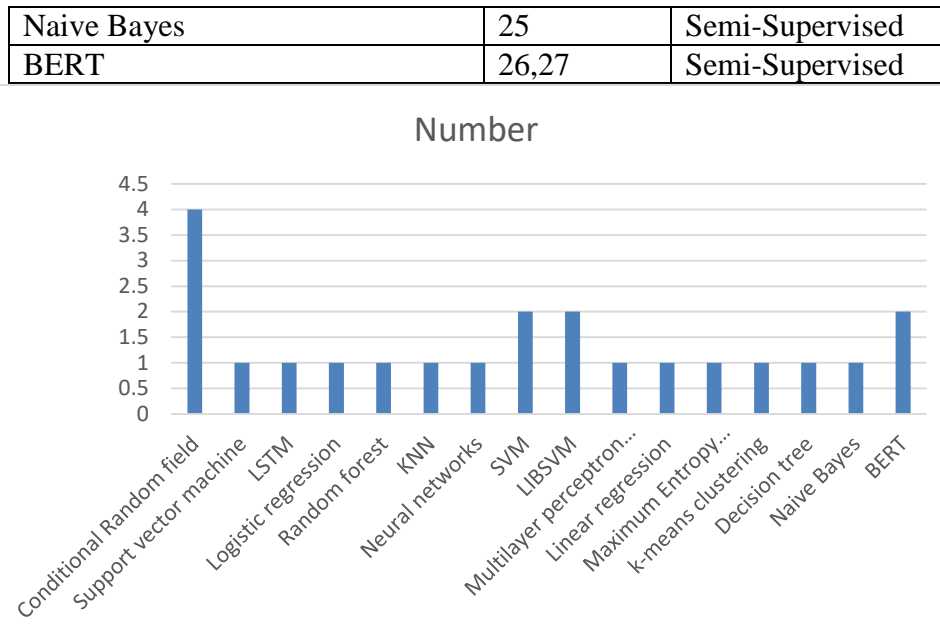


Figure 10. Number of machine learning approaches used.

4.3. Time in documents

A set of documents contains a lot of temporal information. Different categories of temporal information are described in the section 3.1. Section 3.2 describe how the temporal expressions in natural language can be realized.

4.3.1 Types of Temporal Information.

Temporal expressions can be categorized into four types according to TimeML, which is a standard markup language for temporal information: date, time, duration, and set [21].

Date and time expressions represent particular point in time that is specific date and time respectively. Duration expressions give information about the duration of interval. Set expressions provide details about an event's periodic nature.

4.3.2 The possible temporal expressions and their occurrence ways.

There are various ways in which temporal expressions can be expressed. They are as follows:

Explicit temporal expressions represent a particular point in time. The term itself provides all of the necessary information for normalization and is thus completely described. Relative temporal expressions cannot be normalized without considering some contextual information. They need some context for complete information. Implicit temporal expressions represent some events [4].

4.3.3 Temporal Tagging

The goals of temporal tagging are extracting and normalizing the temporal expressions to standard format. TIMEX2 and TIMEX3 are commonly used standard formats. TIMEX2 tags comprises pre as well as post modifiers of temporal expressions. This standard also allows the nested form of temporal expressions.

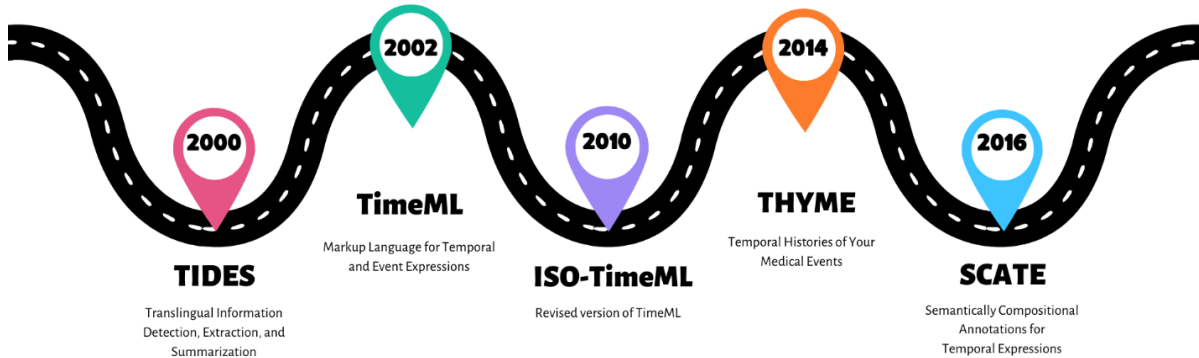


Figure 11. Temporal annotation standards.

4.3.5 TIDES

It provides guidelines for normalized representation annotation of temporal expressions, assisting with research projects funded by the ACE and TIDES programs of Defense Advanced Research Projects Agency (DARPA). The TIDES Temporal Corpus was created by preparing annotated corpora in both Spanish and English via manual annotations and translations. Program areas of TIDES are; Document understanding, text retrieving, topic detection as well as tracking. Time expression annotation standard TIMEX2 was created under TIDES program. Text can be marked with temporal expressions using just one XML tag, TIMEX2 [32].

4.3.6 TIME ML (Time Markup Language)

Understanding the temporal nature of a text requires an understanding of the relationships between temporal events and expressions which are crucial components that were not addressed in the previous schema are addressed in the new schema. Three unique aspects of temporal markup captured by TimeML are; it consistently links event predicates to a wide variety of temporally denoting expressions; it arranges textual event expressions in relation to the others, the both discourse as well as intrasentential; it permits underspecified interpretation of temporal expressions that are only partially determined [30].

4.3.7 ISO-TimeML

ISO-TimeML is an updated and compatible version of TimeML. Three parts make up the ISO-TimeML specification: 1. the abstract syntax for ISO-TimeML annotation; 2. a concrete syntax format for these annotations in XML; and 3. An ISO-TimeML semantics. The main modification was the switch from in-line annotations to stand-off annotations, which do not modify the text that is being processed. A new MLINK tag with the type measure corresponding to durations is also introduced by the ISO-TimeML standard [31].

4.3.8 THYME

Temporal Histories of Your Medical Events project intends to generate manual annotated corpus regarding medical data that includes explicit annotations of temporal relationships between various dates, events, occurrences, and procedures [29].

4.3.9 SCATE

Semantically Compositional Annotations for Temporal Expression. This annotation scheme improves the limitations of TimeML standard. It allows the representation of time expressions that correspond to multiple calendar units, are defined in terms of events, and are the intersection/ unions of time expressions [28].

Table 6. Comparison of the existing review articles

Title	Contribution	Summary	Results
Temporal Reasoning in Natural Language Processing: A Survey	<ul style="list-style-type: none"> • An overview of various approaches to temporal reasoning in NLP. • Annotation strategies for textual time relations and event identification. • Discussion of tractable classes and temporal reasoning frameworks. 	<ul style="list-style-type: none"> • The study examines temporal reasoning in NLP tasks related to natural language processing. • Focuses on temporal relations, event extraction, and real-world NLP applications. • Describes Temporal reasoning frameworks, annotation schemas, and their applications in NLP. 	<ul style="list-style-type: none"> • Review of temporal models for natural language processing (NLP) uses such as question answering and summarization. • Schemes for reasoning frameworks, temporal relations, and event identification annotations. • Models with constraints for temporal reasoning in natural language processing.
A literature survey on temporal information processing	<ul style="list-style-type: none"> • Extending natural language processing systems to capture language's temporal dimension. • Enhancing NLP systems' capacity for tasks involving the extraction of temporal information. 	<ul style="list-style-type: none"> • Discuss temporal Information Processing for Natural Language Processing tasks such as Question Answering. • Emphasize on extraction and retrieval of temporal information from text. 	<ul style="list-style-type: none"> • A summary of the literature on data extraction and retrieval using temporal information processing. • Approaches to utilize NLP systems to find temporal words in the text. • Techniques for connecting modern search terms with their historical equivalents. • Future event summaries are automatically generated from news archives.
Time expression recognition and normalization: a survey	<ul style="list-style-type: none"> • A summary of TERN development across multiple languages and fields. • TERN system evolution: from rule-driven to 	<ul style="list-style-type: none"> • An overview of the development of Time Expression Recognition and Normalization (TERN). 	<ul style="list-style-type: none"> • Examined the rule-based, machine-learning, and deep-learning TERN development approaches.

	<p>methods for deep learning.</p> <ul style="list-style-type: none"> Investigating unsupervised as well as semi-supervised learning strategies for TERN. 	<ul style="list-style-type: none"> Emphasise on TERN's future research, review, evaluation and methodology. 	<ul style="list-style-type: none"> Evaluated methods for TERN systems Review of TERN's methodologies, languages, domains, and assessment procedures.
Temporal Reasoning in Natural Language Processing: A Survey	<ul style="list-style-type: none"> An overview of temporal reasoning in NLP is presented, including discussion of tractable classes, annotation schemes, and temporal reasoning frameworks. 	<ul style="list-style-type: none"> The study examines temporal reasoning in NLP tasks related to natural language processing. Focuses on temporal relations, event extraction, and real-world NLP applications. Explains temporal reasoning frameworks, annotation schemas, and their uses. 	<ul style="list-style-type: none"> Temporal model survey with appropriate mechanisms for NLP applications. Annotation systems for text that represent temporal information. Frameworks with constraints for reasoning and processing temporal data. <p>Different TIMEX2 tag attributes to identify time-related data.</p>

5. Discussion and Conclusion

Temporal expressions, which incorporate temporal information in documents, present an intriguing way to improve the functionality of existing information retrieval systems.

We found 41 papers for this review in a variety of reliable journals, conferences, and other sources. We presented the results in a number of areas, including, Temporal expression identification techniques, machine learning techniques, and temporal annotation utilized to carry out various tasks. Among the approaches identified in the selected studies, rule-based along with machine learning technique is used in most of the studies followed by rule-based approach along with the machine learning. Results of the studies show that the temporal expression recognition systems are developed for various languages. The majority studies have been conducted for English language. To the best possible extent, our survey has addressed the temporal expression recognition systems developed and presented relevant digital framework. However, comparing different approaches was challenging due to varied languages, insufficient uniformity in annotating the dataset, size of the dataset. It was challenging to compare models fairly due to this lack of standardization, which highlights the need for more research using consistent datasets and assessment techniques.

References

1. T. Vichayakitti and C. Jaruskulchai, Automatic temporal event recognition from Thai news, in: proceeding of IEEE International Symposium on Communications and Information

- Technology, ISCIT, Beijing, China, 2005 pp. 938-942, doi: 10.1109/ISCIT.2005.1567021.
2. S. Shekhar and D. K. Sharma, Computational Intelligence for Temporal Expression Retrieval in Code-Mixed Text, in: proceedings of 2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC), Mathura, India, pp. 386-390, 2020 doi: 10.1109/PARC49193.2020.236634.
 3. B. Yin and B. Jin, A Multi-Label Classification Method on Chinese Temporal Expressions Based on Character Embedding, in: Proceedings of 4th International Conference on Information Science and Control Engineering (ICISCE), Changsha, China, pp. 51-54, 2017. doi: 10.1109/ICISCE.2017.21.
 4. Mingli Wu, Wenjie Li, Qing Chen and Qin Lu, Normalizing Chinese temporal expressions with multi-label classification, in: Proceedings of International Conference on Natural Language Processing and Knowledge Engineering, Wuhan, China, pp. 318-323, 2005. doi: 10.1109/NLPKE.2005.1598755.
 5. M. Najafabadipour et al., Recognition of Time Expressions in Spanish Electronic Health Records, in: Proceedings of 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS), Cordoba, Spain, pp. 69-74, 2019. doi: 10.1109/CBMS.2019.00025.
 6. Galicia-Haro, Sofia, Spanish Temporal Expressions: Some Forms Reinforced by an Adverb, in: Proceedings of the 7th Mexican International Conference on Artificial Intelligence: Advances in Artificial Intelligence, pp.193-203, 2008 10.1007/978-3-540-88636-5_18.
 7. P. Martinez-Barco, E. Saquete and R. Muñoz, A Grammar-Based System to Solve Temporal Expressions in Spanish Texts, in: Proceedings of International Conference for Natural Language Processing in Portugal, 2002.
 8. H. Tissot et al., Normalisation of imprecise temporal expressions extracted from text, Knowledge and Information Systems, 61(3), pp.1361-1394, 2019. <https://doi.org/10.1007/s10115-019-01338-1>.
 9. O. Kolomiyets, M.F. Moens, Comparing Two Approaches for the Recognition of Temporal Expressions, in: Proceedings of Conference KI 2009 Advances in Artificial Intelligence, 32nd Annual German Conference on AI, Paderborn, Germany, pp.225-232 2009. https://doi.org/10.1007/978-3-642-04617-9_29
 10. H. Liorens et al., Data-driven approach based on semantic roles for recognizing temporal expressions and events in Chinese, in: Proceedings of the 16th international conference on Natural language processing and information systems, pp.88-99, 2011.10.1007/978-3-642-22327-3_9.
 11. R. Dale, P.P.Mazur, The Semantic Representation of Temporal Expressions in Text, in: Proceedings of Advances in Artificial Intelligence, 20th Australian Joint Conference on Artificial Intelligence, Gold Coast, Australia, pp.435-444, 2007, 10.1007/978-3-540-76928-6_45.
 12. O. Craveiro, J.H.Macedo, H. Madeira, Use of Co-occurrences for Temporal Expressions Annotation, in: Proceedings of String Processing and Information Retrieval, Finland, 2009, 10.1007/978-3-642-03784-9_15.
 13. B. Blamey, T. Crick, G. Oatley, 'The First Day of Summer': Parsing Temporal Expressions with Distributed Semantics, in: Proceedings of SGAI conferences, pp. 389-402, 2013. 10.1007/978-3-319-02621-3_29.
 14. S. Nunes, C. Ribeiro, G. David, Use of Temporal Expressions in Web Search, in: Proceedings of advances in Information Retrieval, 30th European Conference on IR Research UK, 2008.

- 10.1007/978-3-540-78646-7_59.
15. T. Hao, A. Rusanov, C.Weng, Extracting and Normalizing Temporal Expressions in Clinical Data Requests from Researchers, in: Proceedings of the 2013 international conference on Smart Health, 2013, pp. 41-51. 10.1007/978-3-642-39844-5_7.
 16. S. Hartrumpf, J. Leveling, Interpretation and Normalization of Temporal Expressions for Question Answering, in: Proceedings of Evaluation of Multilingual and Multi-modal Information Retrieval, 7th Workshop of the Cross-Language Evaluation Forum Spain, 2006.10.1007/978-3-540-74999-8_51.
 17. P. Mirza, Paramita, Recognizing and Normalizing Temporal Expressions in Indonesian Texts, PACLING 2015. Communications in Computer and Information Science, vol 593, Springer, Singapore. https://doi.org/10.1007/978-981-10-0515-2_1010/.
 18. B. Mansouri, M.S. Zahedi, R. Campos, M. Farhoodi, ParsTime: Rule-Based Extraction and Normalization of Persian Temporal Expressions, Advances in information retrieval, ECIR, pp.715-721, 2018. 10.1007/978-3-319-76941-7_67.
 19. N. Haffar, E. Hkiri, M. Zrigui, TimeML Annotation of Events and Temporal Expressions in Arabic Texts, in: Proceedings of International Conference on Computational Collective Intelligence, pp.207-2018, 2019.10.1007/978-3-030-28377-3_17.
 20. N. Ramrakhiyani, P. Majumder, Approaches to Temporal Expression Recognition in Hindi, ACM Transactions on Asian and Low-Resource Language Information Processing, 14(1), pp. 1–22, 2015.
 21. W.Li, K.F. Wong, A Word-Based Approach for Modeling and Discovering Temporal Relations Embedded in Chinese Sentences, CM Transactions on Asian Language Information Processing 1(3), pp.173-206, 2002. 10.1145/772755.772756.
 22. P. Kim, S.H.Myaeng, Usefulness of temporal information automatically extracted from news articles for topic tracking, ACM Transactions on Asian Language Information Processing, 3(4), pp.227-242, 2004. 10.1145/ 1039621.1039624.
 23. Mani, G. Wilson, B. Sundheim, and L. Ferro, A multilingual approach to annotating and extracting temporal information, in: Proceedings of the Workshop on Temporal and Spatial Information Processing, Association for Computational Linguistics, 2001.
 24. J. Strotgen et al., Time for More Languages: Temporal Tagging of Arabic, Italian, Spanish, and Vietnamese, ACM Transactions on Asian Language Information Processing, 13(1), pp.1-21, 2014. 13. 10.1145/2540989
 25. S. Kamila, M. Hasanuzzaman, A. Ekbal, P. Bhattacharya, Tempo-HindiWordNet: A Lexical Knowledge-base for Temporal Information Processing, ACM Transactions on Asian and Low-Resource Language Information Processing 18(2), pp.1-22, 2018.1-22. 10.1145/3277504.
 26. U. AHmed, J.C.CW.lin, G.Srivastava, U.Yun, Semi-Supervised Lexicon-Aware Embedding for News Article Time Estimation, ACM Transactions on Asian and Low-Resource Language Information Processing, 2023.10.1145/ 3592604.
 27. N.Haffar, M.Zrigui, A Synergistic Bidirectional LSTM and N-gram Multi-channel CNN Approach Based on BERT and FastText for Arabic Event Identification, ACM Transactions on Asian and Low-Resource Language Information ProcessingVol, 22(11), pp.1-27, 2023, 10.1145/3626568.
 28. S. Bethard, J. Parker, A Semantically Compositional Annotation Scheme for Time Normalization, in: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), pp. 3779-3786, 2016.
 29. THYME (Temporal Histories of Your Medical Events) Temporal Relations Annotation

Guidelines (colorado.edu)

30. J.Pustejovsky, J.M.Castano, R.Ingria, R.Sauri, TimeML: Robust Specification of Event and Temporal Expressions in Text, in: Proceedings of New Directions in Question Answering, Papers from 2003 AAAI Spring Symposium, Stanford University, Stanford, USA,2003.
31. L. Ferro, I. Mani, B. Sundheim, G. Wilson, TIDES Temporal Annotation Guidelines Version 1.0.2 (2001).
32. R. Schwitter, Augmenting an Answer Set Based Controlled Natural Language with Temporal Expressions, in: Proceedings of PRICAI 2019: Trends in Artificial Intelligence, 16th Pacific Rim International Conference on Artificial Intelligence, Cuvu, Yanuca Island, Fiji, pp. 500-513, 2019.
33. O. Alonso,J. Strotgen,R.B. Yates, M. Gertz,Temporal Information Retrieval: Challenges and Opportunities, TAWAW 2011.
34. U. Ahmed, J.C.L. Lin, G. Srivastava, U. Yun, Semi-Supervised Lexicon-Aware Embedding for News Article Time Estimation. ACM Transactions on Asian and Low Resource Language Information Processing, 2023.
35. D. Rattan, R. Bhatia, M. Singh, Software clone detection: A systematic review, Inf. Softw. Technol. 55 (7), pp.1165–1199, 2013.
36. A.K. Peterson, How to do a Structured Literature Review in computer science, Technical Report, Alexandra Instituttet, Copenhagen, Denmark, 2015
37. P. Mazur, Temporal expression tagger, in: Proceedings of 23 Human Language Technology Challenges of the Information Society, Springer, pp. 245–257, 2009
38. M. Filannino , G. Nenadic, Temporal expression extraction with extensive feature type selection and a posteriori label adjustment, Data & Knowledge Engineering, Elsevier, pp. 19 33, 2015..
39. C. Hagège, J. Baptista, N. Mamede, Portuguese Temporal Expressions Recognition: From TE Characterization to an Effective TER Module Implementation, in: proceedings of Seventh Brazilian Symposium in Information and Human Language Technology, 2009.
40. D. Ahn, J. Rantwijk, M. Rijke, A cascaded machine learning approach to interpreting temporal expressions, in: Proceedings of HLT-NAACL, pp. 420–427, 2007.
41. C.S. Smith, The Syntax and Interpretation of Temporal Expressions in English, Linguistics and Philosophy, Springer, 2(1), 2009, pp. 43-99.
42. Negri, Matteo & Marseglia, Luca, Recognition and Normalization of Time Expressions: ITC-irst at TERN 2004.
43. B. A. Kitchenham, S. Charters, Guidelines for performing systematic literature reviews in software engineering. Technical Report EBSE-2007-01, School of Computer Science and Mathematics, Keele University, Keele and Department of Computer Science, University of Durham, Durham, UK, 2007, pp. 65.
44. Brereton, B.A. Kitchenham, D. Budgen, M. Turner, M. Khalil, Lessons from applying the systematic literature review process within the software engineering domain, The Journal of Systems and Software 80 (4), 2007, pp. 571–583.