Artificial Intelligence In Medicine 105 (2020) 101860

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

Reconstructing the patient's natural history from electronic health records☆

Marjan Najafabadipour[a,b,*], Massimiliano Zanin[a], Alejandro Rodríguez-González[a], Maria Torrente[b], Beatriz Nuñez García[b], Juan Luis Cruz Bermudez[b], Mariano Provencio[b], Ernestina Menasalvas[a]

[a] Centro de Tecnología Biomédica, Universidad Politécnica de Madrid, Madrid, Spain
[b] Hospital Universitario Puerta de Hierro Majadahonda, Madrid, Spain

ARTICLE INFO

Keywords:
Electronic Health Records
Natural Language Processing
Temporal Reasoning

ABSTRACT

The automatic extraction of a patient's natural history from Electronic Health Records (EHRs) is a critical step towards building intelligent systems that can reason about clinical variables and support decision making. Although EHRs contain a large amount of valuable information about the patient's medical care, this information can only be fully understood when analyzed in a temporal context. Any intelligent system should then be able to extract medical concepts, date expressions, temporal relations and the temporal ordering of medical events from the free texts of EHRs; yet, this task is hard to tackle, due to the domain specific nature of EHRs, writing quality and lack of structure of these texts, and more generally the presence of redundant information. In this paper, we introduce a new Natural Language Processing (NLP) framework, capable of extracting the aforementioned elements from EHRs written in Spanish using rule-based methods. We focus on building medical timelines, which include disease diagnosis and its progression over time. By using a large dataset of EHRs comprising information about patients suffering from lung cancer, we show that our framework has an adequate level of performance by correctly building the timeline for 843 patients from a pool of 989 patients, achieving a precision of 0.852.

1. Introduction

The treatment of a disease does not only depend on the current condition of a patient, but also on his/her past medical history. This is why it is crucial for clinicians to have a complete and precise knowledge of the patient's natural history, which includes the disease, its progression over time, and any other significant fact in chronological order. As largely recognized in the literature, retrieving the patient's natural history can help improving clinical document summarization [1], clinical trial recruitment [2], clinical decision making [3] and patient's survival time calculation [4]. In addition, accessing this information allows clinicians to evaluate the quality of the provided healthcare, and to identify which of its steps require a special attention.

If such clinical information has traditionally been managed and accessed manually, the last decade has witnessed an increasing need for the digitization of clinical data. For this purpose, the information about the interactions between a patient and clinicians is frequently stored in computerized clinical records, which allow the reconstruction of the patient's natural history – see Fig. 1 for a graphical representation. Whenever a patient visits a hospital, one or more clinical notes can be digitally generated, describing highly detailed information about the patient's past and present medical condition, diagnosis, disease progressions, treatments, lab test results, etc. Note that, while digital in nature, notes are mostly composed of free texts and are therefore unstructured. These clinical notes are always written by a professional (physician, nurse, etc.). Complementary to notes, clinical reports are digitally generated once a medical process is completed, and they consolidate and synthesize the information contained in several clinical notes. They have a more structured format compared to the latter ones by containing different sections (e.g., Family Oncological History, Personal History, Diagnosis, Treatment, etc.) under which the relevant information is provided. For the sake of clarity, throughout this paper we collectively call these two sources "Electronic Health Records (EHRs)", while other non-textual elements usually therein included

☆ This article belongs to the Special issue: Artificial Intelligence in Computer-Based Medical Systems.

* Corresponding author.
E-mail addresses: m.najafabadipour@upm.es (M. Najafabadipour), massimiliano.zanin@gmail.com (M. Zanin), alejandro.rg@upm.es (A. Rodríguez-González), mtorrente80@gmail.com (M. Torrente), beangarcia@gmail.com (B. Nuñez García), jlcruz@idiphim.org (J.L. Cruz Bermudez), mprovenciop@gmail.com (M. Provencio), ernestina.menasalvas@upm.es (E. Menasalvas).

https://doi.org/10.1016/j.artmed.2020.101860
Received 29 October 2019; Received in revised form 6 April 2020; Accepted 6 April 2020

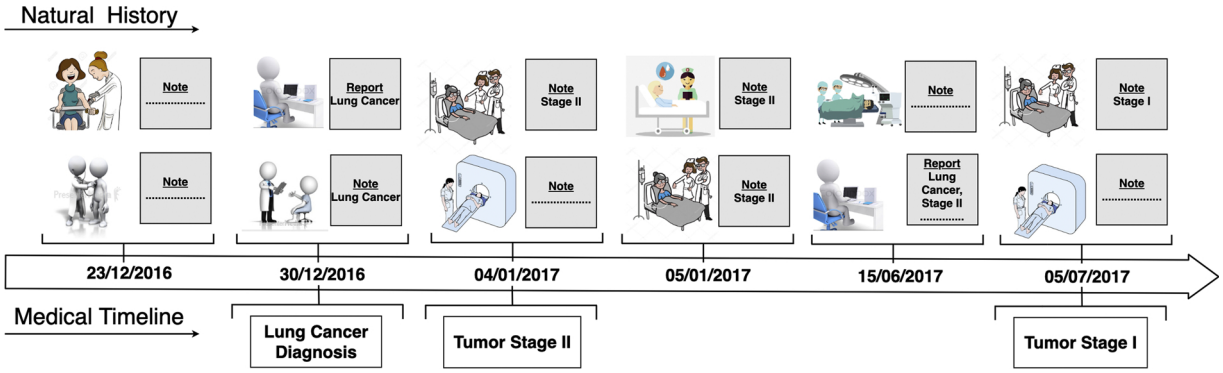


Fig. 1. An example of a natural history of a patient suffering from lung cancer (top part), and the corresponding medical timeline for disease diagnosis and progression (bottom part).

(e.g., echography results) are disregarded for being outside the scope of this study. EHRs are therefore unstructured clinical documents describing various medical events related to the patient's clinical condition and the corresponding chronological sequence.

Although EHRs contain all the information needed to reconstruct the patient's natural history, their manual analysis can be both costly and time-consuming [5]. Oncology provides an ideal case study to show the importance of automatic EHR processing. The different risk factors of cancer, the intra-tumor heterogeneity that implicates the patient's differences in relation to the same cancer type, and that the same patient presents differences between tumor sites, the inter-tumor heterogeneity, the differences in treatment response for the same cancer among patients, and the substantial difficulties in predicting tumor dynamics and the associated outcomes are some of the challenges that oncologists have to face in the daily clinical practice. When the objective becomes the design of a personalized treatment, big data analytics becomes the instrument of choice to tackle this heterogeneity and variability; in turn, this requires the extraction and processing of information coming from EHRs.

One of the types of cancer with higher prevalence and higher mortality worldwide, which encounters all of these difficulties is lung cancer. This is mainly due to the fact that its diagnosis is made in most cases in advanced stages of the disease where surgery is not an option anymore and the tumor burden is high. Diagnosis of lung cancer is usually accidental, during a visit to the emergency department when an image test is performed, and a lung nodule or mass is detected. From this moment, the patient should visit the medical oncology service, where the oncologist will confirm the diagnosis with additional tests if necessary, in order to subsequently receive the most suitable treatment available. Depending on the treatment and its frequency, the patient will visit accordingly both the consultation and the day care hospital where they receive treatments, along with frequent blood tests, imaging tests, and routine controls by the oncology nurses.

In terms of data generation, especially in lung cancer patients, the exploitation of this variety of data stored in EHRs from all lung cancer patients may lead to pattern extraction and better understanding of the disease evolution, toxicity rates and treatment response and outcomes. However, the process of reconstructing a patient's natural history from EHRs requires the extraction of several key elements, like medical concepts, date expressions, temporal relations and the order of medical events from free texts where medical events are medical concepts linked to their occurrence date expressions. Extraction of the aforementioned key elements in turn entails several challenges:

- Challenge 1 – information extraction: EHRs are mainly written in textual format and have no structure, or eventually have a custom structure defined by the hospital, service or the clinician generating them. In addition, clinical texts differentiate themselves from standard texts by containing many specific medical metrics, such as tumor stage codes. Their identification is challenging as these metrics usually include different symbols (e.g., ".", "-", "_", etc.), which are not always used in a standardized way and consequently, limit the use of standard ontologies, taxonomies and controlled vocabularies, such as Unified Medical Language System (UMLS) [6,7] and Systematized Nomenclature of Medicine (SNOMED) [8] in their recognition. EHRs are also very temporal in nature, with frequent mentions of date expressions. These are difficult to annotate, due to: (1) the presence of three categories of expressions, i.e. natural (e.g., "*3 days ago*", "*Today*"), conventional (e.g., "*2016-12-23*", "*December 12, 2016*") and professional (e.g., "*24 h*") date variables, each with their own idiosyncrasies; (2) the existence of domain specific, non-standard and abbreviated date expressions; (3) the presence of ambiguous date variables, having more than one meaning; and (4) the uncertainty inherent the interpretation of relative date expressions.
- Challenge 2 – linkage of medical events to date expressions: language like English and Spanish use the interplay of tense and aspect to encode temporal relations. However, the significance of these features may vary across domains and tasks. As a prototypical example, medical language, specifically EHRs, may ignore many restrictions that are mandatory in standard grammar, such as the fact that each sentence must have a subject. Clinical texts are typically ungrammatical, which make the automatic temporal reasoning a difficult task outside the medical community. In addition, temporal relation identification from clinical texts poses a special problem to NLP, as sentences in EHRs can be complex, including information about more than one medical event occurring at the same or at different time points. The determination of univocal relations between a date expression and the corresponding medical event can be very difficult. Furthermore, another problem emerges from the instant-based representation of medical events. In real applications, it is difficult to relate all medical events to their exact occurrence timing. Free texts include diverse, complex, and sometimes non-standard linguistic mechanisms for mentioning temporal relations. In some cases, the time associated to a medical event is not even explicitly mentioned.
- Challenge 3 – derivation of the order of medical events from the patient's EHRs: in order to generate a comprehensive medical timeline describing the patient's natural history, and thus exploit the temporal succession of medical events, it is firstly necessary to identify the temporal ordering of medical events across EHRs. As discussed in Ref. [9], this is a challenging problem in general NLP as well as the clinical domain, as the texts across patient's EHR lack logical continuity: the narrative goes back and forth in time, describing medical events that have happened at different time points. In addition, in general linguistics, events are often expressed by verbs, and thus tense and aspect are the elements used to temporally order them. This is nevertheless not always true in the context here

tackled, as many medical events are noun phrases [10], and most EHRs are written in past tense. As a last point, it is important to highlight the problem of information redundancy, a fundamental concept associated with EHRs that arises both within and across clinical data sources. The same medical event can be mentioned in multiple EHRs, especially because of two reasons: the tendency to re-use past notes to save time, i.e. by copying and pasting part of a previous note; and the interest to summarize past information in the newly generated EHRs. Two or more medical events are said to be similar (or also coreferential), if they have the same semantical category, value (if any) and have occurred on similar or consecutive time points. To illustrate, consider Fig. 1, in which the same medical event "*Stage II*" is mentioned four times on "*04/01/2017*","*05/01/2017*", "*15/06/2017*"; these mentions are coreferential, i.e. refer to the same event, and it is then necessary to determine its exact and real occurrence date.

Although a great amount of research works has been done on identifying temporal relations from clinical texts, the performance of most proposed systems is far from being adequate for practical applications. Most of these systems perform temporal reasoning with the help of annotated corpora, which are time consuming and costly to build, and whose completeness affects the quality of the analysis. In addition, despite the fact that Spanish is the second most popular language in the world with more than 572 million native speakers [11], little attention has been devoted to temporal relation discovery from Spanish free texts in the general domain. Finally, and to the best of our knowledge, no system has hitherto been proposed for the discovery of temporal relations from Spanish clinical texts.

The main contribution of this paper is to present a novel NLP framework, using rule-based techniques, that firstly, accepts Spanish EHRs annotated with medical concepts and date expressions as input, and then is able to relate these concepts together for reconstructing the patient's natural history. While, we aim to link medical concepts to the document creation dates, section dates and within-sentence dates in EHRs, identification of temporal relations between pairs of medical concepts are disregarded for being outside the scope of this study. In addition, in this paper, we propose to go one step further by extracting the evolution of medical events from these clinical documents, in order to build the patient's medical timeline, which contain the diagnosis of the disease and its progression over time. In particularly, we have applied this framework to generate the medical timelines of disease diagnosis and tumor stage events for patients suffering from lung cancer. Our framework presents a remarkable performance, yielding a precision of 0.852, as validated by using a large set of real EHRs.

Finally, in order to provide our NLP framework with the input of annotated lung cancer diagnosis concepts, tumor stage codes and date expressions, we have used a set of annotators, which are developed over Unstructured Information Management Framework (UIMA) [12–14]. These annotators along with the temporal relation identification process is facilitated with the tokenizer and section recognizer of C-liKES [12], where C-liKES is a text mining system, developed to ingest information written in Spanish free-texts format.

The rest of paper is organized as follows. Firstly, Section 2 explains the main related works on temporal relation discovery and patient's medical timeline construction. Afterwards, Section 3 discusses the details of the proposed framework by providing solutions for identifying temporal relations and building the patient's medical timeline. Section 4 presents the validation of the framework by using a real data set. Finally, Section 5 discusses the main advantages and limitations of the proposed framework, and Section 6 draws some conclusions and outlines future lines of work.

2. Related work

The 21st century has seen a considerable amount of research devoted to processing temporal information from free texts using statistical machine learning techniques and rule-based methods. The rapid development of temporal relation identification algorithms started with the creation of the TimeML [15] annotation schema for general newswire corpus of TimeBank [16]. This corpus contains three types of temporal information: (1) events; (2) time expressions; and (3) temporal relations.

The TimeBank corpus was used in three temporal analysis evaluation tasks in the SemEval competitions, i.e. TempEval-1 [17], TempEval-2 [18], and TempEval-3 [19]. While the Temp-Eval-1 provided the TimeBank corpus for English Language, the TempEval-2 provided this corpus for six languages, including English, Spanish, Italian, French, Chinese, and Korean. In TempEval- 2, the Temporal Information Processing based on Semantic information (TIPSem) algorithm [20] used Conditional Random Field (CRF) models to recognize temporal relation from Spanish free texts. TIPSem achieved a precision of 0.81 in the identification of temporal relations between events and time expressions, and a precision of 0.59 in the discovery of temporal relations between events and document creation time. Furthermore, while Temp-Eval-3 also provided the TimeBank corpus for Spanish, no systems were presented for finding temporal relations from newswire texts.

Although the TimeML group has developed a temporal annotation guideline, it only focuses on the news article domain. In recent years, the interest for temporal information identification from clinical texts has steadily been growing, partly due to the widespread adoption of EHRs [21]. In order to foster research activities on temporal relation discovery in the medical domain, the Integrating Biology and the Bedside (i2b2) NLP Challenge [22] was launched in 2012, providing an English corpus of discharge summaries annotated with events, time expressions and temporal information. Using this corpus, researchers were able to extract a limited set of temporal relations using rule-based and machine learning methods. The highest F1 score of 0.69 for the problem of temporal relation identification was achieved by two organizations: the Vanderbilt University on one hand, which proposed a rule-based pairwise selection with CRF and Support Vector Machine (SVM); and the National Research Council Canada, on the other hand, which implemented Maximum Entropy (ME), SVM and rule-based methods. In 2013, a hybrid system was also designed for the identification of temporal relations from clinical texts, which combined graph reasoning, and SVM and rule-based classification [23]. This system was validated using the test data set (120 clinical notes) of the 2012 i2b2 NLP challenge, obtaining an F1 measure of 0.63.

The authors of Refs. [24,25] modeled the temporal information appeared in clinical discharge summaries, written in English as Simple Temporal Problem. Based upon this work, an architecture was proposed in Ref. [26] for representing, extracting and reasoning about temporal information in clinical narrative texts, which was then incorporated both in the Medical Language Extraction and Encoding System (MedLEE) [27], and in TimeText [28]. This latter system obtained the recall of 79 % in the identification of temporal relations from fourteen discharge summaries, which were obtained from the clinical data repository at Columbia University Medical Center.

The enabling technologies for temporal relation and timeline discovery from clinical narratives were evaluated in Ref. [29]. As a result, an extension of ISO-TimeML guidelines was developed for annotation of a corpus of clinical notes, which was written in English and was provided by the Mayo clinic, named Temporal Histories for Your Medical Events (THYME) [30]. Many systems were developed for extracting events, time expressions and temporal relations using THYME in the context of the Clinical TempEval 2015 [31], Clinical TempEval 2016 [32] and Clinical TempEval 2017 [33]. These systems used rule-based and machine learning models (e.g., SVM, CRF, Recurrent Neural Networks (RNN), logistic regression, etc.) for their implementation.

In Clinical TempEval 2015, BluLab [34] was the only system presented. It used features generated from cTAKES [35] with CRF + +

methods for identifying relations between medical events and document creation time, called DocTimeRel (DR). For DR, BluLab reached an F1 score of 0.702 when the raw texts were used as input, and an F1 score of 0.791 when manually annotated events and time expressions were provided. In addition, BluLab also used features from cTAKES with the integration of CRF++ and rule-based techniques for the discovery of relations between medical events and/or time expressions, called Container Relation (CR). For CR tasks, when raw texts were provided as input, BluLab achieved an F1 score of 0.102 without temporal closure, and an F1 score of 0.123 with it. In addition, for CR with manually annotated events and time expressions as input, BluLab reached F1 scores of 0.143 (with temporal closure) and 0.181 (without temporal closure).

In Clinical TempEval 2016, UTHealth [36] submitted two runs for its implementations based on linear and structural (HMM) SVM, using lexical, morphological, syntactic, discourse, and word representation features. UTHealth run-1 was recognized as the best performing system, with F1 scores of 0.756 and 0.479 for DR and CR respectively, when plain texts were given as input. It also obtained the highest F1 score of 0.573 for CR when the manually annotated medical events and time expressions were provided as input. However, for detection of DR when the input was given as annotated medical events and time expressions, UtahBMI [37] gained the highest recall of 0.843. It implemented CRF and SVM models and used lexical, morphological, syntactic, shape, character pattern, character n-gram, section type, and gazetteer features.

In context of Clinical TempEval 2017, LIMSI−COT [38] used a combination of RNN with character and word embeddings, and SVM models with words and Part of Speech (PoS) tags as features. It obtained the best F1 scores for both unsupervised and supervised domain adaptations. For DR, it achieved F1 scores of 0.60 and 0.66 for unsupervised and supervised domain adaptions, respectively. Furthermore, for CR, it obtained an F1 score of 0.40 for unsupervised domain adaption and an F1 score of 0.43 for supervised domain adaptions.

An extension of Apache Text analysis and Knowledge Extraction System (cTAKES) was also proposed, based on an open-source temporal relation discovery system, and evaluated on the THYME corpus used in Clinical TempEval 2015, and on the 2012 i2b2 corpus [39]. This system used multiple supervised machine learning models for extracting the document creation time and within-sentence temporal relations. It achieved F1 scores of 0.807 and 0.321 for DR and CR, respectively, on the THYME corpus by using an SVM classifier. Also, it obtained an F1 score of 0.695 for the overall evaluation on all types of relations on the 2012 i2b2 corpus by implementing an SVM classifier and rules for coreference pairs. Later, an automated method to generate more high-quality training instances for temporal relation discovery was developed [40]. This method semantically expanded gold medical events based on UMLS using two within sentence temporal relation classification models. It included SVM as the learning algorithm for models. One of the models was developed for the identification of temporal relations between medical events and time expressions, while the second was developed for the detection of temporal relations between medical events. With the presented method, their temporal relation discovery system was evaluated on the colon cancer set of the THYME corpus used in Clinical TempEval 2015 and Clinical TempEval 2016, and achieved the F1 score of 0.594.

In 2018, it was claimed in Ref. [41] that despite of the considerable amount of research done for temporal relation identification in clinical texts, the state of art performance was not high enough for practical applications. As a result, an SVM-based system was developed for identifying direct temporal relations at sentence level in clinical notes written in English. This system is composed of three parts: (1) a pre-processor, which performs tokenization, section identification, PoS tagging, dependency parsing and semantic role labeling; (2) an SVM classifier, which discovers the direct temporal relation between an event and a time expression within a sentence; and (3) a post-processor, which uses deterministic rules to fix common errors emerging in the process. This system was evaluated on 310 discharge summaries and obtained an F1 score of 63.77.

In addition, a tree-based bidirectional Long Short Term Memory Network-RNN end-to-end model proposed in [42] was adapted to extract intra-sentential temporal relations from clinical texts [43]. This model was evaluated on the Clinical TempEval 2016 THYME corpus and obtained an F1 score of 0.629 for the identification of CR.

Furthermore, the authors of Ref. [44] proposed two metrics of Mean Squared Error (MSE) and Pairwise Ordering Accuracy (POA) to temporally order medical events mentioned within each clinical note in a listwise fashion. They also used ListNet [45], which implements neural networks as a listwise temporal ranking model in order to generate the timeline of medical events from clinical notes. By testing their model on THYME corpus, they achieved the MSE of 0.072 and POA of 0.517.

To construct a linear timeline from a set of annotated temporal relations, a set of two models were proposed in Ref. [46], known as Simple Context Independent Model (S-TLM) and Contextual Model (C-TLM). While S-TLM uses a linear projection method to determine the start and end point of temporal relations, C-TLM uses RNN. By testing S-TLM and C-TLM on the TempEval-3 dataset, the highest F measures obtained were 58.6 and 58.4, respectively.

Also, a pipeline for data driven medical event detection from social media using minimal supervision was developed to generate the medical time-line of events [47]. This pipeline includes two components of medical event extractor and temporal resolver, which mainly uses a set of rules for its implementation. It was validated on two datasets of Ravelry [48] and user's post from an online breast cancer community [49], and achieved F measure of 66.73.

Discourse structure, logical flow of sentences and context play a great role in the ordering of events based on temporal relations. However, the cross-document temporal ordering is challenging. The following research works focus on the generation of event timeline from multiple documents written in English.

In context of SemEval-2015 task 4 [50] on cross-document event ordering in general domain, a set of 4 teams submitted their results on the provided dataset by the task. WHUNLP team used the Stanford CoreNLP [51] and a set of rules for its implementations, and achieved the highest F1 score of 7.28 % when the raw texts were provided as input. However, when the gold event mentions were available, GPLSIUA team by using OpeNER language analysis toolchain [52], the TIPSem tool and Semantic Role Labeller from SENNA [53] obtained the best F1 score of 25.36 %.

In addition, a timeline extraction approach was proposed in Ref. [54], which generated noisy training data to anchor events to entities and temporal expressions by using distant supervision. This approach was validated on SemEval-2015 task 4 dataset and obtained F1 score of 28.58 when the gold event mentions were available.

For medical domain, an annotation schema was developed in the Ohio state university to extend the TimeML annotation guidelines to capture medical events from clinical texts written in English [55]. Then, using linear-chain CRF, each medical event was also anchored to a coarse time-bin (e.g., before admission, on admission, after admission, etc.) [56]. The temporal ordering of medical events mentioned in a single clinical narrative was then implemented using SVM-rank and based on the medical events proximity to the admission date [57]. Finally, a framework for aligning medical event sequences across clinical narratives was developed based on coreference and temporal relation information using cascaded Weight Finite-State Transducer (WFST) [58]. This framework was evaluated on a set of 7 patients (80 clinical

narratives overall) and obtained an accuracy of 78.9 %.

Furthermore, to generate a deep phenotype of individual cancer patients from English Clinical documents, a multi-scale information model, known as Deephe, was built on top of Apache cTAKES NLP system [59]. Deep phenotype refers to a set of attributes representing the clinical expression of a disease over time.

Finally, many of the current state-of-art systems implemented machine learning techniques for temporal reasoning tasks by using annotated corpora, provided by the shared tasks. However, the limited size of such corpora can unavoidably affect the quality of processing. In addition, only one work is presented for temporal relation discovery from Spanish newswire texts and to the best of our knowledge, no systems have been introduced for identification of temporal relations from Spanish clinical texts. Also, the extraction of evolution of medical events from the patient's EHRs remains an unsolved problem. Therefore, in line of this research, we propose an NLP framework, which is capable of extracting temporal relations from Spanish clinical texts and building the evolution of the medical events mentioned across patient's EHRs on a timeline.

3. Methods

EHRs are rich clinical data sources, containing information about the patient's medical care. Therefore, we introduce an NLP framework to mine EHRs in order to build the patient's medical timeline. This framework accepts the XML Metadata Interchange (XMI) files annotated with medical concepts and date expressions as input and yields the natural history of the patient using a medical timeline, which starts with the diagnosis event and includes the evolution of the patient's medical condition. The framework is composed of two components (Fig. 2): (1) the Temporal Reasoning System, which links medical events to their corresponding date expressions in the EHR; and (2) the Timeline Constructor, which generates the patient's medical timeline.

To better understand how our NLP framework builds the patient's medical timeline, consider the example of a patient's EHRs annotated with lung cancer diagnosis concepts, tumor stage codes and date expressions, and provided in form of XMI files as input to our framework. The Temporal Reasoning System reads these XMI files to identify the temporal relations between lung cancer diagnosis concepts and tumor stage codes with their corresponding occurrence date expressions. Once, all the temporal relations are extracted, they will be stored in a database. Afterwards, the Timeline Constructor reads this structured information and generates the patient's medical timeline, which starts with lung cancer diagnosis event and includes the evolution of tumor stage events. In the following subsections, we provide the detailed information about the components of our NLP framework.

3.1. Temporal Reasoning System

To construct the medical timelines, the medical concepts of interest and date expressions should be connected together, by finding temporal relations in the corresponding clinical texts at sentence, section or document level. In order to achieve this latter step, we developed a Temporal Reasoning System. This system accepts as input XMI documents, containing the annotations of medical concepts and date expressions. Then, it identifies temporal relations by building the dependency parse trees using the Universal Dependency Pipe (UDPipe) [60,61] tool and implementing a rule-based approach. Finally, once temporal relations are identified, the Temporal Reasoning System stores them into a MySQL structured relational database, named Document. The following Section 3.3.1 describes the details of UDPipe for building the dependency parse trees. Then, Section 3.3.2 explains the details of four rules implemented in order to identify temporal relations from clinical texts of EHRs.

3.1.1. UDPipe

UDPipe [60,61] is an open-source NLP tool, containing a pipeline of components such as tokenization, PoS tagging and universal dependency parsing for processing raw texts in multiple languages including Spanish. To generate the dependency parse trees, UDPipe uses a fast transition-based neural dependency parser, composed by a simple neural network with just one hidden layer and without any recurrent connections, and using locally normalized scores. The dependency parser builds an individual parse tree for each individual sentence in the clinical texts – See Appendix A.

3.1.2. Temporal relation identification

Our objective is to unequivocally identify the date expression of

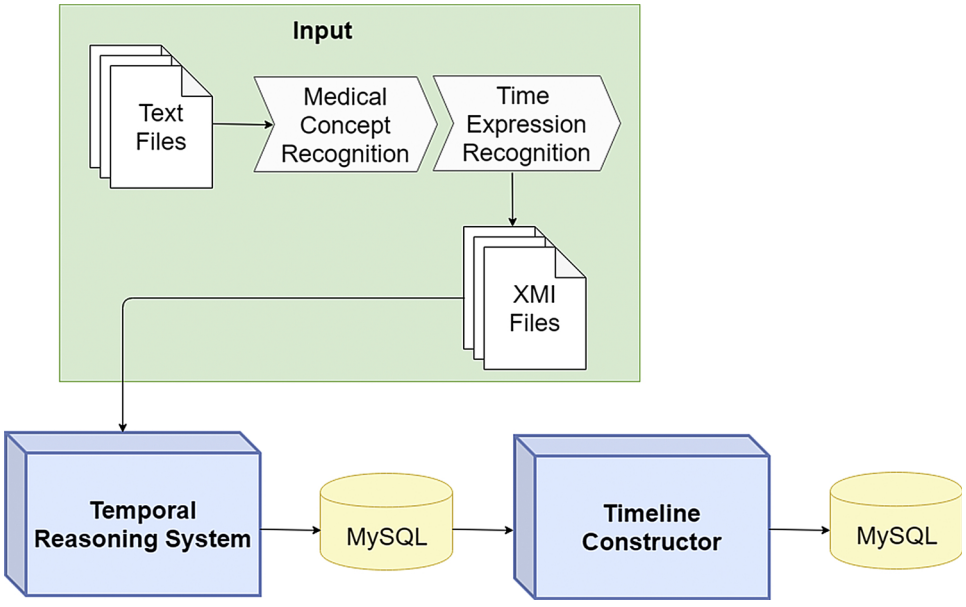


Fig. 2. Architecture of the proposed framework.

Table 1
Examples for the rules of temporal relation identification on the medical event "lung cancer".

Rules of temporal relation identification	An example that satisfies the rule	An example that does not satisfy the rule
Rule – 1	*Clinical Judgment: lung cancer on16/09/2011.* – See Fig. 3	*The patient was diagnosed with lung cancer (stage IIB) and had a surgery on 23/12/2016.* – See Fig. 4
Rule – 2	*The patient is diagnosed with lung cancer on 16/09/2011.* – See Fig. 5	*The patient has had the lung cancer (stage IIB) and a surgery on 23/12/2016.* – See Fig. 6
Rule –3	*Clinical Judgment – 16/09/2011: Lung cancer, stage IIB.*	*Treatment – 23/12/2016: The patient had a surgery due to diagnosis of lung cancer on 16/09/2011.*
Rule – 4	*Document Creation Date: 16/09/2011: The 50 years old patient is diagnosed with lung cancer and stage IIB.*	*Document Creation Date: 23/12/2016: The patient had a surgery due to diagnosis of lung cancer on 16/09/2011.*

each medical event. For this purpose, we define a temporal relation as:

1. A relation whose date expression modifies the medical event mention at the sentence level. This implies a syntactic construction where a medical event mention is directly accompanied by a date expression.
2. A relation whose date expression and medical event mention are arguments of the same predicate at the sentence level. When both elements are arguments of the same predicate, they are considered to be temporally related. A predicate is usually defined in linguistics as a verb or a noun. A predicate requires one or more arguments in different syntactic rules to complete its meaning. Adjunct is another type of grammatical components, which modify or complete the meaning of a predicate. However, opposed to arguments, we can remove adjunct from the sentence without making it grammatically wrong.
3. A relation whose section date expression explains the occurrence timing of a medical event mention. Due to clinician's time limitations, it is a common practice to write patient information without specifically mentioning the date expression for each and every medical event in the same sentence; instead, this information is provided as a date in the section headings.
4. A relation whose document creation date expression determines the occurrence timing of a medical event mention. Medical events happen in time, i.e. they are temporally associated objects. To keep record of these events, clinicians write them in EHRs during each patient's visit in the hospital. Since clinicians usually have limited time for writing the patient-clinician encounter, they usually mention the medical processes followed or the conditions discovered on the same day of the patient's visit without a mention of a date expression at sentence or section level. At the coarsest level, medical events are temporally related to the EHR creation date.

Based on the above definitions, the Temporal Reasoning System follows four rules to link medical event mentions to the date expressions (examples are provided in Table 1 to identify the occurrence date of the medical event "*lung cancer*"):

Rule – 1: If, at the sentence level, a medical event mention is an ancestor of a date expression in the dependency path or vice versa i.e., if a date expression is an ancestor of a medical event in the dependency path, they form a temporal relation based on syntactic structure, and therefore the medical event mention should be linked to that date expression. Otherwise, move to the next rule.

Rule – 2: If, at the sentence level, the medical event mention and the date expression are arguments or adjunct of the same predicate i.e., if the medical event mention and the date expression have the same parent node of type noun or verb in the dependency path, they form a temporal relation based on predicate argument structure, and therefore the medical event mention should be linked to that date expression. Otherwise, move to the next rule.

Rule – 3: If the section in which the medical event mention appeared contains a date expression in heading, they form a temporal relation, and should hence be linked. Otherwise, move to the next rule.

Rule – 4: If a medical event mention appeared in an EHR, it forms a temporal relation with the document creation date and should be linked to it.

3.2. Timeline Constructor

To approach the problem of dealing with redundant medical events across patient's EHRs and ordering them into a medical timeline, we developed a component named Timeline Constructor. It accepts as input the structured information of Document database. It then processes this information to produce the natural history of the patient on a medical timeline, which starts with the diagnosis event and includes the evolution of patient's clinical condition. Finally, the timeline is stored into a relational database.

As mentioned in Section 1, when two or more instances of medical events have similar semantical category and value (in case of medical events that are metrics e.g., tumor '*Stage IIB*' is a metric having value "*IIB*") and have occurred on the same or consecutive time points, they are said to be coreferential. To identify medical events, which are coreferential across a patient's EHRs, a specific time range (e.g., difference of one month) for defining consecutive time points is not considered rather a combination of similarity in terms of semantical category and value on the same time points or those occurred one after another until the value is changed, is considered. To determine whether two or more medical events are coreferential (See Listing 1), firstly, the Timeline Constructor temporally orders these events, for then evaluating their semantic similarity. Since it accepts the structured information of the Document database as input, it does not need to explicitly create complex procedures to discover semantic similarities between medical events. All the medical events with the same semantic category and with/without different notation of words or continuous groups of words are stored within the same table in the Document database. Consider the example of the lung cancer diagnosis and the tumor stage events, which are provided as structured information to the Timeline Constructor. Since this structured information is fed as two separate tables to the Timeline Constructor, this component can identify all the medical events belonging to the lung cancer diagnosis table are semantically similar to each other and semantically dissimilar to the tumor stage events.

Secondly, if the medical event is a metric, value similarity is the second factor to be considered. Therefore, at this stage, decisions made by the Timeline Constructor are binary, meaning either multiple instances of medical events from the same semantic category have similar values or not. Consider Fig. 1 as example where tumor stage events are "*Stage II – 04/01/2017*", "*Stage II – 05/01/2017*", "*Stage II – 05/01/2017*", "*Stage II – 15/06/2017*" and "*Stage I – 05/07/2017*". In here, the first four tumor stages, have value similarity as they hold II value.

```
1. Start;
2.    Set input = read the tables of Document database;
3.    For tables in input {
4.          Order medical events based on their date expressions in ascending order;
5.          If the semantic category of table == 'Diagnosis' {
6.                Timeline = keep the first medical event and discard the rest;
7.          } Else {
8.                Timeline = update Timeline by keeping the unique and the earliest instances of those medical events,
                  which have similar value and occurred on the same or consecutive time points;
9.          } Move to the next table;
10.   } Set output = Time-line;
11. End;
```

Listing 1. Pseudocode of Time-line Constructor.

Thirdly, in order for medical events to be coreferential, the last factor considered is the occurrence time point. Therefore, the Timeline Constructor determines whether two or more medical events with similar semantic and value are overlapped or occurred on consecutive time points. For example, in Fig. 1, the same medical event "*Stage II*" occurred four times on "*04/01/2017*","*05/01/2017*", "*15/06/2017*"; these mentions are coreferential, i.e. refer to the same event.

Finally, to generate the timeline of the patient's medical care, the Timeline Constructor selects the diagnosis event with the earliest time point and discards the rest. Then, for the rest of medical events, the Timeline Constructor keeps the unique and the earliest instance of the same medical event on the timeline, and discards the rest of redundant medical events, as they do not introduce any change of state.

4. Validation

In Europe, lung cancer led to the death of 266,000 persons, i.e. 20.8 % of all cancer deaths in 2011 [63], and to the greatest economic cost of 18.8 billion, i.e. 15 % of all cancer cost in 2009 [64], therefore, we here focus on validating our framework to reconstruct the natural history of patients, who suffer from lung cancer. To evaluate our framework, we used a dataset, containing the information of 989 lung cancer patients,

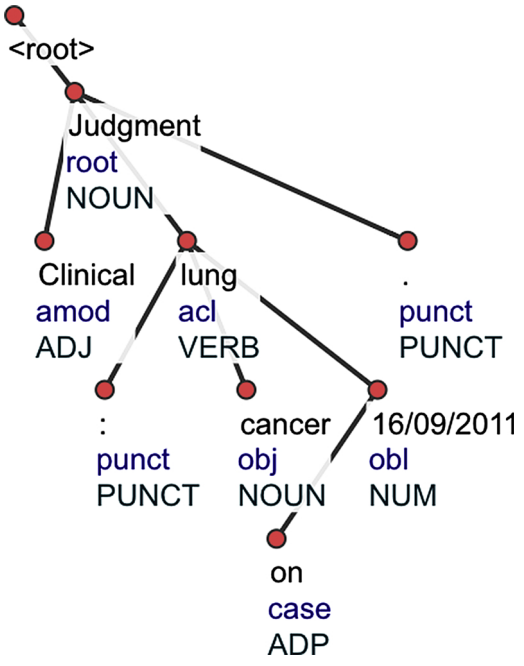


Fig. 3. Dependency parse tree analysis for an example that satisfies Rule – 1 [62].

which corresponds to 296,003 EHRs. The average number of EHRs per patient was 300. These EHRs were written in Spanish and were provided by the Hospital Universitario Puerta de Hierro Majadahonda (HUPHM) of Madrid. They contained the document creation date in structured format and were divided into two main sources of data, clinical notes (corresponding to 281,308 EHRs) and clinical reports (14,695 EHRs). The average lengths of clinical notes and reports were approximately 550 and 2300 words, respectively.

From these EHRs, clinicians are interested in finding specific patterns in long surviving lung cancer patients. The identification of these patterns will help them to detect unknown associations of family history, treatments, response to treatments, toxicities, comorbidities and molecular mechanisms, with the patient's outcome. In order to identify long surviving patients, the detection of the date of the lung cancer diagnosis and the evolution of tumor stage codes are the key factors.

Therefore, we here evaluate our framework by building the patient's medical timelines, starting with the lung cancer diagnosis event and include the evolution of tumor stage events. However, the first step toward generating such a timeline from multiple EHRs is to encode, structure and extract lung cancer diagnosis and tumor stage concepts along with the date expressions from clinical texts, as these represent the basis of medical events. This is where the need for the NLP annotators comes into play. The input to these annotators is the plain texts of EHRs and the output is formatted as a set of XMI files containing annotation results.

To identify the diagnosis concepts, we use the rule-based UMLS Annotator of C-liKES, which is built upon the UIMA framework. This annotator is designed to annotate noun and noun phrase concepts found in clinical texts that have contextually relevant matches in the UMLS as medical concepts. Example of such concepts are "*Lung cancer*", "*Cancer of lung*", "*Ca lung*", etc.

To recognize the tumor stage codes of lung cancer from clinical texts, we use the Stage Annotator and the TNM Annotator presented in a previous work of the authors [13]. These annotators are pattern-based extraction NLP modules, which are built over UIMA framework and using the American Joint Committee on Cancer Staging (AJCC) 8th manual [65] – see Fig. 7 – and the manual studies of EHRs. The Stage Annotator is capable of identifying the tumor stage grouping codes, which are written using roman numerals mixed with alphabets and numbers (e.g., "*I-A1*", "*IIB*", "*I(VA)*", etc.). On the other hand, the TNM Annotator can recognize those stage concepts, which contain three parameters of T (the size of tumor), N (the number of lymph nodes) and M (the presence of metastasis), modulated by suffixes and/or prefixes for a finer tuning of the tumor stages (e.g., "*pT1aN0M0*", "*cT3_cN1_cM0*", "*cT3-N0-M0*", etc.).

Finally, to extract and normalize date variables appeared in clinical texts, we use a rule-based NLP annotator built over UIMA framework, named Temporal Tagger, which is presented in a previous work of the authors [14]. The Temporal Tagger is developed based on the DATE type of TIMEX3 tag in TimeML annotation guidelines and by manually studying EHRs. This tagger is capable of:

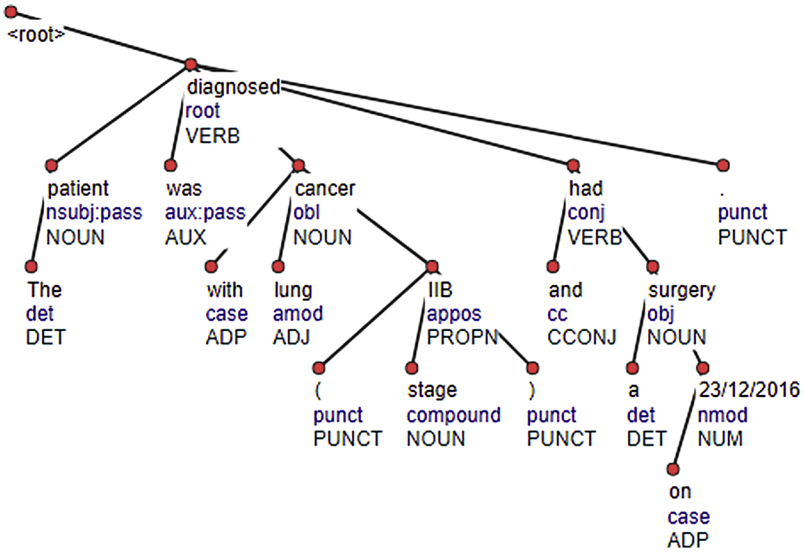


Fig. 4. Dependency parse tree analysis for an example that does not satisfy Rule – 1 [62].

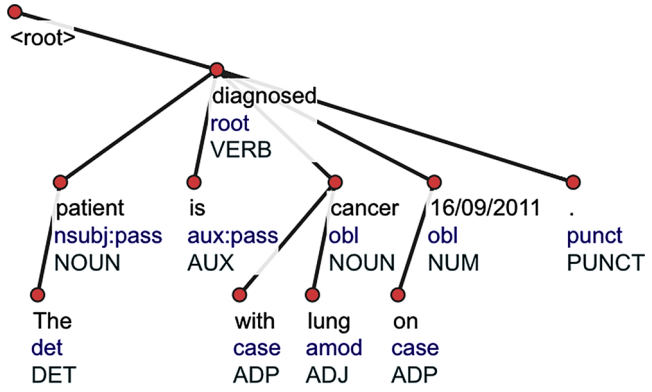


Fig. 5. Dependency parse tree analysis for an example that satisfies Rule – 2 [62].

1. Extracting various date expressions, i.e., natural (e.g., "*3 days ago*", "*Today*"), conventional (e.g., "*2016 – 12-23*", "*December 12, 2016*") and professional (e.g., "24 h") variables, three common ways a date can be written in Spanish. Also, it is able to annotate date expressions written in different formats (e.g., DD-MM-YYYY, MM-DD-YYYY) and styles (numerical, alphabetical, mixed – alphabetical and numerical or even abbreviated date expressions). For example, the date "*23/12/2016*" can be written in different formats and styles as "*23 – 12-2016*", "*Dec 23, 2016*", "*23rd of December 2016*", etc. Since our Temporal Tagger is optimized for Spanish, in which the standard date expression is written as DD-MM-YYYY or YYYY-MM-DD, we give priority to these two rules over the alternative MM-DD-YYYY and YYYY-DD-MM.
2. Filtrating date expressions that are not likely to be date using their PoS tags. For example, the Spanish word "*Tarde*" has two meaning, "late" and "afternoon". In this case, the Temporal Tagger ignores from annotating a single word "*Tarde*" if its PoS tag is not a noun.
3. Resolving date expressions with respect to the section date (if any) or document creation date. For example, for an EHR with the document creation date of "*23/12/2016*", the Temporal Tagger would resolve the date referred to by "*3 days ago*" into "*20/12/2016*". If there is a confusion about the time point which an expression refers to (e.g., "*Tuesday*"), the verb tense of the sentence is used to resolve the ambiguity. However, since the clinical texts do not follow the standard grammar and may not include a verb and the clinical narratives mostly provide information about past, then the relative date expression refers to the past by default.
4. Normalizing date expressions to a standard date format of YYYY-MM-DD.

The following sub-sections explain the conducted experiments, the

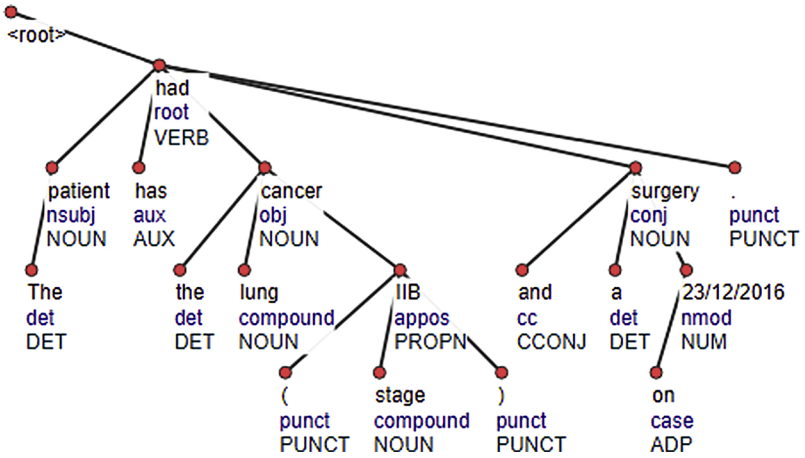


Fig. 6. Dependency parse tree analysis for an example that does not satisfy Rule – 2 [62].

T/M	Subcategory	N0	N1	N2	N3
T1	T1a	IA1	IIB	IIIA	IIIB
	T1b	IA2	IIB	IIIA	IIIB
	T1c	IA3	IIB	IIIA	IIIB
T2	T2a	IB	IIB	IIIA	IIIB
	T2b	IIA	IIB	IIIA	IIIB
T3	T3	IIB	IIIA	IIIB	IIIC
T4	T4	IIIA	IIIA	IIIB	IIIC
M1	M1a	IVA	IVA	IVA	IVA
	M1b	IVA	IVA	IVA	IVA
	M1c	IVB	IVB	IVB	IVB

Fig. 7. AJCC 8th edition – lung cancer stage grouping and TNM system [65].

details of the selected dataset samples used in these studies, and their results.

4.1. Experiments

To evaluate our developed framework, we designed four evaluation tasks: (1) validation of the outputs of the Stage Annotator and the TNM Annotator; (2) comparison of the Temporal Tagger with the Spanish versions of SUTime and HeidelTime; (3) validation of the output of the Temporal Reasoning System and its comparison with Spanish Version of TIPSem; and (4) validation of the output of the Timeline Constructor. For the first three evaluation tasks, two computer scientists served as the evaluation domain experts under the supervision of clinicians from HUPHM. They were native Spanish speakers and participated neither in the design nor in the development of the NLP framework. Furthermore, for the fourth evaluation task, four clinicians from HUPHM conducted the experiments. The details of these evaluation tasks are discussed in the following subsections.

4.1.1. First evaluation task – validation of the outputs of the stage annotator and the TNM annotator

As an extension to our previous work [13], the evaluations of the Stage Annotator and the TNM Annotator were done by manually analyzing the outputs generated by them. To validate the former, for each stage grouping codes mentioned in the clinical texts of EHRs, a comparison was done between the list of the codes automatically provided by the annotator and the list of expressions manually extracted by the evaluation domain experts.

Once the comparison is completed for each stage grouping code, the evaluation domain expert rated: (1) True Positive (TP), if the tumor stage code was correctly classified by the annotator; (2) False Positive (FP), if the tumor stage code was incorrectly classified by the annotator; and (3) False Negative (FN), if the annotator did not classify the tumor stage code when it should have. Given the TP, FP and FN, then the Precision = TP/(TP + FP), the Recall = TP/(TP + FN), and the F1 = (2 × Precision × Recall)/(Precision + Recall) were measured. In the case of precision and recall, confidence intervals were calculated by considering a binomial distribution, with confidence levels of 95 %. The same validation procedure was followed for the validation of the output of the TNM Annotator.

4.1.2. Second evaluation task – comparison of the Temporal Tagger with the Spanish versions of SUTime and HeidelTime

The aim of this evaluation task is to compare the performance of our Temporal Tagger [14] with the Spanish versions of Stanford SUTime [66,67] and HeidelTime [68–70]. Both SUTime and HeidelTime are pattern-based extraction annotators, capable of recognizing and normalizing date expressions written in textual documents. To perform this evaluation task, for each of annotators, a comparison is performed between the list of date expressions automatically extracted and the list of expressions manually extracted from EHRs. Once this comparison is completed, the values of TP, FP and FN were calculated in order to determine the precision, recall and F1 score values. For recall and precision, the confidence intervals were calculated. Finally, a comparison was performed between the results obtained from our Temporal Tagger, SUTime and HeidelTime. In previous work of the authors [14], the comparison was only performed between the Temporal Tagger and SUTime. Here, the goal is to also compare the Temporal Tagger with HeidelTime.

4.1.3. Third evaluation task – validation of the output of the Temporal Reasoning System

The evaluation of our Temporal Reasoning System in processing EHRs involved the verification of its output temporal constraints. For each temporal relation generated by our Temporal Reasoning System, the verification was done by manually analyzing each EHR for the corresponding temporal relation in order to compute TP and FP for determining the precision. Furthermore, as part of this task, we compare our Temporal Reasoning System with the Spanish version of TIPSem for identifying temporal relations from texts based on their precision. This comparison was done using the Spanish test dataset of TempEval-2 and based on its Task C, which focuses on the determination of temporal relations between an event and a time expression mentioned in the same sentence. This task is limited by requiring that either the event dominates the time expression syntactically or the event and time expression are mentioned in the same noun phrase.

4.1.4. Fourth evaluation task – validation of the output of the Timeline Constructor

The idea of our fourth evaluation task is to measure the precision of our Timeline Constructor in generating the timelines of medical events. The validation process was performed by manually studying all the EHRs for every patient and extracting the corresponding timeline, such that its starting point is the diagnosis, followed by the evolution of the tumor stage events. A comparison was done between the timeline manually extracted by the evaluation domain experts and the timeline generated by our Timeline Constructor. An error was counted for the Timeline Constructor when: (1) it generated a timeline with a different number of medical events (each associated to a date expression) compared to the one extracted by the evaluation domain experts; and (2) it

generated medical events with different date expressions for the same medical event identified by the evaluation domain experts.

4.2. Dataset sample selection

Due to the large pool of EHRs, performing the manual validation on the entire dataset was not feasible for the first three evaluation tasks. Therefore, to conduct our experiments, we have decided to perform individual random selection of EHRs from the original dataset for each of these individual tasks.

For the first evaluation task, after conducting a study, we realized that the tumor stage grouping and TNM codes have appeared in only 8 % and 9 % of EHRs, respectively. Due to low number of EHRs containing these codes, the size of the random sample of EHRs which would yield statistically significant result was too large to be practical. We have therefore decided to randomly select a sample of 550 EHRs from the original dataset, such that 50 of them (including 25 clinical notes and 25 clinical reports) contain annotations extracted by the Stage Annotator, and 500 of them (including 250 clinical notes and 250 clinical reports) do not contain any stage annotation. By construction, the maximum number of false positive is 50 (i.e. the 50 found annotations), as is the maximum number of false negatives (supposing that the annotator fails half of the times). Consequently, this selection ensures that both false positives and false negatives are tested with similar precision. The same procedure was followed for the dataset sample selection of the TNM Annotator. This practice allowed us to calculate the precision and recall of these annotators individually.

To select a dataset sample for the second evaluation task, 100 EHRs were randomly chosen from the original dataset, including 50 clinical notes and 50 clinical reports. In the context of the third evaluation task, we randomly selected 200 temporal relations generated by our Temporal Reasoning System from 200 EHRs, including 100 clinical notes and 100 clinical reports. The selection of equal number of clinical notes and reports was aimed at keeping both types of documents equally represented in the validation processes.

Finally, a set of chi-squared statistical tests was also performed on the selected samples, to assess their representativeness of the entire population in the original dataset. These tests were performed on four significant variables: (1) patient's sex; (2) patient's age (categorical variable: < 35, 35–40, 45–55, 55–65, 65–75, 75–80, 80 >); (3) local progression of the tumor; and (4) systemic progression of the tumor. For the dataset samples in the first and the third evaluation tasks, the sample dataset was representative of the entire population (p-value < 0.01). For the second evaluation task, in all cases, except for the systemic progression, the dataset sample was representative of the entire population. A small amount of bias was found in the systematic progression of tumor between the patients of the selected dataset sample and the patients of the original sample; note that this is not expected to affect the types of date expressions found.

4.3. Results

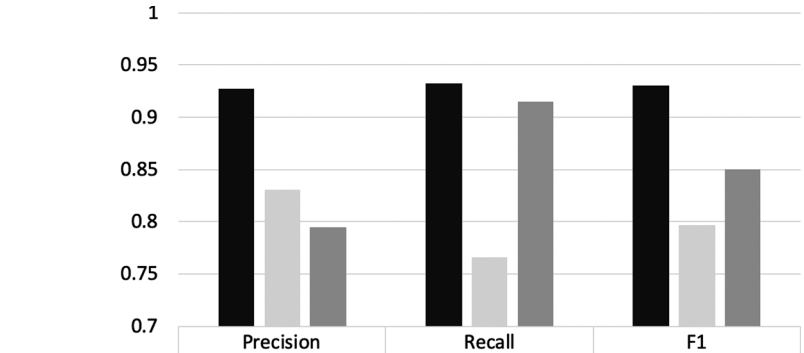
The results of the first evaluation task show that the Stage Annotator achieved a precision of 1.000 within a confidence interval of (0.952, 1.00), recall of 0.872 ± 0.089 and F1 of 0.932. To find the errors occurred in the annotation process, we analyzed its output extensively. By examining FNs, we realized that there are two main reasons for such errors. Firstly, this was due to ambiguous ways of writing tumor stage codes, i.e. writing the value of the tumor stage without mentioning that this value is referring to the stage of the tumor. For example, the annotator failed to annotate "*IV*" because there was no context word around it mentioning that this value is referring to the tumor stage. Secondly, it also happened that the standard system for writing tumor stage codes was not used in clinical texts. For example, instead of "*Stage IIIA*", the clinicians used "*Stage 3A*".

By validating the TNM Annotator, we have seen that it obtained a precision of 0.961 ± 0.071, a recall of 0.881 ± 0.089 and an F1 score of 0.919. By examining the FPs, we realized that main reason for such errors were incomplete usages of the TNM system (e.g., "*cT4*" instead of e.g., "*cT4 cN0 cM0*"). Likewise, the analysis of the FNs revealed that these errors occurred due to mentions of TNM codes in combination with some explanations about the tumor stage given by the clinician (e.g., "*pT2a (pleura) pN1 (fragmented hilar ganglion; margins probably +) M0*").

The results of our second evaluation task are discussed here. Firstly, it is worth mentioning that the conduction of this evaluation task revealed that the average numbers of date expressions found per clinical note and report were 6 and 20, respectively. As it can be seen in Fig. 8 and also as presented in the previous work of the authors [14], the Temporal Tagger obtained a precision of 0.927 ± 0.021, recall of 0.932 ± 0.021 and F1 score of 0.93. It outperformed SUTime in terms of precision, recall and F1 score. SUTime achieved the precision of 0.831 ± 0.033, recall of 0.766 ± 0.036 and F1 score of 0.797. In addition, the Temporal Tagger also outperformed HeidelTime, which obtained the precision of 0.795 ± 0.032, recall of 0.915 ± 0.025 and F1 Score of 0.85. Our Temporal Tagger had a better performance in identifying the various formats a date expression can be written in and in normalizing relative date expressions. Furthermore, by performing filtration, our temporal tagger obtained more correct results compared to SUTime and HeidelTime.

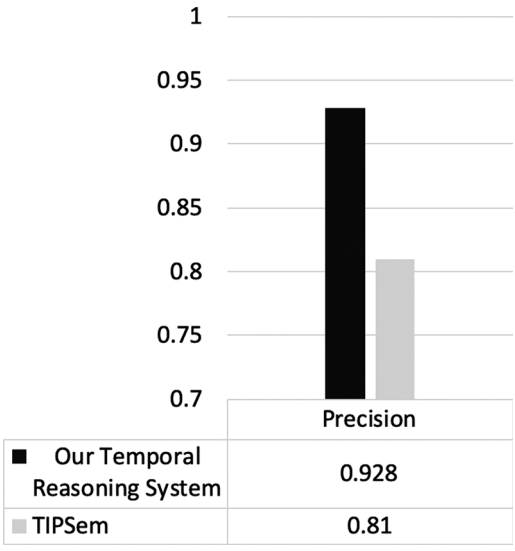
The results of the third evaluation task show that from the sample of 200 temporal relations, our Temporal Reasoning System correctly identified 178 temporal relations, being wrong in 22 instances, which correspond to a precision of 0.89. In this sample, 63, 51 and 86 of the temporal relations were related to the sentence, section and document creation dates, respectively. In order to understand the nature of the detected errors, we analyzed the obtained results to determine which kinds of temporal relations are difficult to be detected through our Temporal Reasoning System. It revealed that identifying temporal relations at sentence level is the most complex one, leading to 12 errors, which were the consequence of: (1) usage of very complex sentences and ambiguous temporal relations, as for instance, "*Lung cancer cT2N3M1b stage IV, due to carcinomatous lymphangitis, EGFR mutated (L858R of exón 21), diagnosed in November 2014*", which led to 5 errors; (2) missing usage of the dot "." or new lines at the end of a sentence, which led to the mixing of two or more sentences and resulted in 6 errors; and (3) missing annotation of date expressions by Temporal Tagger, which led to 1 error. In addition, 10 errors occurred due to the identification of section level temporal relations, where the medical events were actually related to the date expressions in the previous sentences (4 observations) or the document creation date (6 observations). Also, as it can be seen in Fig. 9, our Temporal Reasoning System achieved a precision of 0.928 for task C of TempEval-2, which outperformed TIPSem with a precision of 0.81. Since, the sentences in TempEval-2 dataset are grammatically more accurate than the clinical texts, our Temporal Reasoning system achieved a higher precision on this dataset compared to the clinical dataset provided by HUPHM.

The results generated from the fourth evaluation task show that the Timeline Constructor correctly extracted the complete medical timeline for 843 patients, while failing in 146 instances, which indicates the achievement of precision 0.852. We have seen that a typical patient's timeline contains 3 entries, including one diagnosis and two tumor stage events with a usual temporal range of 84 months. To find the nature of errors occurred, we analyzed the output extensively. We realized that the major reason was related to the incorrect temporal relations fed as input to the Timeline Constructor. However, we have seen that these errors are not catastrophic as they affect only parts of the timelines of patients. For instance, in the clinical text "*Treatment – 27 May 2014: A 48-year-old woman with lung cancer (Stage IIIA), will be treated with superior lobectomy and 4 cycles of adjuvant QT.*", the Temporal Reasoning System inaccurately assigned the section date "*27 May 2014*" to the event "*Stage IIIA*", while this stage should have been



	Precision	Recall	F1
Temporal Tagger	0.927	0.932	0.93
SUTime	0.831	0.766	0.797
HeidelTime	0.795	0.915	0.85

Fig. 8. Comparison of the results of our Temporal Tagger with SUTime and HeidelTime.



	Precision
Our Temporal Reasoning System	0.928
TIPSem	0.81

Fig. 9. Comparison of the results of our Temporal Reasoning System with TIPSem.

linked to the document creation date "*2014−05-27*". While this error caused the event "*Stage IIIA*" to be placed 12 days later than its actual date on the timeline, it did not lead to any errors in relation to the diagnosis event and the rest of stage events. In addition, a few errors have been observed due to the limitation of our framework to detect negations and probabilistic terms (e.g., "likely to have lung cancer", "can be excluded from having lung cancer").

5. Discussion

NLP technologies are helping researchers to extract new insights from large clinical and molecular datasets. Methodological pitfalls notwithstanding, NLP techniques are already beginning to affect cancer research and clinical care, such as early diagnosis and prevention [71], drug discovery [72], matching patients to clinical trials and treatment decisions [73].

In an era where data are being generated at an enormous pace – by 2020, it is expected that clinical data will double every 3 months, and that the average person will generate more than 1 million gigabytes of health-related data in their lifetime – it is increasingly difficult for clinicians to process all the available information that could influence treatment decisions. Furthermore, traditional analytics and machine technology have limited capability to utilize large complex datasets such as the so-called 'big data', and a change of paradigm is needed today to make the most from these potential sources of information. A major challenge is determining how to extract valuable information from the enormous amount of data available in EHRs, and research is ongoing to determine the best methodology to analyze data and reduce/eliminate unhelpful 'noise' [74].

In this paper, our aim was to reconstruct the natural history of patient from EHRs, written in Spanish, using rule-based approaches. The alternative, i.e. the adaption of machine learning approaches, requires annotated corpora, whose construction is costly and time-consuming. Also, their small size can significantly affect the processing quality. We have here shown that the rule-based approach is a viable alternative, which can yield very good results while avoiding the aforementioned problems. However, unlike rule-based approaches, machine learning techniques have learnability capability, and able to generalize and deal with new cases. Therefore, by obtaining the gold standard annotations to evaluate the performance of four tasks (extraction of medical concepts, date expressions, temporal relations and the order of medical events), it is now possible to use these annotated EHRs to create a Spanish clinical corpus to be used by machine learning models.

The previous results obtained from the Stage Annotator and the TNM Annotator support the idea that these annotators have shown an adequate level of performance in the NER process. We observed that the precision of these annotators is higher than their recall. However, since we have applied the pattern-based extraction approaches for annotation of tumor stage codes from clinical texts, the behavior of these annotators could be improved by extending their regular expressions.

We observed that our Temporal Tagger by supporting a set of regular expressions for annotation of natural, conventional and professional date expressions yields better results compared to SUTime and HeidelTime. However, the performance of this Temporal Tagger could be improved by recognizing and normalizing relative date expressions, which can appear in the previous sentences (e.g., "*Two months after surgery*", "*three days after the CT scan from last month*").

Reasoning with temporal information in texts is crucial to the field of NLP and biomedical informatics. In this paper, we proposed a Temporal Reasoning System for extracting temporal relations from Spanish clinical texts using dependency parsing and a set of four rules. We found that most of the temporal relations generated by the Temporal Reasoning System are correct. We have also seen a few errors due to the presence of very complex sentences, usage of ambiguous temporal relations, missing of end of sentence indicators, and the limitation of our system to identify temporal relations that span over a single sentence. We have also observed that our Temporal Reasoning System achieved a better performance on task C of TempEval-2 compared to the Spanish version of TIPSem.

In addition, extracting the evolution of medical events from a patient's EHRs helps in improving the quality of healthcare and

treatments. By implementing a rule-based approach, we have generated medical timelines, containing the evolution of patient's medical conditions. Many of the timelines generated by our Timeline Constructor were accurate. A few errors were observed due to feeding incorrect temporal relations into the Timeline Constructor. However, these errors only effected some parts of the patient's timelines. In addition, the behavior of the Timeline Constructor could be improved by annotating negations and probabilistic terms.

Finally, although our NLP framework has been developed general enough so that it can be used with no further adaptations in every medical domain to extract temporal relations and generate medical timelines form EHRs written in Spanish, its use in other languages would require the translation of the developed rules.

6. Conclusion and future work

The availability of larger volumes of data is already a reality in healthcare and represents an opportunity for clinicians to improve cancer care. However, technical and socio-cultural issues limit their use in practice. The challenge is to find a way of processing all the variables that provides simple and useful answers. Another challenge is understanding whether a computing tool can adapt to geographical differences in attitudes to healthcare, availability of medicines, etc. Importantly, a concerted effort is needed from all stakeholders (healthcare professionals, programmers, AI vendors, etc.) to discuss and agree their ideas, attitudes and goals for computing in oncology. This is vital to ensure mutual commitment to the development and integration of clinically useful tools and achieving the best outcomes for patients.

Medical data are being generated and captured in many ways, and at a pace we can no longer process as humans. This includes highly controlled structured data from clinical trials, which currently forms the basis for most decision-making. However, most of this data is generated and captured in a less controlled way and in unstructured forms, including from registries, electronic patient files, and social media (e.g. patient blogs). Unstructured data are much harder to process. Clinical trials will always be important; however, many questions cannot be answered by trials, for example the best sequence of treatments for each individual patient. Using real-world data could help answer these questions.

In medical Informatics, automatic temporal information extraction from clinical texts has become an active area of research. In the line of this area of research, we presented a novel NLP framework for extracting medical concepts, date expressions, temporal relations and medical timeline from patient EHRs, written in Spanish. For the annotation of medical concepts and date expression from clinical texts, we used a set of rule-based NLP annotators, built upon the Apache UIMA framework. In addition, temporal relations between medical events and date expressions at sentence, section and document level were discovered using a Temporal Reasoning System, which implements the dependency parsing tree and a rule-based method. Since, TIPSem is the only temporal reasoning system presented for Spanish language to identify temporal relations from free texts using machine learning techniques, we have shown that our rule-based approach is an alternative system, which can obtain very accurate results while avoiding the problems of costly and time-consuming process of creating the annotated corpora and also having low processing quality due to the small size of corpora. Furthermore, to generate the patient's medical timeline from multiple EHRs, a Timeline Constructor component was developed for dealing with information redundancy issues using rule-based methods. However, this work is an on-going research, in which future efforts will be aimed at the derivation of the cross-EHRs evolution of treatment events, which usually occur in a time interval, i.e. they have both starting and ending time points, and which are highly dependent on the dosage.

Conflict of interest

No conflicts of interest.

Acknowledgments

This paper is supported by European Union's Horizon 2020 research and innovation programme [grant agreement number 727658], project IASIS (Integration and analysis of heterogeneous big data for precision medicine and suggested treatments for different types of patients); MN is also supported by UPM (Universidad Politécnica de Madrid) Programa Propio of PhD grants.

Appendix A

UDPipe employs nearly unmodified Parsito parser [75] for building dependency parse trees. Parsito is a transition-based and non-projective dependency parser, which parses both non-projective and projective sentences. This parser uses a neural network classifier to perform prediction and does not require feature engineering. It contains a search-based oracle that improves the accuracy of parsing as dynamic oracle, but it can be applied to any transition systems (e.g., fully non-projective swap system). To improve the accuracy of Parsito, UDpipe added an optional beam-search decoding.

To generate a dependency parse tree for a sentence, firstly, UDpipe accepts the texts as input and then, it assigns PoS (e.g., DET, NOUN, VERB, etc.) and dependency relation (e.g., det, nsubj:pass) tags to the tokens. Afterwards, using these tags, it determines the root node of the dependency

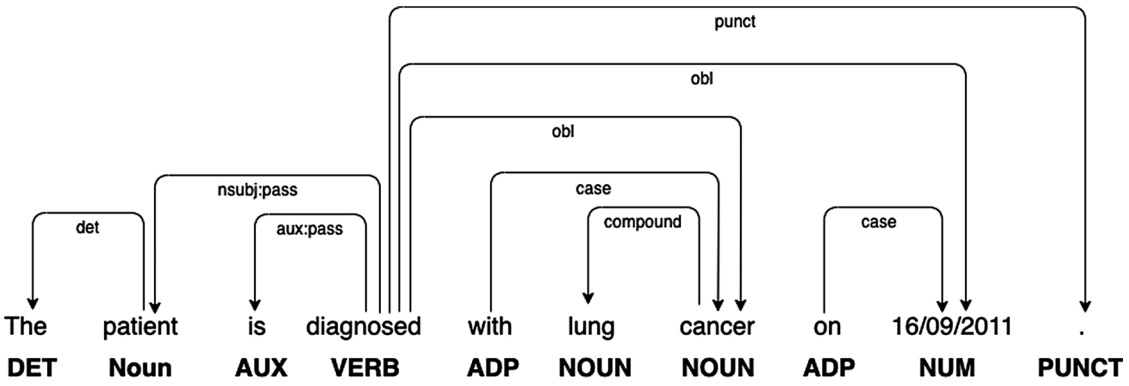


Fig. 10. Example of dependency parsing by UDpipe.

tree and generates its child nodes. For example, consider Fig. 10, where the input sentence to UDPipe is "*The patient is diagnosed with lung cancer on 16/09/2011*.". After assigning the PoS and relation dependency tags to each word in the sentence, UDPipe determines the token of "*diagnosed*" as the root node. It then moves to the next level of the tree, where the child nodes are "*patient*", "*is*", "*cancer*", "*16/09/2011*", and ".". Afterwards, it generates level 2 of the tree with child nodes of "*the*", "*with*", "*lung*", and "*on*".

Appendix B. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.artmed.2020.101860.

References

[1] Reichert D, Kaufman D, Bloxham B, Chase H, Elhadad N. Cognitive analysis of the summarization of longitudinal patient records. AMIA Annu Symp Proc 2010;2010:667–71.

[2] Luo Z, Johnson SB, Lai AM, Weng C. Extracting temporal constraints from clinical research eligibility criteria using conditional random fields. AMIA Annu Symp Proc 2011;2011:843–52.

[3] Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? J Biomed Inform 2009;42:760–72. https://doi.org/10.1016/j.jbi.2009.08.007.

[4] Zadnik V, Žagar T, Žakelj MP. Cancer patients' survival: standard calculation methods and some considerations regarding their interpretation. Zdr Varst 2016;55:144–51. https://doi.org/10.1515/sjph-2016-0012.

[5] Najafabadipour M, Tuñas JM, Rodríguez-González A, Menasalvas E. Analysis of electronic health records to identify the patient's treatment lines: challenges and opportunities. Artificial intelligence XXXVI. 2019. p. 437–42. https://doi.org/10.1007/978-3-030-34885-4_33.

[6] Humphreys BL, Lindberg DA. The UMLS project: making the conceptual connection between users and the information they need. Bull Med Libr Assoc 1993;81:170–7.

[7] Unified Medical Language System (UMLS). Available: https://www.nlm.nih.gov/research/umls/index.html. [Accessed 27 October 2019].

[8] SNOMED Home Page. Available: http://www.snomed.org/. [Accessed 27 October 2019].

[9] Radev DR. A common theory of information fusion from multiple text sources step one: cross-document structure. Proceedings of the 1st SIGdial Workshop on Discourse and Dialogue - Volume 10 2000:74–83. https://doi.org/10.3115/1117736.1117745.

[10] Zhou L, Hripcsak G. Temporal reasoning with medical data—a review with emphasis on medical natural language processing. J Biomed Inform 2007;40:183–202. https://doi.org/10.1016/j.jbi.2006.12.009.

[11] Manuel Bonet J, Rodríguez-Ponga Salamanca R, María Martínez Alonso J, Bueno Hudson R, López-Vega M, Fernández Vítores D. El Español: UNA LENGUA VIVA. Instituto Cervantes; 2018https://cvc.cervantes.es/lengua/espanol_lengua_viva/pdf/espanol_lengua_viva_2018.pdf.

[12] Ruiz EM, Tuñas JM, Bermejo G, Martín CG, Rodríguez-González A, Zanin M. Profiling lung cancer patients using electronic health records. J Med Syst 2018;42:126. https://doi.org/10.1007/s10916-018-0975-9.

[13] Najafabadipour M, Tuñas JM, Rodríguez-González A, Menasalvas E. Lung cancer concept annotation from Spanish clinical narratives. Data integration in the life sciences. 2019. p. 153–63. https://doi.org/10.1007/978-3-030-06016-9_15.

[14] Najafabadipour M, Zanin M, Rodríguez-González A, Gonzalo-Martín C, García BN, Calvo V. Recognition of time expressions in Spanish electronic health records. 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS) 2019:69–74. https://doi.org/10.1109/CBMS.2019.00025.

[15] Pustejovsky J, et al. TimeML: robust specification of event and temporal expressions in text. New directions in question answering. 2003. p. 1–12.

[16] Pustejovsky J, Hanks P, Saurí R, See A, Gaizauskas R, Setzer A. The TimeBank corpus. Proceedings of Corpus Linguistics 2003:647–56.

[17] Verhagen M, Gaizauskas R, Schilder F, Hepple M, Katz G, Pustejovsky J. SemEval-2007 task 15: TempEval temporal relation identification. Proceedings of the 4th International Workshop on Semantic Evaluations - SemEval' 07 2007:75–80. https://doi.org/10.3115/1621474.1621488.

[18] Verhagen M, Sauri R, Caselli T, Pustejovsky J. SemEval-2010 task 13: TempEval-2. Proceedings of the 5th International Workshop on Semantic Evaluation 2010:57–62.

[19] UzZaman N, Llorens H, Derczynski L, Allen J, Verhagen M, Pustejovsky J. SemEval-2013 task 1: TempEval-3: evaluating time expressions, events, and temporal relations. Second Joint Conference on Lexical and Computational Semantics (*SEM) Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013) 2013:1–9.

[20] Llorens H, Saquete E, Navarro B. TIPSem (English and Spanish): evaluating CRFs and semantic roles in TempEval-2. Proceedings of the 5th International Workshop on Semantic Evaluation 2010:284–91.

[21] Moharasar G, Ho TB. A semi-supervised approach for temporal information extraction from clinical text. 2016 IEEE RIVF International Conference on Computing Communication Technologies, Research, Innovation, and Vision for the Future (RIVF) 2016:7–12. https://doi.org/10.1109/RIVF.2016.7800261.

[22] Sun W, Rumshisky A, Uzuner O. Evaluating temporal relations in clinical text: 2012 i2b2 challenge. J Am Med Inform Assoc 2013;20:806–13. https://doi.org/10.1136/amiajnl-2013-001628.

[23] Nikfarjam A, Emadzadeh E, Gonzalez G. Towards generating a patient's timeline: extracting temporal relationships from clinical notes. J Biomed Inform 2013;46:S40–7. https://doi.org/10.1016/j.jbi.2013.11.001.

[24] Hripcsak G, Zhou L, Parsons S, Das AK, Johnson SB. Modeling electronic discharge summaries as a simple temporal constraint satisfaction problem. J Am Med Inform Assoc 2005;12:55–63. https://doi.org/10.1197/jamia.M1623.

[25] Zhou L, Melton GB, Parsons S, Hripcsak G. A temporal constraint structure for extracting temporal information from clinical narrative. J Biomed Inform 2006;39:424–39. https://doi.org/10.1016/j.jbi.2005.07.002.

[26] Zhou L, Friedman C, Parsons S, Hripcsak G. System architecture for temporal information extraction, representation and reasoning in clinical narrative reports. AMIA Annu Symp Proc 2005;2005:869–73.

[27] Friedman C, Hripcsak G, DuMouchel W, Johnson SB, Clayton PD. Natural language processing in an operational clinical information system. Nat Lang Eng 1995;1:83–108. https://doi.org/10.1017/S1351324900000061.

[28] Zhou L, Parsons S, Hripcsak G. The evaluation of a temporal reasoning system in processing clinical discharge summaries. J Am Med Inform Assoc 2008;15:99–106. https://doi.org/10.1197/jamia.M2467.

[29] Savova G, Bethard S, Styler W, Martin J, Palmer M, Masanz J. Towards temporal relation discovery from the clinical narrative. AMIA Annu Symp Proc 2009;2009:568–72.

[30] Styler WF, Bethard S, Finan S, Palmer M, Pradhan S, Groen PC. Temporal annotation in the clinical domain. Trans Assoc Comput Linguist 2014;2:143–54.

[31] Bethard S, Derczynski L, Savova G, Pustejovsky J, Verhagen M. SemEval-2015 task 6: clinical TempEval. Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015) 2015:806–14.

[32] Bethard S, Savova G, Chen W-T, Derczynski L, Pustejovsky J, Verhagen M. SemEval-2016 task 12: clinical TempEval. Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016) 2016:1052–62. https://doi.org/10.18653/v1/S16-1165.

[33] Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC. SemEval-2017 task 12: clinical TempEval. Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017) 2017:565–72. https://doi.org/10.18653/v1/S17-2093.

[34] Velupillai S, Mowery DL, Abdelrahman S, Christensen L, Chapman W. BluLab: temporal information extraction for the 2015 clinical TempEval challenge. Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015) 2015:815–9. https://doi.org/10.18653/v1/S15-2137.

[35] Savova GK, et al. Mayo clinical text analysis and knowledge extraction system (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc 2010;17:507–13. https://doi.org/10.1136/jamia.2009.001560.

[36] Lee H-J, Zhang Y, Xu J, Moon S, Wang J, Wu Y. UTHealth at SemEval-2016 task 12: an End-to-End system for temporal information extraction from clinical notes. Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016) 2016:1292–7. https://doi.org/10.18653/v1/S16-1201.

[37] AAl Abdulsalam A, Velupillai S, Meystre S. UtahBMI at SemEval-2016 task 12: extracting temporal information from clinical text. Presented at the 10th International Workshop on Semantic Evaluation (SemEval-2016) 2016:1256–62.

[38] Tourille J, Ferret O, Tannier X, Névéol A. LIMSI-COT at SemEval-2017 task 12: neural architecture for temporal information extraction from clinical narratives. Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017) 2017:597–602. https://doi.org/10.18653/v1/S17-2098.

[39] Lin C, Dligach D, Miller TA, Bethard S, Savova GK. Multilayered temporal modeling for the clinical domain. J Am Med Inform Assoc 2016;23:387–95. https://doi.org/10.1093/jamia/ocv113.

[40] Lin C, Miller T, Dligach D, Bethard S, Savova G. Improving temporal relation extraction with training instance augmentation. Proceedings of the 15th Workshop on Biomedical Natural Language Processing 2016:108–13. https://doi.org/10.18653/v1/W16-2914.

[41] Lee H-J, Zhang Y, Jiang M, Xu J, Tao C, Xu H. Identifying direct temporal relations between time and events from clinical notes. BMC Med Inform Decis Making 2018;18:49. https://doi.org/10.1186/s12911-018-0627-5.

[42] Miwa M, Bansal M. End-to-end relation extraction using LSTMs on sequences and tree structures. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics 2016:1105–16. https://doi.org/10.18653/v1/P16-1105. (Volume 1: Long Papers).

[43] Galvan D, Okazaki N, Matsuda K, Inui K. Investigating the challenges of temporal relation extraction from clinical text. Proceedings of the Ninth International Workshop on Health Text Mining and Information Analysis 2018:55–64. https://doi.org/10.18653/v1/W18-5607.

[44] Jeblee S, Hirst G. Listwise temporal ordering of events in clinical notes. Proceedings of the Ninth International Workshop on Health Text Mining and Information Analysis 2018:177–82. https://doi.org/10.18653/v1/W18-5620.

[45] shiba24/learning2rank. Available: https://github.com/shiba24/learning2rank

[Accessed 18 February 2020].
[46] Leeuwenberg A, Moens M-F. Temporal information extraction by predicting relative time-lines. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing 2018:1237–46. https://doi.org/10.18653/v1/D18-1155.
[47] Naik A, Bogart C, Rose C. Extracting personal medical events for user timeline construction using minimal supervision. BioNLP 2017 2017:356–64. https://doi.org/10.18653/v1/W17-2346.
[48] Ravelry — a knit and crochet community. Available: https://www.ravelry.com/account/login. [Accessed 19 February 2020].
[49] Breastcancer.org — breast cancer information and support. Available: https://www.breastcancer.org/. [Accessed 19 February 2020].
[50] Minard A-L, Speranza M, Agirre E, Aldabe I, van Erp M, Magnini B. SemEval-2015 task 4: TimeLine: cross-document event ordering. Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015) 2015:778–86. https://doi.org/10.18653/v1/S15-2132.
[51] Manning C, Surdeanu M, Bauer J, Finkel J, Bethard S, McClosky D. The stanford CoreNLP natural language processing toolkit. Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations 2014:55–60. https://doi.org/10.3115/v1/P14-5010.
[52] OpeNER suite of tools — ELRC-SHARE. Available: https://elrc-share.eu/repository/browse/opener-suite-of-tools/1c0425724b2b11e9a7e100155d026706a830b480a23f40deb20750a2b6937309/. [Accessed 19 February 2020].
[53] SENNA. Available: https://ronan.collobert.com/senna/. [Accessed 19 February 2020].
[54] Cornegruta S, Vlachos A. Timeline extraction using distant supervision and joint inference. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing 2016:1936–42. https://doi.org/10.18653/v1/D16-1200.
[55] Raghavan P, Fosler-Lussier E, Lai AM. Inter-annotator reliability of medical events, coreferences and temporal relations in clinical narratives by annotators with varying levels of clinical expertise. AMIA Annu Symp Proc 2012;2012:1366–74.
[56] Raghavan P, Fosler-Lussier E, Lai AM. Temporal classification of medical events. Proceedings of the 2012 Workshop on Biomedical Natural Language Processing 2012:29–37.
[57] Raghavan P, Lai A, Fosler-Lussier E. Learning to temporally order medical events in clinical text. Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics 2012:70–4 (Volume 2: Short Papers).
[58] Raghavan P, Fosler-Lussier E, Elhadad N, Lai AM. Cross-narrative temporal ordering of medical events. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics 2014:998–1008. https://doi.org/10.3115/v1/P14-1094. (Volume 1: Long Papers).
[59] Hochheiser H, Castine M, Harris D, Savova G, Jacobson RS. An information model for computable cancer phenotypes. BMC Med Inform Decis Making 2016;16:121. https://doi.org/10.1186/s12911-016-0358-4.
[60] Straka M, Hajič J, Straková J. UDPipe: trainable pipeline for processing CoNLL-U files performing tokenization, morphological analysis, POS tagging and parsing. Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16) 2016:4290–7.
[61] Straka M, Straková J. Tokenizing, POS tagging, lemmatizing and parsing UD 2.0 with UDPipe. Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies 2017:88–99. https://doi.org/10.18653/v1/K17-3009.
[62] UDPipe. Available: http://lindat.mff.cuni.cz/services/udpipe/. [Accessed 22 October 2019].
[63] 1 in 4 deaths caused by cancer in the EU28. Available: https://ec.europa.eu/eurostat/web/products-press-releases/-/3-25112014-BP. [Accessed 21 October 2019].
[64] Luengo-Fernandez R, Leal J, Gray A, Sullivan R. Economic burden of cancer across the European Union: a population-based cost analysis. Lancet Oncol 2013;14:1165–74. https://doi.org/10.1016/S1470-2045(13)70442-X.
[65] Detterbeck FC. The eighth edition TNM stage classification for lung cancer: what does it mean on main street? J Thorac Cardiovasc Surg 2018;155:356–9. https://doi.org/10.1016/j.jtcvs.2017.08.138.
[66] Chang AX, Manning CD. SUTIME: a library for recognizing and normalizing time expressions. Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012) 2012.
[67] The Stanford Natural Language Processing Group. Available: https://nlp.stanford.edu/software/sutime.html. [Accessed 27 October 2019].
[68] Strötgen J, Gertz M. Heideltime: high quality rule-based extraction and normalization of temporal expressions. Proceedings of the 5th International Workshop on Semantic Evaluation, Sem-Eval, 2010 2010:321–4.
[69] Strötgen J, Zell J, Gertz M. HeidelTime: tuning English and developing Spanish resources for TempEval-3. Second Joint Conference on Lexical and Computational Semantics (*SEM) Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013) 2013:15–9.
[70] Database Research Group: HeidelTime demonstration. Available: https://heideltime.ifi.uni-heidelberg.de/heideltime/. [Accessed 27 October 2019].
[71] Strauss JA, Chao CR, Kwan ML, Ahmed SA, Schottinger JE, Quinn VP. Identifying primary and recurrent cancers using a SAS-based natural language processing algorithm. J Am Med Inform Assoc 2013;20:349–55. https://doi.org/10.1136/amiajnl-2012-000928.
[72] Begoli E, Kusnezov D. Artificial intelligence's essential role in the process of drug discovery. Future Drug Discovery 2019;1. https://doi.org/10.4155/fdd-2019-0026. FDD21.
[73] Maguire FB, Morris CR, Parikh-Patel A, Cress RD, Keegan THM, Li C-S. A text-mining approach to obtain detailed treatment information from free-text fields in population-based cancer registries: a study of non-small cell lung cancer in California. PLoS One 2019;14. https://doi.org/10.1371/journal.pone.0212454.
[74] Velupillai S, Suominen H, Liakata M, Roberts A, Shah AD, Morley K. Using clinical natural language processing for health outcomes research: overview and actionable suggestions for future advances. J Biomed Inform 2018;88:11–9. https://doi.org/10.1016/j.jbi.2018.10.005.
[75] Straka Milan, Hajic J, Strakova J, Hajic Jr. J. Parsing universal dependency treebanks using neural networks and search-based oracle Milan. 14th International Workshop on Treebanks and Linguistic Theories (TLT 2015) 2015.