Mining, Representation and Reasoning with Temporal Expressions in the Legal Domain

María Navas-Loro*

Ontology Engineering Group, Universidad Politcnica de Madrid, Madrid, Spain

Abstract. Reasoning on temporal references present in free text documents is a challenging problem demanding the synergy of several disciplines in Artificial Intelligence. First, Natural Language Processing (NLP) techniques are necessary to mine temporal expressions (referring to instants and events) and their relations; then, knowledge representation systems able to support more than trivial temporal information must be defined; finally, temporal logic systems exposing practical functionality are much demanded. Documents in the legal domain are rich in temporal expressions, often linked to a norm. Permissions, obligations and prohibitions are often bounded by temporal constraints. While there exist different rules and policy languages capable of representing temporal constraints to some extent, they are far from being able to express the complex temporal relations present in these legal texts. Reaching a formalized expression of temporal rules able to accommodate the findings of the NLP algorithms and enabling temporal reasoning in an integrated manner would unleash new possibilities in the area of legal reasoning. As this thesis is still in an early phase, Doctoral Consortium feedback will be extremely valuable for topic delimitation and additional research lines.

Keywords: Temporal Constraints · Temporal Reasoning · Temporal Representation · Legal Domain · Natural Language Processing

1 Motivation

We live in a world where a lot of public documents are freely accessible but also hard to understand for a human due to concrete and technical expressions existing in the legal domain. Such is the case of Eur-Lex[1], a portal full of publicly-accessible legal documents of the European Union available in several languages.

Already uneasy for humans, making machines to understand legal documents is a challenging and multidisciplinary task involving several fields of Artificial

* This Ph.D. thesis is supervised by Dr. Asunción Gómez-Pérez and Dr. Víctor Rodríguez-Doncel, and it has been partially supported by LPS-BIGGER (IDI-20141259, Ministerio de Economía y Competitividad) and a research assistant grant by the Consejo de Educación, Juventud y Deporte de la Comunidad de Madrid partially founded by the European Social Fund (PEJ16/TIC/AI-1984).

[1] `http://eur-lex.europa.eu/homepage.html`

Intelligence: on the one hand, Natural Language Processing is required to analyze documents and queries to extract relevant information; this information needs to be properly represented. Reasoning methods are also required to answer to queries considering collected information. Additionally, human intervention is required to detect the main lacks on current reasoners and the usual legal understanding issues; human contribution is also useful for final validation.

Focusing on the legal domain, treatment of time remains as one of main challenges, as confirmed by a recent contribution to the field, the MIREL project[2]. Context information, such as the current date at the moment of the query, is also important when reasoning with temporal constraints. Being aware of the high complexity of the problem, we consider that the way to address it involves the integration of all the steps in a single framework, from document processing to reasoning, since they feed off intrinsically one another. This Ph.D. thesis will, therefore, propose new representation and reasoning options, along with resources related to time processing (such as sets of rules or tagged corpora), in order to build a framework able to process, represent and reason on legal documents to answer queries related to time.

2 State of the Art

Temporal reasoning when applied to the legal domain has not got a lot of attention lately. Some knowledge bases such as Yago [13] have some mechanisms to process this kind of information, but one of the last deep studies about the role of time in Legal Reasoning was performed in the late 90's [27]. Reasoning on time requires research on multiple fields, highlighting among them the representation to be used (along with the possible restrictions applicable), the temporal reasoning mechanisms existing and the linguistic processing needed to obtain the information. The state of the art presented is therefore divided in Natural Language Processing of temporal expressions, Representation of temporal expressions and Temporal Reasoning.

2.1 Natural Language Processing of temporal expressions

There are several aspects intervening in NLP of temporal expressions, such as annotation and temporal relation extraction and classification. Works in the latter [8] usually rely on aspects like tense or modality of events and concrete keywords, using both Machine Learning and knowledge-based methods to tackle it [6]. Regarding annotation, the most recent standard is TimeML [22]; although this is a generic language, sometimes it requires modification for its application to a concrete domain.

Despite the major boom in NLP in recent times, both Legal and Temporal NLP present still open issues, since recent achievements have not been applied in these fields. Even when temporal tagging is currently covered by several tools (such as HeidelTime [23] or the Tarsqi Toolkit[3]), in the annotation step (and

[2] http://www.mirelproject.eu/publications/D1.1.pdf

[3] http://www.timeml.org/tarsqi/toolkit/index.html

regarding already the reasoning phase) implicit knowledge must also be inferred, made explicit and processed; also localization derived issues must be handled both from legal and temporal points of view. Legal tagged corpora and related resources (such as sets of rules and expressions) are also scarce.

2.2 Representation of temporal expressions

Regarding formal representations, a lot of different representing options exist, especially oriented to concrete fields and applications. There are for instance formal temporal languages (such as Tokio [9] or TProlog [14]) and representations, mainly thought for planning or verification purposes [20]; we can find among them TimeGraphs [10], Kripke structures [15] or Temporal Constraint Networks [7]. If we limit the scope to general purpose representations in the computer science field, the most used ones are ontologies, finding an extensive bunch of them, such as the W3C OWL-Time Ontology or the Reusable Time ontology [30]. Other proposals suggest adding a temporal layer to concrete ontological representations [11], by using for instance Named Graphs [5], or using temporal querying specific languages (such as τ-SPARQL [24]). Also new ideas for representations arose lately, such as the four-dimensionalist approach [28] or the representation used by Yago in its second version [13].

Finally, for the legal application, OASIS-LegalRuleML [2] and ODRL[4] can be used for document representation, at least partially. For constraints and validation, SHACL[5] and ODRL can be highlighted among other languages available, such as ShEx[6]; unfortunately, none of them seems to be able to handle time with enough complexity. Therefore either a temporal expansion or some kind of connection with temporal resources should be needed.

2.3 Temporal Reasoning

There exist several classical theories on Temporal Logic (TL) [21,26], such as Shoham's [12] or McDermott's [17]. They are usually chosen for different applications depending on concrete ontological aspects such as the basic ontological unit to model time (points or intervals) or if time would be relative or absolute [3]; this choice can differ even when applying to the same domain depending on concrete uses [29]. Even when there exist subsequent proposals to the classical ones [25], these remain as the reference in the field, stressing amongst these Allen's Interval TL [1], in whose 13 basic temporal binary relations are based the most of the related resources.

Besides the ontological distinctions in TL theories, one of the main problems when reasoning on time is vagueness: sometimes information is not expressed clearly as concrete dates but as imprecise terms uneasy to represent formally. In this context, fuzzy logic ideas have already been applied to deal with vague

[4] `https://www.w3.org/community/odrl/xml/2.1/`
[5] `https://www.w3.org/TR/shacl/`
[6] `https://www.w3.org/2001/sw/wiki/ShEx`

temporal statements such as *"late thirties"* [19]. Also the existence of multiple possibilities can be expressed by using Fuzzy Branching TL, where different temporal lines allow reasoning about different possible worlds [18].

Temporal reasoning is often related to representation; such is the case of TProlog and other logic languages. When the chosen representation is an ontology, logic is usually carried out by Description Logics (DL); when completing a DL with temporal operators and TL, we obtain a Temporal DL [16] that allow us to represent and reason on time in ontological environments. Among other options for logic in ontologies we find N3Logic [4], a RDF expansion proposed lately. Despite all these recent options, none has prevailed over the rest as a standard, nor has been used in the Legal domain; in fact, new efforts in legal reasoning, such as previously mentioned MIREL project, recognized this lack and focus in the need of adding a temporal dimension as one of the main research challenges in the field.

3 Problem Statement and Contributions

As exposed in the previous section, there is a gap between the theoretical logical formulation and the practical application and representation. The issues found when trying to formalize temporal information in the legal field suggests that current languages can't support temporal reasoning. The main research question to answer to bridge this gap by building a framework able to deal with temporal constraints (Fig. 1), with the sub-questions it leads to, is, therefore:

Are current languages rich enough for formally representing and reasoning on usually key temporal restrictions in legal documents?

- *Are they tuned with the capabilities of state of the art NLP extraction?* NLP is a domain in constant evolution, and recent advancements in its techniques might not be properly reflected and exploited in formal representation and reasoning systems.
- *Are nowadays representation systems able to properly reason for legal purposes? Can they represent temporal constraints? Would a brand new representation be needed or could some extension of an existing one be done?* We will at first focus on chosen languages (OASIS-LegalRuleML, SHACL and ODRL) for temporal expansion, as it is the framework already proposed, but maybe other representations would be helpful or could even replace them.
- *Which time-related queries can reasoners answer? What are the main lacks?* We need to find the current temporal reasoners limitations, identifying the questions that reasoners can deal with on the different representations available and a list of the most important lacking ones that should also be solved.
- *Are there enough semantic resources for legal NLP to use existing tools?* Even when there are already NLP techniques, schemes and tools for all the tasks to do, it will be necessary to study them concretely for the legal domain, that has very specific structures for both events and temporal expressions.

We plan to study them and make a proper resource by systematizing common relations and expressions present in legal texts, starting from judgments in Eur-Lex.

Main research hypothesis to validate are that an extension over constraint languages and that the integration of NLP, representation and reasoning would lead to a more powerful and flexible treatment of time. This would allow to answer queries depending on time and context, such as *"Can I access now to a concrete series or football match on my cable provider?"*, where each series has policies attached related to location, age, release dates and kind of customer, linked also to temporal constraints (often foreign TV series can't be watched in Spain before a concrete period after original broadcasting). Also application to grants or calculations on the amount of fines depending on relative times of infraction and payments could benefit of research on the legal domain.

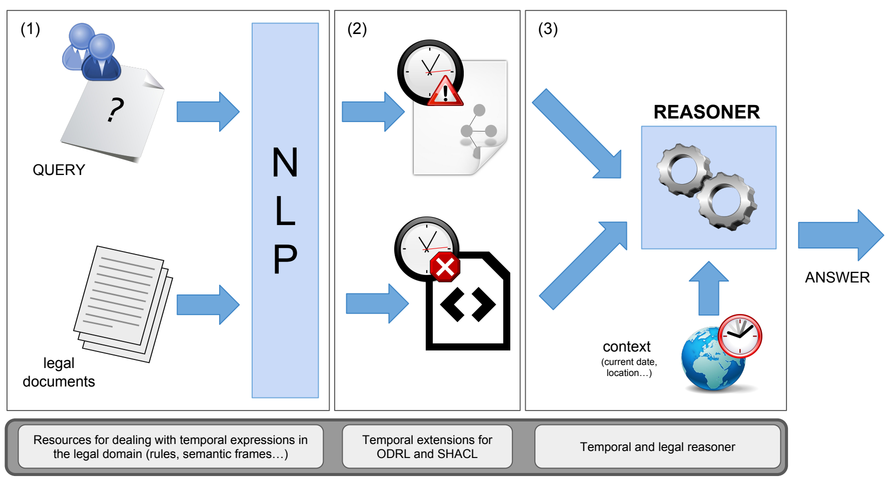

Fig. 1. Legal Framework: At top the parts of the framework (1) NLP, (2) Representation and (3) Reasoning; below, the contributions expected on each part. Images adapted from Wikimedia Commons.

4 Research Methodology and Approach

Research methodology will be incremental regarding the different steps outlined in Fig. 1 and tackling the Research Questions stated in Section 3:

Requirements and lacks In order to build the framework, legal texts will be studied to see what information must be extracted and how to do it; collaboration of both potential users and experts in the field will be required to detect main problematic aspects and the optimal way to handle them. Judgments, one of the type of documents publicly available in Eur-Lex, will be the starting point. Once information is identified, it must be tagged and represented in the already existent languages, studying and solving their main lacks to reach what required by users. Finally, and as exposed in Section 2, temporal reasoning can be tackled in several ways and depends on representation, being available options Temporal Description Logics or specific logic languages such as TProlog. Several proposals will be examined, choosing the one that better fits the representation.

NLP In (Fig. 1 (1)) the framework receives a query from a user, about one or several legal documents. All the input is processed by several NLP tools (annotators, rule systems...). An extensive study has been performed to see the existing tools and choose the adequate ones for temporal purposes, resulting on formerly mentioned HeidelTime as a starting point for its flexibility on domain-dependent rule creation and multilingualism. The contribution in this first step will be concrete resources to feed this legal temporal text processing, built brand new (such as a set of legal patterns and tagged corpora from Eur-Lex) or based on existing semantic resources, such as FrameNet-style set of legal frames representing usual legal situations.

Representation Processed information should be represented in chosen languages (Fig. 1 (2)). Here we will use the previous study about how others representations handle temporal information and later considerations about reasoning to extend our representation and make it able to support temporal information and constraints derived from legal documents and the query. It will also be populated by information obtained in the NLP step (Fig. 1 (1)).

Reasoning The final knowledge representation will be the input, along with contextual information (such as the current date) to the legal reasoner able to deal with time (Fig. 1 (3)); this will be the result of the study of several systems and the application of the best method for legal purposes and for the final representation. It might be an adaptation of a temporal logic to the legal context or a brand new alternative, and its output will be the final answer to the user's query.

5 Results and Evaluation Plan

Since the exposed Ph.D. is an early stage, started in October 2016, there are no results available yet except the problem identification, the partial state of the art already examined and the pinpointing of the different disciplines that will be part of the thesis working plan.

Related to evaluation, as exposed in Fig. 1, steps involved in the framework are sequential, relying on the previous one; each step will have its own evaluation, based on statistical performance measures for NLP and reasoning given answers (such as precision or recall) and quality and completeness for representation. We have already identified several resources available for all the intervening fields that will be useful for training, such as judgments in Eur-Lex portal, COLIEE[7] and TREC [8][9] corpora, resources that use TimeML[10] or Temporal Tracks in SemEval[11]. Finally, the framework will process, represent and reason on different corpora to see if the representation schema can express information properly and if the developed reasoning system can answer correctly queries submitted by end users; if so, the framework would be considered successful.

6 Conclusions

This paper presented the outline of a Ph.D. proposal of adding a temporal dimension to representation and reasoning in the legal field, addressing the problem in an integrated manner. This approach intends to improve legal information processing and perform machine reasoning, making it accessible through a complete framework that takes legal documents and answers queries about them.

References

1. Allen, J.F., Ferguson, G.: Events and actions in interval temporal logic. Journal of Logic and Computation 4(5), 531–579 (1994)
2. Athan, T., Boley, H., Governatori, G., et al.: OASIS LegalRuleML. In: Proceedings of the 14th International Conference on Artificial Intelligence and Law. p. 3 (2013)
3. van Benthem, J.: The logic of time: a model-theoretic investigation into the varieties of temporal ontology and temporal discourse. Kluwe Academic Publishers (1983)
4. Berners-Lee, T., Connolly, D.: N3logic: A logical framework for the world wide web. Theory and Practice of Logic Programming 8(3), 249–269 (2008)
5. Carroll, J.J., Bizer, C., Hayes, P., Stickler, P.: Named Graphs , Provenance and Trust. Digital Media (2004)
6. Chambers, N., Wang, S., Jurafsky, D.: Classifying temporal relations between events. In: Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions. pp. 173–176. No. 2005 (2007)
7. Dechter, R., Meiri, I., Pearl, J.: Temporal constraint networks. Artificial Intelligence 49(1-3), 61–95 (1991)
8. Derczynski, L.R.: Determining the Types of Temporal Relations in Discourse. Ph.D. thesis, University of Sheffield (2013)
9. Fujita, M., Kono, S., Tanaka, H., Moto-Oka, T.: Tokio: Logic Programming Language Based on Temporal Logic and Its Compilation to Prolog. In: Proceedings on 3rd International Conference on Logic Programming. pp. 695–709 (1986)
10. Gerevini, A., Schubert, L.: Efficient temporal reasoning through timegraphs. In: Proceedings of the 13th International Joint Conference on Artifical Intelligence. pp. 648–654 (1993)

[7] http://webdocs.cs.ualberta.ca/~miyoung2/COLIEE2017/
[8] http://www.trec-ts.org/downloads
[9] https://trec-legal.umiacs.umd.edu/
[10] http://www.timeml.org/timebank/timebank.html
[11] http://alt.qcri.org/semeval2016/task12/

11. Gutierrez, C., Hurtado, C.A., Vaisman, A.: Introducing time into RDF. IEEE Transactions on Knowledge and Data Engineering 19(2), 207–218 (2007)
12. Halpern, J.Y., Shoham, Y.: A propositional modal logic of time intervals. Journal of the ACM 38(4), 935–962 (1991)
13. Hoffart, J., Suchanek, F.M., Berberich, K., Weikum, G.: YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. Artificial Intelligence 194, 28–61 (2013)
14. Hrycej, T.: A Temporal extension of Prolog. The Journal of Logic Programming 15(1-2), 113–145 (1993)
15. Kripke, S.: Semantical Considerations on Modal Logic. In: Acta Philosophica Fennica. vol. 16, pp. 83–94 (1963)
16. Lutz, C., Wolter, F., Zakharyashev, M.: Temporal description logics: A survey. In: Proceedings of the International Workshop on Temporal Representation and Reasoning. pp. 3–14 (2008)
17. McDermott, D.: A temporal logic for reasoning about processes and plans. Cognitive Science 6(2), 101–155 (1982)
18. Moon, S.I., Lee, K.H., Lee, D.: Fuzzy Branching Temporal Logic. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 34(2), 1045–1055 (2004)
19. Nagypál, G., Motik, B.: A Fuzzy Model for Representing Uncertain, Subjective, and Vague Temporal Knowledge in Ontologies. In: On the Move to Meaningful Internet Systems. pp. 906–923 (2003)
20. Orgun, M., Ma, W.: An Overview of Temporal and Modal Logic Programming. In: Proceedings of the 1^{st} International Conference on Temporal Logic - Lecture Notes in Artificial Intelligence. pp. 445–479 (1994)
21. Pani, A.K., Bhattacharjee, G.P.: Temporal representation and reasoning in artificial intelligence: A review. Mathematical and Computer Modelling 34(1-2), 55–80 (2001)
22. Pustejovsky, J., Lee, K., Bunt, H., Romary, L.: ISO-TIMEML: An International Standard For Semantic Annotation. In: Proceedings of the 7^{th} conference on International Language Resources and Evaluation. pp. 394–397 (2010)
23. Strötgen, J., Gertz, M.: Multilingual and cross-domain temporal tagging. Language Resources and Evaluation 47(2), 269–298 (2013)
24. Tappolet, J., Bernstein, A.: Applied Temporal RDF: Efficient Temporal Querying of RDF Data with SPARQL. In: ESWC. pp. 308–322 (2009)
25. Vila, L., Schwalb, E.: A Theory of Time and Temporal Incidence based on Instants and Periods. In: Proceedings of International Workshop on Temporal Representation and Reasoning. pp. 21–28 (1996)
26. Vila, L.: A Survey on temporal reasoning in artificial intelligence. AI Communications 7(1), 4–28 (1994)
27. Vila, L., Yoshino, H.: Time in Automated Legal Reasoning. Information and Communications Technology Law 7(3), 1–30 (1998)
28. Welty, C., Fikes, R.: A Reusable Ontology for Fluents in OWL. In: Proceedings of 4^{th} FOIS. pp. 226–236 (2006)
29. Zhou, L., Hripcsak, G.: Temporal reasoning with medical data-A review with emphasis on medical natural language processing. Journal of Biomedical Informatics 40(2), 183–202 (2007)
30. Zhou, Q., Fikes, R.: A reusable time ontology. In: Proceeding of the AAAI Workshop on Ontologies for the Semantic Web (2002)