

TEMPORAL INFORMATION PROCESSING: A SURVEY

Ines Berrazega

¹LARODEC Laboratory, ISG of Tunis B.P.1088, 2000 Le Bardo, Tunisia
Ines_berrazega@yahoo.fr

ABSTRACT

Temporal Information Processing is a subfield of Natural Language Processing, valuable in many tasks like Question Answering and Summarization. Temporal Information Processing is broadened, ranging from classical theories of time and language to current computational approaches for Temporal Information Extraction. This later trend consists on the automatic extraction of events and temporal expressions. Such issues have attracted great attention especially with the development of annotated corpora and annotations schemes mainly TimeBank and TimeML. In this paper, we give a survey of Temporal Information Extraction from Natural Language texts.

KEYWORDS

Temporal Information Processing, Natural Language Processing, Information Extraction, annotation schemes, TimeML.

1. INTRODUCTION

Information Extraction is gaining increased attention by researchers who seek to acquire knowledge from huge amount of natural language contents. Many approaches have been proposed to extract valuable information from text. However, dynamic facts as temporal information have been neglected, although time is a crucial dimension in any information space [1]. This limitation can be explained by the complexity of such task. For example, the classical techniques used to extract named entities and events from textual contents are unable to identify the temporal relations between events or to infer the chronological ordering of these events. Such processes require a grater effort to analyze how temporal information is conveyed in real texts, especially when temporal information is implicitly expressed.

In this concern, recent researchers have aimed to expand the capabilities of existing Natural Language Processing (NLP) systems to account for the temporal dimension of language. Thus the Temporal Information Extraction tasks first appeared in the scope of MUC-5 when it was asked to assign a calendrical time to a joint venture event [2].

In fact, extracting temporal information from texts is valuable in many NLP tasks like Question Answering, Summarization, and Information Extraction. It is also gainful for many real-world applications like Web Search, Medical-Records, Legal Reasoning, Accounting, Banking, Reservation Systems, and Accident Reports.

Motivated by the expected applications of Temporal Information Extraction in various domains, several computational approaches have been proposed. They are handling two main issues:

automatic recognition of temporal entities in narrative (namely temporal expressions and events), and temporal relations identification between these entities. These tasks are considered quite complex; tackling them cannot be limited to simple pattern recognition. It requires a deep comprehension and studying of natural language contents, especially that temporal information can be expressed in implicit way and some expressions can be ambiguous.

The main purpose of this paper is to present a survey of the Temporal Information Extraction domain. The rest of this paper is organized as follows. We first present a brief review of Information Extraction in section 2. Then, we state the main concepts related to Temporal Information in section 3. We review in section 4 the classical works on Temporal Information Processing. After that, we focus on the current issues on the Temporal Information Extraction domain in section 5. Then, we describe some real-world applications in section 6, followed by a conclusion in section 7.

2. INFORMATION EXTRACTION

Information Extraction (IE) is an important field of Natural Language Processing (NLP). Its aim is to automatically extract from unstructured and/or semi-structured contents predefined types of information such as entities, events or relationships.

Information Extraction dates back to the late 1970s. Later, the scope of IE was strongly influenced by two competitions namely the Message Understanding Conferences (MUC) [3] and the Automatic Content Extraction (ACE) program. In the scope of these competitions, several tasks have been discussed such as named entity recognition, co-reference resolution and event extraction.

It's worth distinguishing Information Extraction from other fields of NLP like Information Retrieval (IR). In the next subsection, we present a brief distinction between these two NLP tasks.

2.1. Information Retrieval vs. Information Extraction

Daily, huge amounts of unstructured data are produced everywhere. These contents need to be exploited and analyzed to acquire relevant information useful in different domains. This issue has been the topic of several fields of research such as Information Retrieval (IR) and Information Extraction (IE).

Information Retrieval (IR) involves searching and finding documents that answer the user's requirement. According to Gaizauskas and Wilks, Information Retrieval is based on document retrieval with the aim to answer the question "how to find, in a set of documents, those that interest me?" [4].

On the other side, Information Extraction (IE) aims to extract interesting information from documents for an automatic analysis by a computer. The extraction techniques have to deal with the understanding of the meaning of natural language.

Sarawagi defines Information Extraction as "the automatic extraction of structured information such as entities, relationships between entities and attributes describing entities from unstructured sources. This enables much richer forms of queries on the abundant unstructured sources than possible with keyword searches alone [5]."

While IR aims to recover from a collection of documents a subset that answers the user's query, the goal of IE is to extract relevant facts from these documents. These facts are then used in further applications.

Information Extraction is useful for many issues. The next subsection presents some of its real-world applications.

2.2. Real-World Applications

Various applications depending on information extraction have been developed especially with the advent of the Internet. Sarawagi lists four categories of applications: Enterprise, Personal, Scientific, and Web-oriented. Some of these applications are presented in the following paragraphs.

Customer Care

These applications are designed to collect unstructured data from customer interaction; and then to integrate it with the databases and the business ontologies of the enterprise. Sarawagi states many problems resolved with these applications: "the identification of product names and product attributes from customer emails, linking of customer emails to a specific transaction in a sales database, the extraction of merchant name and addresses from sales invoices, the extraction of repair records from insurance claim forms, the extraction of customer moods from phone conversation transcripts, and the extraction of product attribute value pairs from textual product descriptions."

Personal information management (PIM)

PIM systems are used to structure personal data like documents, emails, and projects in a structured inter-linked format. For example these systems are able to automatically extract from an email the location of people names and phone numbers.

Bio-informatics

Many medical applications aim to extract biological objects such as proteins and genes from paper repositories such as Pubmed [6] as well as their names and their interaction.

Community Websites

This category of websites tracks information about several topics like researchers, conferences, talks, projects, and events specific to a given community. These websites are used for the creation of structured databases like DBLife [7] based on various extraction operations: locating talk announcements from department pages, extracting names of speakers and titles from them, extracting structured records about a conference from a website, etc.

2.3. Approaches for Information Extraction

In the last two decades, several approaches have been proposed for Information Extraction. There are three families of approaches namely rule-based approaches, machine learning approaches and hybrid approaches.

Rule-Based approaches

Historically, rule-based systems for Information Extraction were first implemented. This family of systems is based on domain-specific extraction rules written by a domain expert. Several rule-based systems were developed to extract information namely AutoSlog by Riloff [8]; Fastus by Appelt et al, [9] and GATE by Cunningham et al, [10].

This family of methods requires a great human effort and a considerable time for data analysis and rule writing. Also rule-based methods lack portability to other domains.

To overcome the deficiencies of the classical rule-based approaches, human effort should be adequately replaced by a learning component. Thus, machine-learning approaches began to be exploited.

Machine-Learning approaches

These approaches have become the enabling technique for many problems in Natural Language Processing, among which are named entity recognition, parsing, tagging and semantic role labeling. For these purposes, several models were used: generative models based on Hidden Markov Models [11, 12, 13], conditional models based on maximum entropy [14, 15, 16] and global conditional models which are Conditional Random Fields [17].

Rule-based methods and machine-learning methods are both used depending on the nature of the extraction task. There is a third family of approaches consisting on hybrid models.

Hybrid approaches.

This third family of approaches leverages the benefits of both machine-learning and rule-based methods [18, 19, 20].

All reviewed approaches on Information Extraction were limited to study static facts of natural language to the detriment of dynamic ones, in spite of the interest of temporal dimension in many cases. For this reason, several researchers have aimed to expand the capabilities of existing Natural Language Processing systems to account for the temporal dimension of language. In the next section, we examine this new trend in Natural Language Processing.

3. BASIC CONCEPTS

When studying temporality, it is important to define basic concepts related to time. In this section we try to provide a clear description of temporal entities and relations in textual contents that enables us to gain useful insights into how temporal information is conveyed in written language.

3.1. Eventualities: Events and States

The first two main temporal entities in texts are *Events* and *States*. These two entities are grouped by Bach in 1986 under the term of *eventualities* [21].

Events are things that happen or occur in the world (like weddings, birthdays, and parties...) at a certain time or over a given period of time and in a given place. They are typically dynamic occurrences that have causes and effects, a clear beginning and end, and bring about some perceptible change in the world [22].

States, on the other hand, are considered as the existence of a set of properties over a given period (like happiness, owning a car...) often without a clearly defined beginning and end.

Bittar mentions that “the occurrence of an event may bring about the existence of a state. Similarly, an event may also bring about the end of a state. In other words, a change of state in the world typically indicates the occurrence of an event [23].”

The main difference between events and states is attributed to the *dynamic/static* distinction, where events are seen as dynamic and resulting in some change or other, while states, on the other hand, do not involve a perceptible.

3.2. Temporal Expression

A temporal expression can be defined as being any element in language that lexicalizes the concept of time in terms of recognized and generally quantifiable temporal units [23]. In accordance with Sauri and al, there are four categories of temporal expressions according to the types of instants or intervals they describe: dates, times and durations, sets [24].

Dates are expressions that refer to a particular period based on the Gregorian calendar. This includes units that are larger than a year, such as centuries and millennia, as well as subintervals of a typical year, such as seasons. The basic unit on which the calendar is based is the day. Dates can be expressed in an absolute form (eg, Tuesday 18th) or in a relative form (eg, last week).

Times are expressions which denote a particular subdivision of a day. They are understood to refer to particular moments that subdivide a day. Also, times represent a relatively simple category of temporal expressions and may correspond to the moments we measure on a clock, or express more general parts of a typical day (eg, in the morning/ at 9 a.m)

Durations are expressions which refer to an extended period of time, very often specifying the temporal extent of an eventuality. They are measured using calendrical units (years, months, days etc.) or clock units (hours, minutes, seconds etc.) of temporal measure (eg, 3 hours last Monday).

Sets are expressions that refer to the regularity or reoccurrence of an eventuality, either in the absolute, or relative to a period of time (eg, *twice a week*).

3.3. Temporal Relations

In natural language, temporal relations hold between temporal entities: between two events, between two temporal expressions or between an event and a temporal expression. Longacre defines a temporal relation as “an inter-propositional relation that communicates the simultaneity or ordering in time of events or states” [25]. On this basis, reasoning about time implies the temporal representation of these relations. In this concern, most of computational models for reasoning about time are based on the Allen’s Interval Algebra [26] to capture the temporal dimension of a narrative. More details about this algebra are presented in the subsection 4.2.

After having stated the main concepts related to time in language, the next section is dedicated to present the earlier work on Temporal Information Processing.

4. CLASSICAL WORK IN TEMPORAL INFORMATION PROCESSING

The scope of research in Temporal Information Processing is broadened, ranging from classical linguistic theories of tense and aspect to current computational approaches for Temporal Information Extraction.

Prior to describe current Temporal Information Processing issues, it is necessary to review the background studies on time and language. The following subsections review the most outstanding classical Temporal Information Processing works [27]. The first subsection presents the most prominent linguistic theories of tense and aspect. The second subsection reviews the temporal reasoning works. The third subsection states the works on the temporal structure of discourse. Finally, the fourth subsection shows the limitations of these classical works.

4.1. Linguistic theories of tense and aspect

Temporal information is conveyed in natural language through grammatical mechanisms to represent time and temporal relations. In this concern, tense and aspect are considered as the most important grammatical categories in language.

The most prominent works on linguistic theories are the Reichenbach's theory for tense called '*The Tenses of Verbs*' [28] and the Vendler's work [29] for aspect which has been the basis for subsequent researchers.

4.1.1. Tense

Tense is a temporal linguistic mechanism that expresses the time at which or during which an event takes place. In this purpose, Reichenbach [28] develops a theory in which tense gives information about the following times:

Speech Time (S): the time related to the time point of the speaking act.

Event Time (E): the time at which the told event happens.

This two times points let express the basic tense classes with use of the operators of precedence (<) and simultaneity (=).

The followings three sentences illustrate basic tense classes with speech and event time points.

- a. I played football: Past tense. ($E < S$)
- b. I play football: Present tense. ($E = S$)
- c. I will play football: Future tense. ($S < E$)

However, combinations between these two times points are unable to express all tenses. Consider the following examples:

- a. I played football.
- b. I had played football.

Even though the two sentences express events which occurred in the past ($E < S$), its incorrect to represent them in the same way. In example (I had played football) the event of playing seems to refer to another event. To handle this situation, Reichenbach introduce a third temporal point into

his model which is the **Reference Time (R)**. According to this extension, the Event Time (E) In example (I had played football) is the time at which I played, and the Reference Time (R) is between the Event Time (E) and the Speech Time (S) “(E <R < S)”.

With the three points defined by Reichenbach, it’s possible to represent all the tenses using a set of relations between these points. Nevertheless, Reichenbach’s relations still ambiguous, so Song and Cohen [30] develop an unambiguous set of relations including a new operator (>) and presenting the relations always in S-R-E order.

The following table illustrates the original Reichenbach’s relations, the unambiguous relations proposed by Song and Cohen, and their corresponding English tenses.

Table1. Tense Temporal Relations

Reichenbach relations	Unambiguous relations	English tense	example
E<R<S (Anterior Past)	S>R>E	Past Perfect	I had played
E=R<S (Simple Past)	S>R=E	Past Simple	I played
R<E<S* (Posterior Past)	S>R<E		[I expect] I would play
R<S=E*			
R<S<E*			
E<S=R (Anterior Present)	S=R>E	Present Perfect	I have played
S=R=E (Simple Present)	S=R=E	Present Simple	I play
S=R<E (Posterior Present)	S=R<E	Future Simple	I will play
S=E<R*			
E<S<R*			
S<E<R* (Anterior Future)	S<R>E	Future Perfect	I will have played
S<R=E (Simple Future)	S<R=E	Future Simple	I will play
S<R<E (Posterior Future)	S<R<E		I shall be going to play

* ambiguous relations

4.1.2. Aspect

Aspect is the second device expressing time in natural language. Two types of aspects are expressed in language namely grammatical aspect and lexical aspect. Grammatical aspect expresses the viewpoint from which a particular eventuality is described, it indicates the phase in which an eventuality is to be perceived. While lexical aspect, distinguishes between different subclasses of events based on its following temporal properties: *dynamicity*, *telicity* and *durativity*.

In the literature of Aspect, Vendler’s work [29] has been the basis for subsequent researchers. Vandler proposes an initial distinction between events and states and then classifies the event expressions into three aspectual categories or Aktionsarten: activities, accomplishments, and achievements.

Activities are events which are durative, or extended in time, but that do not involve an explicit end point. In other words, activities are *durative* and *atelic* events. In the other hand,

accomplishments (durative culminated process) are events that imply a duration with a definite end point in which a state changes. Finally, achievements (non-durative culminations) are punctual or instantaneous events that do not imply a duration, happening at a defined point in which a state changes.

4.2. Temporal Reasoning

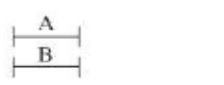
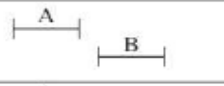
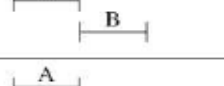
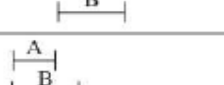
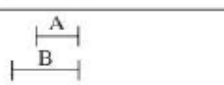


Temporal Reasoning involves the temporal representation of events and their temporal anchoring within natural language text. This topic has attracted great attention due to its potential applications in several NLP tasks like Summarization, Question Answering and so on.

The state-of art of Temporal Reasoning proposes several constraint based models. In a constraint based model, the temporal information is represented as a temporal constraint network (TCN) in which the events are denoted by nodes and the ordering constraints between events are denoted by edges. In this case, reasoning about time becomes a Temporal Constraint Satisfaction Problem (TCSP).

Different TCNs are defined depending on the representation of the temporal entity as time intervals, durations or points, and the class of constraints namely qualitative, quantitative, metrics or its combination [31]. Several researches fall into this framework [26] [32] [33].

James Allen developed an Interval Algebra which provides a conceptual model of time that captures the different ways in which eventualities may be related to each other. Allen considers that every temporal expression or event can be presented as a temporal interval having a start point and an end point on a timeline. In this concern, Allen defines a set of thirteen basic (binary) interval relations, where six are inverses of the other six, excluding equality. The following figure represents those intervals.

Table2. Allen's Temporal Interval Relations [26]

	A is EQUAL to B B is EQUAL to A
	A is BEFORE B B is AFTER A
	A MEETS B B is MET by A
	A OVERLAPS B B is OVERLAPPED by A
	A STARTS B B is STARTED by A
	A FINISHES B B is FINISHED by A
	A DURING B B CONTAINS A

In Allen's model, temporal relations between the intervals are maintained in a network, where the nodes represent individual intervals and the arcs represent the relationship between them.

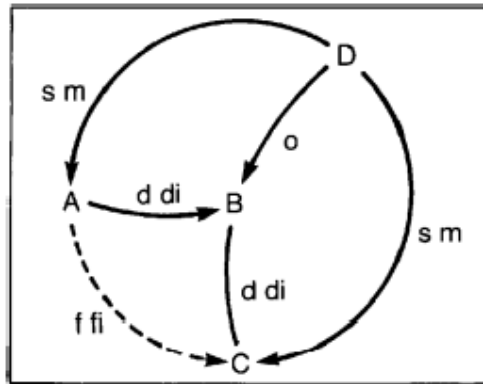


Figure1. Allen's interval-relation network [26]

Filatova and Hovy [32] also propose a model for temporal relations representation known as *the time stamping of events*. This model serves for arranging the contents of news stories into a timeline. This procedure consists on assigning a calendrical time point or interval to all events in a text. However, this model doesn't capture information in many cases and sometimes loses information or misinterpret.

Setzer and Gaizauskas [33] propose a simpler representation able to capture more information. It's a *Time-Event Graph Representation* used for identifying event-event relationship with event-time relationship. This representation is based on the Allen algebra. In fact, after reducing all events and temporal expressions to intervals and after identifying the temporal relations between them, the temporal information in a text can be represented as a graph where events and temporal expressions form the nodes, and the edges are labeled with the temporal relations between them.

4.3. Temporal structure of discourse

The context of a discourse is required to reach the correct language comprehension. In fact, sentences are usually interpreted in context instead of isolation. However, the context in which a sentence is uttered and the situation of the speakers, is important.

In this concern, Kamp proposes an influential work on Discourse Representation Theory (DRT) [34]. This work introduced a new level of mental representations, known as *the discourse representation structures (DRSs)*. In fact, authors consider that a hearer of a given discourse builds up a mental representation for each sentence of this discourse in a cumulative way. This theory gives the possibility to interpret sentences in the context of a discourse, rather than sentences in isolation using a construction procedure extending a given DRS. Other remarkable work on the temporal structure of discourse is the Dowty's *temporal discourse interpretation principle* (TDIP) [35]. Another well known work on temporal structure of discourse is Webber's theory of *anaphoric reference* [36]. In this work, Webber develops an account for noun-phrase anaphora and shows how tensed clauses can be treated as anaphors like noun phrases. This work is also based on the Reichenbach's theory for tense. We can also mention the Hwang and Schubert's *Tense Trees* [37]. Authors tried to consider the compositional semantics of complex sentences. For this purpose, they automatically convert the logical form of sentences into a temporal structure as a tense tree. Other works on temporal structure of discourse are interesting

namely the approach proposed by ter Meulen [38] about *the dynamic aspect trees* (DATs), the Reference Point (Rpt) work presented by Kamp and Reyle [39] and the *temporal conceptual graphs* proposed by Moulin [40].

4.4. Discussion

Even though the presented works highlighted a variety of issues and proposed many solutions for understanding and representing temporal information, they were criticized because they were not analyzed and evaluated over real linguistic data.

In this concern, it's worth noting that Temporal Information Processing and Natural Language Processing in general have undergone a mutation from the rationalist strategy based on such formal theories of language analysis to an empiricist strategy based on the analysis of real language use (i.e., textual corpora) [42]. This mutation has been accompanied by the development of computational semantic models and the establishment of evaluation frameworks to annotate temporal information in a corpus based vision. The next section highlights this new trend.

5. CURRENT ISSUES IN TEMPORAL INFORMATION EXTRACTION

As mentioned in the previous section, Temporal Annotation has become in the last decade the most prominent issue in the Temporal Information Processing area, especially with the development of annotation schemes and annotated corpora. With this new trend, researches are focusing on two tasks: in one hand automatically recognizing and extracting temporal entities in narrative (namely temporal expressions and events), and in the other hand on discovering temporal relations between these entities and inferring the type of each recognized one.

Temporal entity recognition

This task consists on identifying and extracting temporal entities from natural language texts. For events, this means finding which textual entity constitutes an event. For example, in the sentence "*she bought 15% of the shares*", Filatova and Hovy [32] consider that the entire clause represents an event. While for Pustejovsky and al, the appropriate span is just the verb group or just the head of the verb group ("*bought*" in this example) [43]. Furthermore, event recognition involves the identification of some attributes related to each event. Such attributes depend on the topic being studied. Pustejovsky and al define five attributes in their TimeBank news corpus [44] (tense, aspect, modality, polarity and class). Several approaches were proposed to identify events in natural language contents. Faiz propose an approach to identify relevant sentences in news articles for event information extraction [45]. Later, ElKhelifi and Faiz propose a new automatic approach able to annotate events in News texts [46].

For temporal expressions, entity recognition amounts to distinguish its different types (times, dates, durations, frequencies) and to find its corresponding values.

Temporal relations identification

Besides temporal entities, natural language texts contain other devices expressing logical relations between times and events or between events and events. The temporal relation identification task aims to recognize such relations and to infer the type of each recognized one.

Several researches fall into this task. [26, 32, 33].

Annotation schemes and annotated corpora

Handling temporal annotation requires the development of annotations schemes and the construction of annotated corpora. In this concern, a lot of ongoing researches are focusing on the development of annotations schemes to extract, model and interpret temporal information in natural language texts [42].

The most prominent temporal annotation schemes are: MUC-TIMEX proposed by Grishman and Sundheim in 1996 [47], TIDES proposed by Ferro and al in 2001 [48], STAG developed by Setzer and Gaizauskas in the same year [49] and finally the TimeML annotation scheme developed by Pustejovsky and al in 2003 [43]. All of them follow a SGML/XMLbased annotation format.

TimeML (Time Markup Language) is the latest annotation scheme. It was developed under the sponsorship of ARDA as the natural evolution of STAG. This new scheme is considered to be a combination and an extension of preceding schemes, which makes it the most complete and potent one. TimeML has recently been standardized to an ISO international standard for temporal information markup, ISO-TimeML (ISO-TimeML, 2007). Both the TimeML and the ISO-TimeML annotation standards define the following basic XML tags: <EVENT> for the annotation of events, <TIMEX3> for the annotation of time expressions, <SIGNAL> for locating textual elements that indicate a temporal relation, and the tags <TLINK>, <SLINK> and <ALINK> that capture different types of relations.

Regarding the construction of corpora, a corpus is seen as a collection of natural language data contents organized according to a set of criteria with the aim of conducting specific research. These contents are often enriched with descriptive or analytic annotations, in order to explain a particular linguistic phenomenon.

In terms of temporally annotated corpora, several corpora have been developed within the framework of the TimeML annotation scheme. TimeBank 1.1, TimeBank 1.2 and the AQUAINT corpora are all made up of journalistic texts in English and annotated with the TimeML annotation language.

TimeBank 1.2 [44] has become somewhat of a reference for the study of temporal information within the Computational Linguistics community. It's the last major TimeML annotated corpus and it has become the main reference for temporal annotation in English. It's also used as the basis for the TempEval evaluation campaigns' corpora.

After getting insights on current issues of Temporal Information Extraction, it's worth exposing some real world applications.

6. REAL-WORLD APPLICATIONS

Extracting Temporal Information from texts is a key problem for many real-world applications like Web Search, Medical-Records, Legal Reasoning, Accounting, Banking, Reservation Systems, and Accident Reports. Examples from these applications are presented in the next paragraphs.

Web Search

Query log analysis is currently an active topic of research. In fact, there is significant and growing number of contributions to the understanding of online user behavior. In this concern, temporal information has received little attention by researches who want to understand and characterize users' behavior online. Researchers try to identify and characterize the use of temporal expressions in web search queries. Nunes and al, led a query log analysis to investigate

the use of temporal expressions in web search queries [50]. They find that temporal expressions are rarely used (1.5% of queries) and, when used, they are related to current and past events. Also, there are specific topics where the use of temporal expressions is more visible

Medical Domain

Temporal Information Systems are lately used in several medically-related problems and tasks such as *diagnosis*, *monitoring* and *longitudinal guideline-based care*. Several applications in the medical domain are used to extract information about times of clinical investigations (X-rays, ultrasounds, etc.). In this concern, Angus and al, propose an algorithm to extract temporal relations between temporal expressions and clinical investigation events form clinical narratives [51].

Legal Reasoning

In legal domain, extracting temporal information has a great utility to facilitate the work of lawyers. Several applications have been proposed namely *the automatic temporal ordering of legal documents* in a time line, or *the automatic retrieval of a specific event* mentioned in various legal documents according to temporal constraints that may be associated with this event. Schilder presents a prototype system that extracts events from the United States Code on U.S. immigration nationality and links these events to temporal constraints [52].

Financial Accounting

Many applications in the accountancy field have used temporal information to solve several issues. Fisher presents a prototype system to support *the temporal reconstruction of financial accounting standards* (FASs) [53]. This prototype enables a user to specify an FAS along with a date, and permits a dynamic and continuing codification of a particular FAS.

7. CONCLUSION

Temporal Information Extraction is a recent area gaining more interest among researchers. Our literature review reveals the existence of two trends in the domain. In one hand, classical Temporal Information Processing works are arising from the intersection of linguistics, philosophy and symbolic AI. In the other hand, the current works on Temporal Information Extraction are based on annotation schemes and annotated corpora. These works handle several interrelated issues namely the automatic recognition of temporal entities and the temporal relations identification. This area of research is attracting researchers in several domains, which explain its exploitation in several real applications.

REFERENCES

- [1] X.Ling & D. S .Weld, (2010), "Temporal Information Extraction". In Proceedings of the Twenty-Fourth Conference on Artificial Intelligence (AAAI 2010). Atlanta, GA.
- [2] M. Kufman, (1993) "Proceeding of the Fifth Message Understanding Conference, (MUC-5)".
- [3] B. M. Sundheim, (1991). "Overview of the third message understanding evaluation and conference". In Proceedings of the Third Message Understanding Conference (MUC-3), pp 3–16, San Diego, CA.
- [4] R. Gaizauskas & Y. Wilks, (1998). "Information Extraction: Beyond Document Retrieval. Computational Linguistics and Chinese Language Processing. vol. 3, no. 2, August 1998, pp 17-60.
- [5] S. Sarawagi, (2008) "Information extraction". Foundations and Trends in Databases, 1:261–377.

- [6] C. Plake, T. Schiemann, M. Pankalla, J. Hakenberg & U. Leser, (2006). "Alibaba: Pubmed as agraph," *Bioinformatics*, vol. 22, pp. 2444–2445.
- [7] P. DeRose, W. Shen, Y. Lee, D. Burdick, A. Doan & R. Ramakrishnan, (2007) "Dblife: A community information management platform for the database research community (demo)," in *CIDR*, pp. 169–172.
- [8] E. Riloff, (1993) "Automatically constructing a dictionary for information extraction tasks". In *Proceedings of the 11th National Conference on Artificial Intelligence*, pp 811–816.
- [9] D. E. Appelt, J. R. Hobbs, J. Bear, D.J. Israel & M. Tyson, (1993). "Fastus: A finite-state processor for information extraction from real-world text," in *IJCAI*, pp 1172–1178.
- [10] H. Cunningham, D. Maynard, K. Bontcheva & V. Tablan, (2002) "Gate: A framework and graphical development environment for robust nlp tools and applications". In *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics*.
- [11] E. Agichtein & V. Ganti, (2004). "Mining reference tables for automatic text segmentation". In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Seattle, USA.
- [12] K. Seymore, A. McCallum & R. Rosenfeld, (1999) "Learning Hidden Markov Model structure for information extraction," in *Papers from the AAAI- 99 Workshop on Machine Learning for Information Extraction*, pp 37–42.
- [13] Zhang, N. R, (2001) "Hidden Markov Models for Information Extraction".
- [14] A. Borthwick, J. Sterling, E. Agichtein & R. Grishman, (1998) "Exploiting diverse knowledge sources via maximum entropy in named entity recognition," in *Sixth Workshop on Very Large Corpora New Brunswick, New Jersey, Association for Computational Linguistics*.
- [15] A. McCallum, D. Freitag & F. Pereira, (2000) "Maximum entropy markov models for information extraction and segmentation," in *Proceedings of the International Conference on Machine Learning (ICML-2000)*, pp 591–598, Palo Alto, CA.
- [16] R. Malouf, (2002) "A comparison of algorithms for maximum entropy parameter estimation," in *Proceedings of the Sixth Conference on Natural Language Learning (CoNLL-2002)*, pp 49–55.
- [17] F. Peng & A. McCallum, (2004) "Accurate information extraction from research papers using conditional random fields," in *HLT-NAACL*, pp 329–336.
- [18] M. Califf & R. Mooney, (2003) "Bottom-up Relational Learning of Pattern Matching Rules for Information Extraction".
- [19] B. Marthi, B. Milch, & S. Russell, (2003) "First-order probabilistic models for information extraction," in *Working Notes of the IJCAI-2003 Workshop on Learning Statistical Models from Relational Data (SRL-2003)*, (L. Getoor and D. Jensen, eds.), pp 71–78, Acapulco, Mexico.
- [20] Y. Choi, C. Cardie, E. Riloff, & S. Patwardhan, (2005) "Identifying sources of opinions with conditional random fields and extraction patterns," in *HLT/EMNLP*.
- [21] E. Bach, (1986) "The Algebra of Events". chapter 9, pp 5–16. *Linguistics and Philosophy*.
- [22] N. Asher, (1993) "Reference to abstract objects in discourse." *Kluwer Academic Publishers*, Dordrecht.
- [23] A. Bittar, (2010) "Building a TimeBank for French: A Reference Corpus Annotated According to the ISO- TimeML Standard." PhD thesis, Université Paris Diderot EFR de Linguistique, école doctorale de sciences du langage, laboratoire ALPAGE.
- [24] R.Sauri, J. Pustejovsky & J. Littman, (2006) "Argument structure in TimeML". In G. Katz, J. Pustejovsky & F. Schilder, *Annotating, Extracting and Reasoning about Time and Events*, Dagstuhl Seminar Proceedings, Dagstuhl, Germany. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany.
- [25] R. Longacre, (1983) "The Grammar of Discourse". New York. Plenum Press.
- [26] J. Allen, (1983), "Maintaining knowledge about temporal intervals". *Commun. ACM*, 26(11), pp, 832–843.
- [27] I. Mani, J. Pustejovsky & R. Gaizauskas, (2005) "The Language of Time. A Reader" Oxford University Press.
- [28] H. Reichenbach, (1947) "The tenses of verbs". *Elements of Symbolic Logic*. New York. Macmillan & Co.
- [29] Z. Vendler, (1967) "Verbs and Times", chapter 4, pp 97–121. Cornell University Press, Ithaca, New York.
- [30] F. Song & R. Cohen, (1991) "Tense interpretation in the context of narrative". In *AAAI*, pp. 131–136.

- [31] S. K. Sanampudi & G. Kumari, (2010) "Temporal reasoning in natural language processing: A survey." *International Journal of Computer Applications*, 1(4), pp53–57.
- [32] E. Filatova & E. Hovy, (2001) "Assigning time-stamps to event-clauses". In Katz, G., Pustejovsky, J., and Schilder, F., editors, the workshop on Temporal and spatial information processing, Dagstuhl Seminar Proceedings, Morristown, NJ, USA, Association for Computational Linguistics.
- [33] A. Setzer & R. Gaizauskas, (2002.) "On the importance of annotating temporal event-event relations in text". In the Third International Conference on Language Resources and Evaluation. LREC2002.
- [34] H. Kamp, (1981) "A theory of truth and semantic representation". In *Formal methods in the study of language*.
- [35] D.R. Dowty, (1986) "The effects of aspectual class on the temporal structure of discourse: "Semantics or pragmatics?", chapter 9, pp 37–61. *Linguistics and Philosophy*.
- [36] Webber, B. L, (1988), "Tense as discourse anaphor". Chapter 14, pp 61–73. *Linguistics and Philosophy*.
- [37] C. H. Hwang & L. K. Schubert, (1992) "Tense trees as the "fine structure" of discourse". In the 30th annual meeting on Association for Computational Linguistics, Morristown, NJ, USA, ACL.
- [38] A. F. ter Meulen, (1995) "Representing time in natural language: the dynamic interpretation of tense and aspect". Mit Press, Cambridge.
- [39] H. Kamp & U. Reyle, (1993) "From Discourse to Logic. Introduction to Model-theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory." *Studies in Linguistics and Philosophy*. Dordrecht, The Netherlands.
- [40] B. Moulin, (1997) "Temporal contexts for discourse representation: An extension of the conceptual graph approach". *Applied Intelligence*, 7 (3), pp 227–255.
- [41] C. D. Manning, & H. Schutze, (1999) "Foundations of statistical natural language processing". Mit Press, Cambridge.
- [42] J. Pustejovsky, J. Castano, R. Ingria, R. Sauri, R. Gaizauskas, A. Setzer & G. Katz, (2003a) "TimeML: Robust specification of event and temporal expressions in text", In Fifth International Workshop on Computational Semantics. IWCS-5.
- [43] J. Pustejovsky, R. Hanks, A. Sauri, R. See, A. Gaizauskas, A. Setzer, B. Radev, D. Sundheim, L. Day, L. Ferro & M. Lazo, (2003b) "The timebank corpus". In *Corpus Linguistics*. Lancaster.
- [44] R. Faiz, (2006) "Identifying relevant sentences in news articles for event information extraction". *International Journal of Computer Processing of Oriental Languages (IJCPOL)*, World Scientific, Vol. 19, No. 1, pp. 1-19.
- [45] A. Elkhilfi & R. Faiz, (2009) "Automatic Annotation Approach of Events in News Article", *International Journal of Computing and Information Sciences (IJCIS)*, Vol.7, N°1, pp. 19-28.
- [46] R. Grishman B. Sundheim, (1996) "Message understanding conference-6: A brief history". In the 16th Conference on Computational Linguistics, pp466–471, USA, Morristown, NJ, 1996. Association for Computational Linguistics.
- [47] L. Ferro, I. Mani, B. Sundheim & G. Wilson, (2001) "Tides temporal annotation guidelines". Technical Report 1.0.2, The MITRE Corporation, Washington C3 Center McLean, Virginia.
- [48] A. Setzer & R. Gaizauskas, (2000) "Annotating events and temporal information in newswire texts". In the Second International Conference on Language Resources and Evaluation, pp1287–1294. LREC2000.
- [49] S. Nunes, C. Ribeiroand & G. David, (2008) "Use of temporal expressions in web search". In ECIR 2008, pp 580–584. Springer-Verlag Berlin Heidelberg.
- [50] A. Roberts, R. Gaizauskas & M. Hepple, (2008) "Extracting clinical relationships from patient narratives", *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing BioNLP 2008*, pp 10–18, Columbus, Ohio, USA.
- [51] F. Schilder, (2007) "Event extraction and temporal reasoning in legal documents". In the proceedings of Annotating, Extracting and Reasoning about Time, pp59–71. LNAI Volume 4795/2007.
- [52] I.E. Fisher, (2007) "A prototype system for temporal reconstruction of financial accounting standards", *International Journal of Accounting Information Systems Volume 8, Issue 3, September 2007*, pp139–164