TimeLex: A Suite of Tools for Processing Temporal Information in Legal Texts

María Navas-Loro(✉) 📧 and Víctor Rodríguez-Doncel 📧

Ontology Engineering Group, Universidad Politécnica de Madrid, Madrid, Spain
mnavas@fi.upm.es

Abstract. In this paper we present a suite of tools named TimeLex, that includes different systems able to process temporal information from legal texts. The first tool, called *lawORdate*, helps preprocessing legal references in texts in Spanish that can be misleading when trying to find dates in texts. The second one, *Añotador*, is a temporal tagger (this is, a tool that finds temporal expressions, such as dates or durations) that identifies temporal expressions in texts and provides a standard value for each of them. Finally, a third tool, called *WhenTheFact*, extracts relevant events from judgments, allowing a full processing of the temporal dimension of this kind of texts, and being a first step towards the complete temporal information processing in the legal domain.

Keywords: Temporal expressions · Events · Timeline generation · Legal texts

1 Introduction

Temporal information is a very important dimension in documents. Being able to extract it would enable higher level functionalities, such as event-based summarization or search, pattern detection in cases, and timeline generation, that would facilitate the understanding of legal documents, usually difficult to comprehend by layman users, as well as enhance other NLP tasks over legal documents. Nevertheless, not a lot of research has been done in the legal domain in the field of temporal information.

TimeLex [1] is a suite of tools that aims to cover this gap in the domain, providing approaches to several parts of the temporal information extraction task. In this paper we briefly present the different contributions we have created in order to process this kind of texts from the temporal perspective.

The remaining of this paper is structured as follows. Section 2 presents related work in previous literature. Section 3 introduces *lawORdate*, a preprocessing tool that deals with legal references in order to facilitate latter temporal tagging task. Section 4 presents *Añotador*[1], a temporal tagger designed to find and normalize temporal expressions in legal texts. Section 5 shows a first approach for event extraction, introducing the

---

[1] *Añotador* is a pun: "Año" means "Year" in Spanish, while "Anotador" is the person or tool that performs the task of annotation. *Añotador* is a merge of the two concepts, and would therefore can be understood as "*What annotates years*".

tool *WhenTheFact*, which is able to generate a timeline of events from the information extracted from a document from the European Court of Human Rights[2]. Finally, Sect. 6 presents the conclusions and details the next steps in this research, targeting the semantic representation of the temporal annotations for further applications.

2 Related Work

Most effort related to temporal information in the legal domain has been done in relation to normative texts. This is the case of CronoLex [2], that aims to help lawyers by representing the legal norms in Spanish storing information about their life cycle, among others. Also in this direction, Akoma Ntoso [3] allows to represent several types of legal text in a standard way, including temporal information in the metadata.

Regarding the processing of temporal expressions and events, Schilder [4] analyzed the different types of legal documents with regard to temporal information, and divided them in statutes or regulations (where temporal information usually are constraints), transactional documents (including documents for legal transactions like contracts) and case law. In this paper, Schilder deeply studies the two first types of legal documents, but case law narrative structure was considered similar to the narratives in news, and received no dedicated attention.

Again in normative texts, Isemann et al. [5] used Named Entitiy Recognition and temporal processing in order to process the temporal dimension of regulations. This work, on the other hand, also described usual problems found by temporal taggers find in legal texts (not only in normative texts). Among them we can highlight the similar pattern of legal references and dates, that tend to be misleading to temporal taggers (e.g. "Directive 2012/33/EC"), or the distinction between *generic* events and *episodic* events. While the first refer to abstract events, general truths, rules, expectations or laws, *episodic* events are those that actually happened. Finally, also works on transactional documents [6, 7] and reasoning in legal evidence [8] can be found in literature.

3 LawORdate

lawORdate is a tool that cleans legal references with a date form from text documents. It addresses an important problem when processing legal documents from the temporal perspective, since common legal references in Spanish tend to include dates or patterns that can be misleading to temporal taggers. For instance, in the following excerpt:

> *".. creado via el Real Decreto **2093/2008**$_1$, de **19 de diciembre**$_2$. Ha sido actualizado por ultima vez el __13 de agosto de 2017__$_3$."*

Most temporal taggers would find in this excerpt the three expressions in bold. Nevertheless, expression number one is not a date (despite of following a date-ish pattern) and expression number two is a date but does not belong to the narrative of the text (is part of a legal reference), so they should not be tagged. Therefore, the only one that should be tagged is the one underlined.

---

[2] https://www.echr.coe.int/Pages/home.aspx?p=home.

LawORdate is currently available both as a webapp [9] and as a GitHub repository [10], and finds and replaces misleading legal references in the texts, storing the original references. Once the temporal tagging is done, the references are restored in the text. Figure 1 shows the pipeline of use of lawORdate.
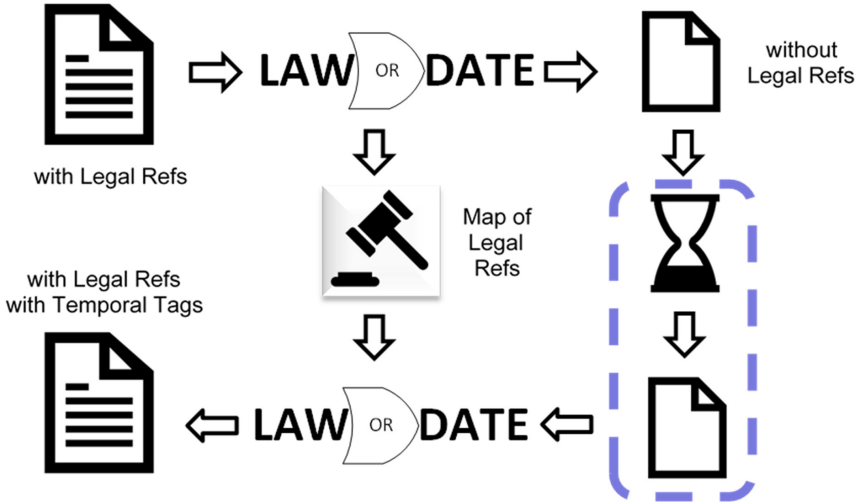

Fig. 1. Pipeline of use of lawORdate.

In the pipeline in Fig. 1, a text with legal references is first sent to the service. Then it finds all the misleading legal references that could affect to the precision of a temporal tagger and replaces them with inoquous expressions, storing the original references for further restoring. The output of this first step is to be used in a temporal tagger (in the demo, *HeidelTime* is offered, but any other can be used). Then, the output of the tagger (in TimeML) is sent back to *lawORdate*, that restores the original legal references. We therefore obtain the original text, but tagged without the interference of any legal references in it.

4   Añotador

Añotador [11] is a temporal tagger for Spanish and English able to find temporal expressions in texts, specially targeted to the legal domain. *Añotador* can detect different types of temporal expressions included in the TimeML standard, namely *dates*, *times*, sets (this is, expressions that repeat over time such as "*every Thursday*" or "*twice a week*") and *durations*, and some additional temporal expressions developed for the legal domain, such as specific expressions (e.g., "*business days*") and the type *interval*. *Añotador* outperforms the available state-of-the-art temporal taggers for Spanish [12]. It receives as input the text to annotate and optionally a reference date, called *anchor date* (if no date were introduced, the current date would be considered). With this information, the system is able to both find and normalize temporal expressions, this is, express them as a

standard value, usually normalizing with regard to a reference date. If we had for instance the sentence "*I went to the park **yesterday***" with "2019-09-20" as reference date, we would consider that the *normalized value* of '*yesterday'* is "2019-09-19". Nevertheless, not every temporal expression is normalized with regard to this initial *anchor date*. Once the temporal expressions in the text are identified using some hand-made rules specifically developed for the Spanish language, we apply a normalization algorithm that takes into account previous dates in the text for normalizing temporal expressions (see Fig. 2).
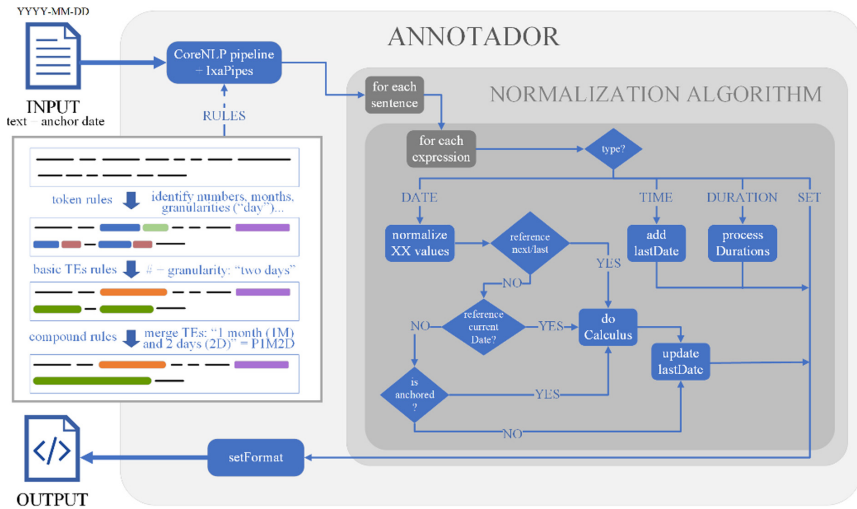
Fig. 2. Pipeline of *Añotador*. The user introduces the text to annotate and optionally a reference date.

Figure 2 shows how the text is first preprocessed using *CoreNLP* [13] and some *IxaPipes* [14] models. Then, different rules apply at different stages in order to detect the temporal expressions in the text. Once we have them, a normalization algorithm is applied in order to find their value. Finally, the system returns the text tagged.

Añotador has been tested against different state-of-the-art temporal taggers, both for legal English and for the Spanish language. Updated results of these evaluations can be found in its website [11].

5   WhenTheFact: Dealing with Events

After being able to identify temporal expressions using *Añotador* and *lawORdate*, the next logical step would be to detect events. Our current work focuses in detecting legal events in judgments, not covering just in the mention of the event (as most temporal taggers do), but also considering all the surrounding information available, such as the parts involved, when and where it happened or the jurisdiction involved. This was already done for a different type of legal document in a previous work, detecting events related to the lifecycle of a contract [15], but while in that case a rule-based approach was

successful, taking into account the limited amount of events targeted, for legal judgments the amount of relevant events demands a more flexible approach.

To this aim, we are considering different lines of research in parallel, in order to detect the different types of events we can find in a judgment (e.g., the facts under judgment, that change from case to case, and the legal events, such as applications or decisions, that are court-related and tend to occur in all cases). To test the different approaches, a corpus of legal documents annotated with events (the first publicly available of its kind as far as the authors know) has been built [16], in collaboration with experts from other institutions, based on previous related works [17, 18].
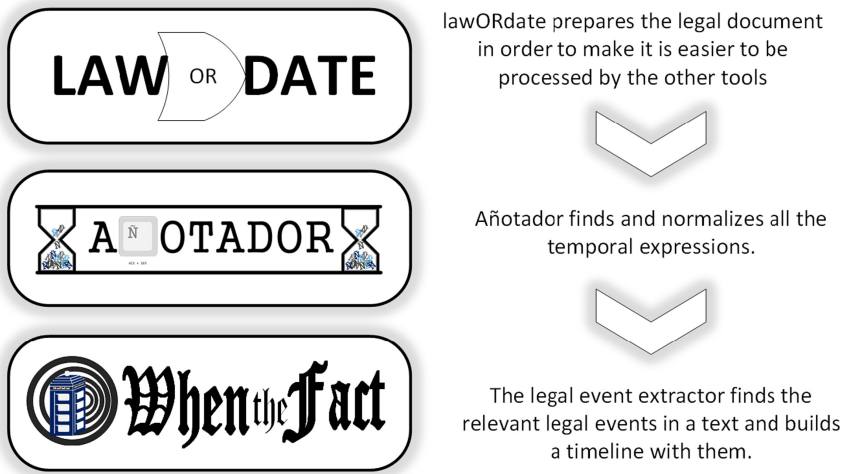
Fig. 3. Different tools available in for temporal processing of legal texts.

In Fig. 3, first, the text is preprocessed by *lawORdate*; then, the temporal expressions can be more accurately found by *Añotador* (once done, *lawORdate* would return the original legal references to the text). Finally, *WhenTheFact* detects the relevant events in the text (current online implementation already includes *Añotador* in order to perform the full processing).

6   Conclusions

In this paper we presented the different tools created for processing temporal information from legal texts. The first service introduced, *lawORdate*, "cleans" the document of misleading legal references; then, *Añotador* is able to tag and normalize the temporal expressions in the text. *WhenTheFact* detects events and builds a timeline from it. The suite therefore covers a full processing from the temporal perspective.

Although *WhenTheFact* is still an ongoing tool, able just to extract events from very specific types of texts whose structure is already known and with room for improvement. Additionally, *WhenTheFact* builds a timeline, but we consider further applications would

be useful for the legal domain, such as event-based summarization or pattern recognition. To facilitate these potential applications, represent temporal information in a standard and NLP focused manner would be extremely helpful. For this reason, next steps include the definition of this representation option by gathering different already available ontologies and schemas. Additionally, *WhenTheFact* is currently being expanded to cover more languages and types of documents. All advances in these directions will be reflected in the website of *TimeLex* [1].

References

1. Website of TimeLex. https://mnavasloro.github.io/timelex/
2. Marin, R.H., Hernandez, J.L., Delgado, J.J.I.: Cronolex: a computing representation of law dynamics. In: XXII World congress of Philosophy of Law and Socialphilosophy, Granada, Spain (2005)
3. Palmirani, M., Vitali, F.: Akoma-Ntoso for legal documents. In: Sartor, G., Palmirani, M., Francesconi, E., Biasiotti, M. (eds.) Legislative XML for the Semantic Web. Law, Governance and Technology Series, vol. 4, pp. 75–100. Springer, Dordrecht (2011). https://doi.org/10.1007/978-94-007-1887-6_6
4. Schilder, F., McCulloh, A.: Temporal information extraction from legal documents. In: Dagstuhl Seminar Proceedings. Schloss Dagstuhl-Leibniz-Zentrum für Informatik (2005)
5. Isemann, D., Ahmad, K., Fernando, T., Vogel, C.: Temporal dependence in legal documents. In: Yin, H., et al. (eds.) IDEAL 2013. LNCS, vol. 8206, pp. 497–504. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-41278-3_60
6. Guda, V., Srujana, I., Naik, M.V.: Reasoning in legal text documents with extracted event information. Int. J. Comput. Appl. **28**(7), 8–13 (2011)
7. Ramakrishna, K., Guda, V., Padmaja Rani, B., Chakati, V.: A novel model for timed event extraction and temporal reasoning in legal text documents. Int. J. Comput. Sci. Eng. Surv. **2**, 39–48 (2011)
8. Vlek, C.S., Prakken, H., Renooij, S., Verheij, B.: Representing and evaluating legal narratives with subscenarios in a Bayesian network. In: 2013 Workshop on Computational Models of Narrative. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik (2013)
9. LegalWhen demo. http://legalwhen.appspot.com/
10. LawORDate GitHub. https://github.com/mnavasloro/LawORDate
11. Añotador website. https://annotador.oeg.fi.upm.es/
12. Navas-Loro, M., Rodríguez-Doncel, V.: Annotador: a temporal tagger for Spanish. J. Intell. Fuzzy Syst. **39**(2), 1979–1991 (2020)
13. Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J.R., Bethard, S., McClosky, D.: The Stanford CoreNLP natural language processing toolkit. In: Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55–60 (2014).
14. Agerri, R., Bermudez, J., Rigau, G.: IXA pipeline: efficient and ready to use multilingual NLP tools. In: LREC, vol. 2014, pp. 3823–3828, May 2014
15. Navas-Loro, M., Satoh, K., Rodríguez-Doncel, V.: ContractFrames: bridging the gap between natural language and logics in contract law. In: Kojima, K., Sakamoto, M., Mineshima, K., Satoh, K. (eds.) JSAI-isAI 2018. LNCS, vol. 11717, pp. 101–114. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-31605-1_9
16. Filtz, E., Navas-Loro, M., Santos, C., Polleres, A., Kirrane, S.: Events matter: extraction of events from court decisions. In: Serena, V., Harasta, J., Kremen, P. (eds.) Legal Knowledge and Information Systems – JURIX 2020: The Thirty-third Annual Conference, Brno, Czech Republic, December 9–11 2020, volume 334 of Frontiers in Artificial Intelligence and Applications, pp. 33–42. IOS Press (2020)

17. Navas-Loro, M., Santos, C.: Events in the legal domain: first impressions. In: Proceedings of the 2nd Workshop on Technologies for Regulatory Compliance co-located with the 31st International Conference on Legal Knowledge and Information Systems (JURIX 2018), Groningen, The Netherlands, pp. 45–57, 12 December 2018 (2018)
18. Navas-Loro, M., Filtz, E., Rodríguez-Doncel, V., Polleres, A., Kirrane, S.: TempCourt: evaluation of temporal taggers on a new corpus of court decisions. Knowl. Eng. Rev. **34** (2019)