Recognition of Time Expressions in Spanish Electronic Health Records

Marjan Najafabadipour
Centro de Tecnología Biomédica
Universidad Politécnica de Madrid
Madrid, Spain
m.najafabadipour@upm.es
ORCID: 0000-0002-1428-9330

Massimiliano Zanin
Centro de Tecnología Biomédica
Universidad Politécnica de Madrid
Madrid, Spain
massimiliano.zanin@ctb.upm.es
ORCID: 0000-0002-5839-0393

Alejandro Rodríguez-González
Centro de Tecnología Biomédica
Universidad Politécnica de Madrid
Madrid, Spain
alejandro.rg@upm.es
ORCID: 0000-0001-8801-4762

Consuelo Gonzalo-Martín
Centro de Tecnología Biomédica
Universidad Politécnica de Madrid
Madrid, Spain
consuelo.gonzalo@upm.es

Beatriz Nuñez García
Hospital Universitario Puerta de Hierro Majadahonda
Madrid, Spain
beangarcia@gmail.com

Virginia Calvo
Hospital Universitario Puerta de Hierro Majadahonda
Madrid, Spain
vircalvo@hotmail.com

Juan Luis Cruz Bermudez
Hospital Universitario Puerta de Hierro Majadahonda
Madrid, Spain
jlcruz@idiphim.org

Mariano Provencio
Hospital Universitario Puerta de Hierro Majadahonda
Madrid, Spain
mprovenciop@gmail.com

Ernestina Menasalvas
Centro de Tecnología Biomédica
Universidad Politécnica de Madrid
Madrid, Spain
ernestina.menasalvas@upm.es
ORCID: 0000-0002-5615-6798

Abstract—**The widespread adoption of Electronic Health Records (EHRs) is generating an ever-increasing amount of unstructured clinical texts. Processing time expressions from these domain-specific-texts is crucial for the discovery of patterns that can help in the detection of medical events and building the patient's natural history. In medical domain, the recognition of time information from texts is challenging due to their lack of structure; usage of various formats, styles and abbreviations; their domain specific nature; writing quality; and the presence of ambiguous expressions. Furthermore, despite of Spanish occupying the second position in the world ranking of number of speakers, to the best of our knowledge, no Natural Language Processing (NLP) tools have been introduced for the recognition of time expressions from clinical texts, written in this particular language. Therefore, in this paper we propose a Temporal Tagger for identifying and normalizing time expressions appeared in Spanish clinical texts. We further compare our Temporal Tagger with the Spanish version of SUTime. By using a large dataset comprising EHRs of people suffering from lung cancer, we show that our developed Temporal Tagger, with an F1 score of 0.93, outperforms SUTime, with an F1 score of 0.797.**

Keywords—Electronic Health Records, Natural Language Processing, Named Entity Recognition, Time Expression Extraction

I. Introduction

Time, the concept that mankind associates to the changes in the world, is crucial in biomedical informatics and EHRs [1]. Clinicians chronologically record the progress of a disease or a hospital course in clinical texts and store this information with time points in EHRs. To identify medical events and build the natural history of the patient, the retrieval of time expressions from clinical texts is thus essential. In recent years, the annotation of such time information has received a great attention by researchers, due to the richness of time expressions, their importance in medical care, and the great availability of EHRs [2]. However, this information remains hidden within unstructured text of EHRs and requires the development of specific NLP techniques in order to be accessed.

Within the clinical context, time annotation presents four major challenges. First of all, the specific time notations can fall within three categories, i.e. natural, conventional and professional time (e.g. "*5 days ago*" vs. "*2011-09-16*" vs. "*24hr*"), each one presenting their own idiosyncrasies. Secondly, physician usually have limited time to write the details of patient-clinician encounters, and therefore have to resort to domain specific and non-standard expressions and abbreviations. This makes clinical texts hard to understand outside the medical community, let alone by automated systems. Furthermore, the presence of ambiguous expressions, having more than one semantical meaning in this domain-specific texts, adds an additional layer of complexity. Thirdly, the interpretation of relative time expressions can be uncertain, being difficult to automatically identify which time point a relative time refers to. Note that, as opposed to standard texts, relative time expressions are here more prevalent compared with absolute ones. Finally, granularity is an integral part of time information, both in absolute and relative expressions; it is thus not always clear what measurement unit should be adopted to represent different granularities. To illustrate, expressions like *"5 days ago"* or *"September"* are very common, and yet are difficult to decode in an exact way.

Although a considerable amount of research has been performed on processing time expressions in clinical texts, most of the existing systems focus on English texts and perform their annotation with the help of annotated corpora. The creation of such corpora is firstly costly and time consuming, and secondly, their completeness directly affects the processing quality. In addition, while Spanish has acquired the second position in the world ranking of number of speakers (with more than 572 million people) [3], to the best of our knowledge, no NLP system has been presented for processing time expressions in Spanish clinical texts.

Processing time information requires the ability to recognize its expressions and convert them from text to a normalized form, to simplify subsequent processing. Therefore, the main contribution of this paper is to introduce a rule-based Temporal Tagger, capable of extracting and normalizing time expressions written in Spanish clinical texts. When compared with a standard alternative tool, i.e. the Spanish version of SUTime, our solution presents a significantly higher F1 score, as we demonstrate by using a large collection of real EHRs. At the same time, it is both conceptually and technically simpler than machine learning solutions, as it does not require any annotated corpora.

The remainder of the paper is organized as follows. Firstly, the main related works are explained in Section II, with Section III focusing on SUTime, the main system we used for benchmarking. Afterwards, the details of our developed Temporal Tagger for annotation of time expressions in Spanish clinical texts are described in Section IV, and its validation using a large collection of EHRs is presented in Section V. Finally, Section. VI discusses the main results here presented, and Section. VII draws some conclusions and outlines about future lines of work.

II. Related Work

Within the general domain, numerous systems have been proposed for processing time expressions, in both English and Spanish texts, using rule-based and machine learning approaches. However, these systems are not generally flexible, i.e. they are not designed to work with the various styles and formats a date can be written in. Additionally, systems developed for processing time information from free texts in the general domain may not be efficient enough to be applied to the medical one. As discussed in [1], this statement is theoretically supported by the sub-language theory, which shows that a restricted domain is more well-defined than a general one, and can more accurately be characterized by specific vocabularies, semantic relations and, in some cases, syntax. For the medical domain, most of the state-of-the-art systems were developed using annotated corpora, provided in English by the shared tasks. Yet, the limited size of such corpora can significantly affect the quality of the processing. In addition, to the best of our knowledge, no system has been proposed for Spanish. For the sake of completeness, in this section, we review the most important of them.

Time expression recognition and normalization from free-form texts has seen a great deal of interest in the last decades, especially with the development of the TimeML annotation schema [4] and release of TimeBank [5] newswire corpus. The TimeBank corpus was used in three temporal analysis evaluation tasks in the SemEval competitions, TempEval-1 [6], TempEval-2 [7], and TempEval-3 [8]. In Temp-Eval-2 for Spanish language, the Temporal Information Processing based on Semantic information (TIPSem) [9] used Conditional Random Field (CRF) models for extracting time expressions, and applied CRF and rule-based methods for normalizing time information. While TIPSem achieved the best F1 score of 0.91, TIPSem-B and UC3M [10] obtained the second-best F1 score of 0.88. UC3M applied a rule-based approach for its implementations.

For English language in the context of TempEval-2, HeidelTime [11] gained the best F1 score of 0.86 by introducing a rule based system, which used regular expressions for extracting time expression and knowledge resources as well as linguistic clues for their normalization. TIPSem, TRIPS and TRIOS [12] were the second-best systems with F1 score of 0.85. TRIPS and TRIOS used a combination of deep semantic parsing, Markov Logic Networks, CRF classifiers and a set of rules for recognizing and normalizing time expressions.

A Perl temporal tagger was also developed as part of the TARSQI toolkit [13], named GUTime [14]. GUTime was built for processing time expressions appearing in English texts only. It was an extension of TempEx tagger [15], which handled both absolute and relative times and has been applied to different corpora, including broadcast news, print news, and meeting scheduling dialogs.

In addition, the Stanford university also built a rule-based temporal tagger, named SUTime [16] [17]. SUTime was developed upon regular expressions for recognizing and normalizing time expressions, written in both English and Spanish texts. Both SUTime and GUTime were evaluated on the English dataset of TempEval-2, achieving the F1 scores of 0.92 and 0.84, respectively. Thus, SUTime outperformed the tools presented in TempEval-2 for annotation of time expressions.

In TempEval-3, HeidelTime [18] was presented as a multi-lingual temporal tagger for extraction and normalization of time expressions mentioned in English and Spanish texts. For both languages, it achieved the highest F1 scores with values of 77.61 in English texts using HeidelTime-t and 90.1 in Spanish texts. In addition, the Spanish version of TIPSem, named as TIPSemB-Freeling (TIPSemB-F) was ranked as the second-best performing system with an F1 score of 87.4 in time expression tasks.

Within the medical domain, the extraction and normalization of time expressions has been the topic of many shared tasks over the past few years. Among them, it is worth highlighting the Integrating Biology and the Bedside (i2b2) NLP Challenge [19], clinical Temp-Eval 2015 [20], clinical Temp-Eval 2016 [21] and clinical Temp-Eval 2017 [22]. The i2b2 NLP Challenge provided the researchers with an English corpus of discharge summaries, which was annotated with temporal information in the year 2012. Using ISO-TimeML annotation guidelines and by implementing the regular expressions and machine learning methods, researchers were able to perform extraction and normalization of time information from clinical texts. Among all participants, the Mayo clinic system achieved the highest accuracy of 0.73 by using regular expressions for performing the time expression tasks.

In 2014, an extension of ISO-TimeML guidelines was developed to annotate an English corpus of clinical notes, provided by the Mayo clinic, known as Temporal Histories for Your Medical Events (THYME) corpus [23]. This dataset has hitherto been used in several competitions. In clinical Temp-Eval 2015, BlueLab [24] used SVM classifiers with features generated by the Apache clinical Text Analysis and Knowledge Extraction System (cTAKES). It obtained the highest F1 score of 0.709 for identifying the span and class (DATE, TIME, DURATION, QUANTIFIER, PREPOSTEXP or SET) of time expressions. In addition, in clinical Temp-Eval 2016, UTHealth [25] implemented linear and structural (HMM) SVMs using lexical, morphological, syntactic, discourse, and word representation features. Its run

1 gained the best F1 score of 0.772 for span and class of time information.

In the context of clinical Temp-Eval 2017, KULeuven-LIIR [26] used linear SVM classifiers with features including words and part-of-speech to find time expressions mentioned in clinical texts. KULeuven-LIIR achieved the highest F1 score of 0.53 for time span and class information in Unsupervised domain adaptation. Nevertheless, within the specific task of supervised domain adaptation, GUIR [27] achieved the best F1 score of 0.56 for time span and class information. GUIR used supervised learning algorithms with lexical, syntactic, semantic, distributional, and rule-based features.

Finally, The NLP system of cTAKES [28] [29] was also extended with a temporal module, employing forward and backward search algorithms and multiple learning methods, like SVM and CRF, for annotation of time expressions from clinical narratives written in English [30] [31].

III. PRELIMINARIES

As we here benchmark our proposed Temporal Tagger against the Spanish version of SUTime, for the sake of completeness, in this Section we provide an overview of the structure and performance of the latter. SUTime [16] [17], a pattern-based extraction annotator, is basically developed for retrieval of time information from English free texts. It supports four basic types of temporal expressions, namely TIME, DURATION, INTERVAL and SET. This annotator is implemented under the Stanford CoreNLP pipeline [32], which supports tokenization of free texts, making it convenient for SUTime to specify regular expression over tokens.

Given a tokenized text, to extract time expressions SUTime follows a three-fold strategy: (1) building patterns over individual words for recognizing numeric expressions; (2) using patterns over words and numerical expressions to recognize simple time expressions; and (3) forming compound patterns on discovered temporal expressions. Once time expressions are recognized, ambiguous expressions that are not likely to be time-related are removed from the list of candidates. Then, each of the expressions is associated to a temporal object. If there is a relative temporal object, it will be resolved based on document date. Also, if there is a confusion about the time point which the relative temporal object refers to (e.g., "*Monday*"), the verb tense of the clause is used to help in resolving the ambiguity. Finally, SUTime performs the internal time representation of all the temporal objects and produces a TIMEX3 [4] annotation for each temporal objects.

At a later stage, SUTime was extended for annotation and normalization of TIME expressions written in Spanish. However, this annotator contains a limited set of rules for performing the named entity recognition process in Spanish, does not support the disambiguation feature for removal of expressions that are not likely to be related to time, and does not deal with ambiguities about the time to which a relative expression object may refer.

IV. THE PROPOSED TEMPORAL TAGGER

Every time expression t (1) can be viewed as a two-tuple:

$$t = (ti,vi) \tag{1}$$

where ti is the time expression itself and vi is the normalized value. Our goal is to extract every time expression ti and to accurately assign the value attribute vi. For this purpose, we developed a rule-based NLP module using the Apache Unstructured Information Management Architecture (UIMA) framework, named Temporal Tagger. Given a tokenized and Part of Speech (PoS) labeled text, the Temporal Tagger is capable of processing time expressions, written in Spanish, within clinical texts. To provide the temporal tagger with a tokenized and PoS labeled text, we have used the NLP pipeline of the Clinical Knowledge Extraction System (C-liKES) [33]. C-liKES is a text mining system that has been developed on top of Apache UIMA framework.

The main steps of our Temporal Tagger are: (1) extraction of various time expressions; (2) filtration of time expressions; (3) resolution of time expressions with respect to a reference date; and (4) normalization of time expressions to a standard date format. The following sub-sections provide the detailed information about each one of these steps.

A. Extraction of various time expressions

Given a tokenized text, the Temporal Tagger identifies time expressions and outputs annotations for further manipulations and interpretations. Its output includes annotations in form of XML Metadata Interchange (XMI) files. The Temporal Tagger is capable of structuring natural (e.g., "*Hace 5 días*" meaning 5 days ago, "*Hoy*" meaning Today), conventional (e.g., "*Septiembre dieciséis, 2011*", meaning September sixteen, 2011) and professional (e.g., "*24hr*") time expressions, the three common ways in which a date can be written in Spanish. It also supports annotation of time expressions written in different formats and styles. For instance, some of them include DD-MM-YYYY, MM-DD-YYYY, YYYY-MM-DD, YYYY-DD-MM, DD-MM-YY, MM-DD-YY, dia (meaning of day) DD de mes (meaning of month) MM de año (meaning of year) YYYY, etc. Since our Temporal Tagger is optimized for Spanish, in which the standard time expression is written as DD-MM-YYYY or YYYY-MM-DD, we assign priority to these two rules over the alternative MM-DD-YYYY and YYYY-DD-MM. In addition, by various styles we mean numerical, alphabetical, mixed (alphabetical and numerical) or even abbreviated time expressions. For example, the date "*16/09/2011*" can be written in different formats and styles as "*09/16/2011*", "*16-9-2011*", "dia *16 de mes 9 de año 2011*", "*Septiembre dieciséis, 2011*", "*Sep 16, 2011*" etc. Finally, the Temporal Tagger supports the extraction of time points, indicating a particular instance on a time line.

B. Filtration of time expressions

Given a text labeled with PoS tags, the Temporal Tagger removes those time expressions from the list that are not likely to be such. For example, the word "*Tarde*" has two meanings in Spanish, "afternoon" and "late". In this specific case, we specified a rule according to which, if a candidate time expression is a single word "*Tarde*" and its PoS tag is not a noun, then it is not referring to a time point; hence, the Temporal Tagger ignores from annotating it as a time expression.

C. Resolution of time expressions with respect to a reference date

Processing relative time expressions (e.g., "*Hace 5 días*") requires a reference date on which the statement was made. The Temporal Tagger uses the section date (if any) or the document date of the EHR as references. For example, for an EHR with the document date of "*16/09/2011*", the Temporal

Tagger would resolve the date referred to by "*Hace 5 días*" into "*11/09/2011*". However, there could be confusion about the time point which an expression refers to. For example, for the time expression "*Martes*" (meaning Tuesday) from a reference date like "*16/09/2011*", it is not clear whether it refers to "*13/09/2011*" (i.e. in the past) or "*20/09/2011*" (i.e. in the future). In this case, the verb tense of the sentence is used to resolve the ambiguity. However, since the clinical text is more compact and may not include a verb and the clinical narratives mostly provide information about past, then "*Martes*" refers to "*2011-09-13*" by default.

D. Normalization of time expressions to a standard date format

Once the time expressions have been identified and processed, the Temporal Tagger normalizes them into the standard date format of YYYY-MM-DD. For example, for an extracted time expression of "*Septiembre dieciséis, 2011*", the Temporal Tagger converts the date to "*2011-09-16*". However, if the month and/or the day of the time expression is not identified, the Temporal Tagger automatically assigns the value 00 to the corresponding MM or DD value. Note that this value is used to record the uncertainty about the exact date. For example, for the time expression "*Septiembre, 2011*", the Temporal Tagger would convert the date to "*2011-09-00*". In Table. I, we give examples of some expressions recognized and normalized by our Temporal Tagger.

V. Validation

The evaluation of our developed Temporal Tagger consists of two use cases: (1) verification of the output of the Temporal Tagger; and (2) comparison of the Temporal Tagger with the Spanish version of SUTime. The following sub-sections discuss the details of the dataset used in these studies, the conducted experiments and their outputs.

A. Dataset

The dataset used in our experiments includes information for 749 patients suffering from lung cancer. This dataset contains 199,835 EHRs, which are written in Spanish and are provided by the Hospital Puerta de Hierro Majadahonda (HUPHM) of Madrid. These EHRs are divided into two main categories, clinical notes (191,311 of them) and clinical reports (8,524 of them). Clinical notes can be generated for a patient by different services and personnel in the hospital at each patient's hospital visit. They are always written by a professional like physicians, nurses, social services people, etc. Clinical notes contain highly detailed information about the patient's personal and clinical status, processes followed and their results, and the services visited by the patient. On the other hand, clinical reports are generated when a medical process is completed, and they provide a summary of the corresponding clinical notes. Compared to the latter ones, they have a more structured format as they contain different sections (e.g., Personal History, Family Oncological History, Diagnosis, Treatment, etc.) under which the relevant information is provided.

Due to the huge amounts of EHRs provided by HUPHM, it was not feasible to perform a manual validation on the entire dataset. Therefore, to conduct our experiments, we have decided to perform a random selection of 100 EHRs from the original dataset, including 50 clinical notes and 50 clinical reports. The selection of equal number of notes and reports was aimed at keeping both types of clinical documents significantly present in the validation phases.

TABLE I. Examples of time expressions annotated by our Temporal Tagger (with 2016-12-23 as the reference date)

Time Expression	Meaning	Normalized value
16/09/2011	16/09/2011	2011-09-16
Septiembre dieciséis, 2011	September sixteen, 2011	2011-09-16
Hace 5 días	5 days ago	2016-12-18
Hoy	Today	2016-12-23
Pasado mañana	The day after tomorrow	2016-12-25
Martes por la noche	Tuesday night	2016-12-20
Anoche	Last nigh	2016-12-22
Septiembre	September	2016-09-00
2011	2011	2011-00-00
24hr	24 hour	2016-12-22

Using chi-squared tests, we have also performed a set of statistical tests on the selected sample, to assess its representativeness of the entire population in the original dataset. These chi-squared tests were performed on the patient's sex, age (categorical variable: < 35, 35-40, 45-55, 55-65, 65-75, 75-80, 80 >), stage of tumor, and local and systemic progression of the tumor. In all cases, except for the systemic progression, the sample dataset was representative of the entire population (p-value < 0.01). However, a small amount of bias was observed in the systematic progression of tumor between the patients of the selected sample and the patients of the original sample; note that this is not expected to affect the types of temporal expressions found.

B. Experiments

Two computer scientists, who were native Spanish speakers, served as evaluation domain experts under the supervision of clinicians from the HUPHM and helped in conducting the experiments. None of them participated in the design or development of the Temporal Tagger.

1) First use case – verification of the output of the Temporal Tagger

In our first use case, the evaluation was done by manually analyzing the output provided by the Temporal Tagger. For each time expression written in the free text of EHRs, a comparison was done between the list of the expressions automatically provided and the list of expressions manually extracted. After the comparison was completed, confusion matrices were calculated, for then obtaining the True Positive (TP), False Positive (FP) and False Negative (FN) values, in order to determine precision, recall and F1 score values. To obtain TP, FP and FN, for each time expression appeared in the clinical texts, the evaluation domain experts rated:

- TP: if the time expression was correctly classified and normalized by the Temporal Tagger.
- FP: if the time expression was incorrectly classified or normalized by the Temporal Tagger.
- FN: if the Temporal Tagger did not classify the time expression when it should have.

Given TP, FP and FN, the following equations were used for determining precision (2), recall (3) and F1 score (4):

$$Precision = TP / (TP + FP) \quad (2)$$

$$Recall = TP / (TP + FN) \quad (3)$$

$$F1 = (2 \times Precision \times Recall) / (Precision + Recall) \quad (4)$$

In the case of precision and recall, confidence intervals were calculated by considering a binomial distribution, with confidence levels of 95%.

2) Second use case – comparison of the Temporal Tagger with the Spanish version of SUTime

The idea of our second use case it to benchmark our solution with an existing one, specifically the Spanish version of SUTime. By implementing the first use case, the results of our Temporal Tagger were obtained. To measure the accuracy of SUTime for retrieval of time variables, the first use case was also repeated by the evaluation domain experts on the same dataset of 100 EHRs. Finally, the results were manually analyzed, and a comparison was performed between our Temporal Tagger and SUTime.

C. Results

The results of the first use case show that the Temporal Tagger achieves a precision of 0.927 ± 0.021, recall of 0.932 ±0.021 and F1 score of 0.93. To find the errors occurred in the annotation by our Temporal Tagger, we analyzed its output extensively. By examining FPs, we realized that the majority of errors are caused by incorrect normalization of relative time expressions. The main reasons for such errors were: (1) usage of compact clinical sentences without mention of verbs in them. As a result, the Temporal Tagger referred the relative time expressions to the past time while they were referring to the future time; (2) existence of ambiguous time variables such as "*Mañana*", which has two meanings in Spanish, "morning" and "tomorrow". This leads to incorrect normalization of these time variables in some cases; and (3) few errors occurred due to limitation of the Temporal Tagger to cover time durations like "*esta semana*", which means "this week", or "*este año*", which means "this year". The analysis of FNs revealed that most of the errors were the consequence of mentioning combined time expressions (e.g., "*1, 4, 8 y 14-sep*") and the usage of dots "." within time expressions (e.g., "*17.1.14*"). As the C-liKES tokenizer uses dots as the indicators of end of a sentence, this precludes the Temporal Tagger to capture the complete pattern.

The results of our second use case are discussed here. As it can be seen in Fig. 1, the experimental results show that our Temporal Tagger outperformed SUTime in terms of precision, recall and F1 score. SUTime obtained the precision of 0.831 ±0.033, recall of 0.766 ±0.036 and F1 score of 0.797. The main differences were found in the recognition of the various formats a time expression can be written in, as well as in the annotation of natural and professional time expressions. Our Temporal Tagger had a better performance in processing such time variables compared to SUTime. Regarding the detection of numerical conventional time expressions, both taggers have shown closely equal performances, but for the annotation of mixed and abbreviated conventional time variables, the Temporal Tagger showed a higher accuracy. In addition, by assigning priorities to the rules of DD-MM-YYYY and YYYY-MM-DD and by performing filtration, our Temporal

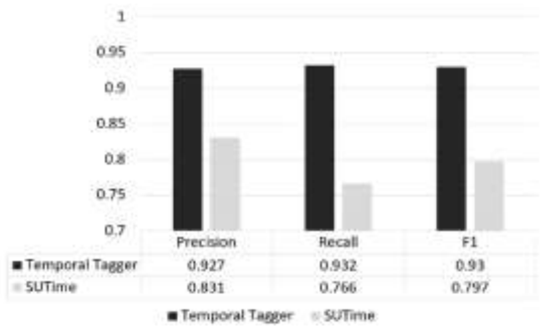

Fig. 1. Comparison of the results of our Temporal Tagger with SUTime

Tagger obtained more precise normalization results compared to SUTime.

VI. Discussion

In synthesis, all previous results support the idea that our Temporal Tagger had higher precision, recall and F1 score than the Spanish version of SUTime in retrieving time expressions from clinical texts. We applied rule-based approaches for extraction and normalization of time variables similar to SUTime, which is introduced as the best performing tool compared to the systems presented in TempEval-2.

We observed that by supporting regular expressions for annotation of various styles and formats a time variable can be written in, and by supporting natural, conventional and professional time expressions, our Temporal Tagger has obtained a high accuracy. In addition, the assignment of priorities to some rules for Spanish language and the processes of filtration and resolution with respect to a reference date played a great role in providing accurate normalized values.

However, we have also seen that the relative time expressions were the most difficult ones for our Temporal Tagger to normalize. Also, the tagger faced some errors due to its limitation to annotate the duration, the combined time expressions and those time variables that used dot in their format, like DD.MM.YYY.

Finally, our aim was to benchmark the existing SUTime tool and rule-based methods to annotate time expressions from clinical texts written in Spanish. Adaptions of machine learning based methods requires annotated corpora, whose building is both time-consuming and costly. In addition, the small size of their annotated corpora could unavoidably affect the processing quality.

VII. Conclusion and future work

Annotation of time expressions is essential to the task of investigating problems requiring temporal information, such as event extraction and temporal ordering of events. Before using time expressions in any other applications, the first step is to extract and normalize time variables appearing in free texts. Therefore, in this paper, we presented a Temporal Tagger, capable of processing time information appearing in Spanish clinical texts. As one of our primary objectives was to yield a generic Temporal Tagger that could be applied in other scenarios with minimal adaptions, it has been developed as an UIMA component. Beyond what here presented, future works will be focused on improvements and integration. On one hand, improvements will include the annotation of combined time expressions and durations from Spanish clinical texts. On the other hand, we plan on building a complete NLP pipeline with the Temporal Tagger being a part of it, capable of detecting medical events and reconstructing the patient's natural history.

VIII. Acknowledgment

This paper is supported by European Union's Horizon 2020 research and innovation programme under grant agreement No. 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.

References

[1] L. Zhou, C. Friedman, S. Parsons, and G. Hripcsak, "System architecture for temporal information extraction, representation and reasoning in clinical narrative reports," *AMIA Annu. Symp. Proc. AMIA Symp.*, pp. 869–873, 2005.

[2] G. Moharasar and T. B. Ho, "A semi-supervised approach for temporal information extraction from clinical text," in *2016 IEEE RIVF International Conference on Computing Communication Technologies, Research, Innovation, and Vision for the Future (RIVF)*, 2016, pp. 7–12.

[3] J. Manuel Bonet, R. Rodríguez-Ponga Salamanca, J. María Martínez Alonso, R. Bueno Hudson, M. López-Vega, and D. Fernández Vítores, "EL ESPAÑOL: UNA LENGUA VIVA.," Instituto Cervantes, 2018.

[4] J. Pustejovsky *et al.*, "TimeML: Robust Specification of Event and Temporal Expressions in Text," in *New directions in question answering*, 2003, vol. 3, pp. 1–12.

[5] J. Pustejovsky *et al.*, "The TIMEBANK Corpus," in *Proceedings of Corpus Linguistics*, 2003, pp. 647–656.

[6] M. Verhagen, R. Gaizauskas, F. Schilder, M. Hepple, G. Katz, and J. Pustejovsky, "SemEval-2007 task 15: TempEval temporal relation identification," in *Proceedings of the 4th International Workshop on Semantic Evaluations - SemEval '07*, Prague, Czech Republic, 2007, pp. 75–80.

[7] M. Verhagen, R. Sauri, T. Caselli, and J. Pustejovsky, "SemEval-2010 Task 13: TempEval-2," in *Proceedings of the 5th International Workshop on Semantic Evaluation*, Uppsala, Sweden, 2010, pp. 57–62.

[8] N. UzZaman, H. Llorens, L. Derczynski, J. Allen, M. Verhagen, and J. Pustejovsky, "SemEval-2013 Task 1: TempEval-3: Evaluating Time Expressions, Events, and Temporal Relations," in *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, Atlanta, Georgia, USA, 2013, pp. 1–9.

[9] H. Llorens, E. Saquete, and B. Navarro, "TIPSem (English and Spanish): Evaluating CRFs and Semantic Roles in TempEval-2," in *Proceedings of the 5th International Workshop on Semantic Evaluation*, Uppsala, Sweden, 2010, pp. 284–291.

[10] M. T. Vicente-Díez, J. M. Schneider, and P. Martínez, "UC3M system: Determining the extent, type and value of time expressions in TempEval-2," presented at the SemEval '10 Proceedings of the 5th International Workshop on Semantic Evaluation, Los Angeles, California, 2010, pp. 329–332.

[11] J. Strötgen and M. Gertz, "Heideltime: High quality rule-based extraction and normalization of temporal expressions," in *In Proceedings of the 5th International Workshop on Semantic Evaluation, Sem-Eval*, 2010, pp. 321–324.

[12] N. UzZaman and J. Allen, "TRIPS and TRIOS System for TempEval-2: Extracting Temporal Information from Text," in *Proceedings of the 5th International Workshop on Semantic Evaluation*, Uppsala, Sweden, 2010, pp. 276–283.

[13] M. Verhagen and J. Pustejovsky, "Temporal Processing with the TARSQI Toolkit," in *COLING '08 22nd International Conference on on Computational Linguistics: Demonstration Papers*, Manchester, United Kingdom, 2008, pp. 189–192.

[14] I. Mani, "Recent Developments in Temporal Information Extraction," in *Recent Advances in Natural Language Processing III*, Borovets, Bulgaria, 2003, pp. 45–60.

[15] I. Mani and G. Wilson, "Robust Temporal Processing of News," in *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, Hong Kong, 2000, pp. 69–76.

[16] A. X. Chang and C. D. Manning, "SUTIME: A Library for Recognizing and Normalizing Time Expressions," in *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*, 2012.

[17] "The Stanford Natural Language Processing Group." [Online]. Available: https://nlp.stanford.edu/software/sutime.html. [Accessed: 27-Sep-2018].

[18] J. Strötgen, J. Zell, and M. Gertz, "HeidelTime: Tuning English and Developing Spanish Resources for TempEval-3," in *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, Atlanta, Georgia, USA, 2013, pp. 15–19.

[19] W. Sun, A. Rumshisky, and O. Uzuner, "Evaluating temporal relations in clinical text: 2012 i2b2 Challenge," *J. Am. Med. Inform. Assoc.*, vol. 20, no. 5, pp. 806–813, Sep. 2013.

[20] S. Bethard, L. Derczynski, G. Savova, J. Pustejovsky, and M. Verhagen, "SemEval-2015 Task 6: Clinical TempEval," in *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, Denver, Colorado, 2015, pp. 806–814.

[21] S. Bethard, G. Savova, W.-T. Chen, L. Derczynski, J. Pustejovsky, and M. Verhagen, "SemEval-2016 Task 12: Clinical TempEval," in *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, San Diego, California, 2016, pp. 1052–1062.

[22] S. Bethard, G. Savova, M. Palmer, and J. Pustejovsky, "SemEval-2017 Task 12: Clinical TempEval," in *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, Vancouver, Canada, 2017, pp. 565–572.

[23] W. F. S. Iv *et al.*, "Temporal Annotation in the Clinical Domain," *Trans. Assoc. Comput. Linguist.*, vol. 2, no. 0, pp. 143–154, Apr. 2014.

[24] S. Velupillai, D. L. Mowery, S. Abdelrahman, L. Christensen, and W. Chapman, "BluLab: Temporal Information Extraction for the 2015 Clinical TempEval Challenge," in *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, Denver, Colorado, 2015, pp. 815–819.

[25] H.-J. Lee *et al.*, "UTHealth at SemEval-2016 Task 12: an End-to-End System for Temporal Information Extraction from Clinical Notes," in *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, San Diego, California, 2016, pp. 1292–1297.

[26] A. Leeuwenberg and M.-F. Moens, "KULeuven-LIIR at SemEval-2017 Task 12: Cross-Domain Temporal Information Extraction from Clinical Records," in *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, Vancouver, Canada, 2017, pp. 1030–1034.

[27] S. MacAvaney, A. Cohan, and N. Goharian, "GUIR at SemEval-2017 Task 12: A Framework for Cross-Domain Clinical Temporal Information Extraction," in *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, Vancouver, Canada, 2017, pp. 1024–1029.

[28] G. K. Savova *et al.*, "Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications," *J. Am. Med. Inform. Assoc.*, vol. 17, no. 5, pp. 507–513, Sep. 2010.

[29] "Apache cTAKES™ - clinical Text Analysis Knowledge Extraction System." [Online]. Available: http://ctakes.apache.org/. [Accessed: 11-Apr-2018].

[30] G. Savova *et al.*, "Towards Temporal Relation Discovery from the Clinical Narrative," *AMIA. Annu. Symp. Proc.*, vol. 2009, pp. 568–572, 2009.

[31] "cTAKES 4.0 - Temporal Module - Apache cTAKES - Apache Software Foundation." [Online]. Available: https://cwiki.apache.org/confluence/display/CTAKES/cTAKES+4.0+-+Temporal+Module. [Accessed: 14-May-2018].

[32] "Stanford CoreNLP – Natural language software | Stanford CoreNLP." [Online]. Available: https://stanfordnlp.github.io/CoreNLP/. [Accessed: 20-Dec-2018].

[33] E. Menasalvas Ruiz *et al.*, "Profiling Lung Cancer Patients Using Electronic Health Records," *J. Med. Syst.*, vol. 42, no. 7, p. 126, May 2018.