

Article

Semantic Annotation of Legal Contracts with ContrattoA

Michele Soavi ^{1,*} , Nicola Zeni ¹ , John Mylopoulos ² and Luisa Mich ¹ ¹ Department of Industrial Engineering, University of Trento, Via Sommarive 14, 38123 Trento, TN, Italy² School of Electrical Engineering and Computer Science, University of Ottawa, 800 King Edward Ave, Ottawa, ON K1N 6N5, Canada

* Correspondence: michele.soavi@unitn.it

Abstract: The aim of the research is to semi-automate the process of generating formal specifications from legal contracts in natural language text form. Towards this end, the paper presents a tool, named ContrattoA, that semi-automatically conducts semantic annotation of legal contract text using an ontology for legal contracts. ContrattoA was developed through two iterations where lexical patterns were defined for legal concepts and their effectiveness was evaluated with experiments. The first iteration was based on a handful of sample contracts and resulted in defining lexical patterns for recognizing concepts in the ontology; these were evaluated with an empirical study where one group of subjects was asked to annotate legal text manually, while a second group edited the annotations generated by ContrattoA. The second iteration focused on the lexical patterns for the core contract concepts of obligation and power where results of the first iteration were mixed. On the basis of an extended set of sample contracts, new lexical patterns were derived and those were shown to substantially improve the performance of ContrattoA, nearing in quality the performance of experts. The experiments suggest that good quality annotations can be generated for a broad range of contracts with minor refinements to the lexical patterns.



Citation: Soavi, M.; Zeni, N.; Mylopoulos, J.; Mich, L. Semantic Annotation of Legal Contracts with ContrattoA. *Informatics* **2022**, *9*, 72. <https://doi.org/10.3390/informatics9040072>

Academic Editors: Sanjay Misra, Robertas Damaševičius and Bharti Suri

Received: 9 August 2022

Accepted: 15 September 2022

Published: 20 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: legal contract; semantic annotation; structural annotation; contract ontology; semantic annotation tool

1. Introduction

Legal contracts constitute for millennia the main vehicle for conducting business transactions worldwide. They are established (aka ‘formed’ in Law) through a systematic negotiation process, followed by an execution (aka ‘performance’) supported by legal dispute resolution mechanisms. Contracts exist as natural language (NL) text using legal terminology grounded on legal concepts, such as those of obligation and power.

The aim of the research is to transform legal contract text into formal specifications for two reasons. Firstly, there is much interest in Law in the algorithmic analysis of legal contracts to ensure they are consistent with the expectations and interests of contracting parties. Formal analysis tools, such as model checkers [1] and SMT/OMT solvers [2], have come of age in the past decade and are used routinely to analyze various kinds of artifacts, including hardware, software and business process designs. However, such tools can only be used with a formal specification of the artifact to be analyzed. Secondly, there is a new class of software systems called smart contracts [3] that partially automate, monitor and control the execution of legal contracts. Formal specifications of legal contracts can serve as a starting point for the systematic tool-supported process of generating smart contract code. The formality of specifications is essential to avoid ambiguity, a ubiquitous trait of natural language documents and a critical issue for legal contracts [4].

Based on experiences from our earlier work [5,6] we envision the generation of a formal specification from NL text as a five-step process: (a) identify domain terms in the text; (b) annotate the text using an ontology for legal contracts, to determine text fragments that describe concepts such as ‘role’, ‘obligation’, ‘power’ and ‘asset’; (c) mine relationships

among the concepts identified in (b); (d) generate a domain model for domain terms identified in (a) and define parameters and local variables to serve as atomic building blocks for expressions; and (e) generate formal expressions for the constituents of obligations and powers from legal text fragments using parameters and local variables defined in step (d). Considering that formalization of legal text is a laborious and error-prone process, we envision our task as one of semi-automating the generation of formal specifications from natural language with tools that improve the quality of generated specifications while substantially reducing manual effort.

The main goal of this paper is to report results on step (b). Towards this end, we adopt GaiusT [5], a semantic annotation platform for legal documents, to build ContrattoA, a semantic annotation tool for contracts. The implementation requires: (1) a semantic annotation ontology for contracts that indicates what the concepts are that we are trying to identify in text, (2) a structural model of contracts—e.g., definitions, clauses, etc.—expressed as an extended BNF (eBNF) grammar that defines the structure of legal contracts, and (3) lexical patterns for recognizing elements of the ontology in text. We then experiment with the tool to determine how well it performs annotation. Experimentation was conducted in two iterations where lexical patterns were defined for legal concepts, given a set of sample contracts, and their effectiveness was evaluated. The first iteration resulted in version one of the tool, ContrattoA v1.0, and involved a handful of sample contracts. The resulting patterns were evaluated with an empirical study where one group of subjects was asked to annotate legal text manually, while a second edited the annotations generated by ContrattoA v1.0. The results of the empirical study suggest that ContrattoA v1.0 substantially improves the quality of annotations relative to manual annotation. The second iteration—which resulted in ContrattoA v2.0—focused on lexical patterns for the concepts of ‘obligation’ and ‘power’, where results from the first iteration were mixed. On the basis of an extended set of sample contracts, new patterns were proposed and those were experimentally evaluated and shown to substantially improve the performance of ContrattoA v2.0 over its predecessor, approaching in quality the performance of experts.

An early version of this work appeared in a conference publication [7] that presented ContrattoA (previously named ContracT) along with its performance evaluation. In this paper we added the second iteration results and established experimentally an improvement that approaches the performance of expert annotators. In addition, we extended and improved related work through a systematic literature review that covers the annotation of legal contracts, resulting in 44% more content over its conference precursor. The publication of [7] offers evidence for the originality of the contributions of this work, as well as the literature review presented in Section 6.

The rest of the paper is organized as follows. Section 2 describes the research baseline for this work, including legal contracts, an ontology for legal contracts and GaiusT, the semantic annotation platform we have adopted. Section 3 describes the prototype version of ContrattoA v1.0, including lexical and formatting patterns for identifying instances of the concepts of the ontology, as well as elements of contract structure. Section 4 presents the evaluation results of an empirical study that uses ContrattoA to annotate two contracts, while Section 5 presents improved lexical patterns for obligations and powers as well as the results of an empirical evaluation of these patterns. Section 6 discusses related work and Section 7 concludes and presents planned future work.

2. Research Baseline

2.1. Legal Contracts

A *legal contract* (or simply *contract*) is a collection of obligations and powers between contracting parties. As legal artifacts, contracts come about through a negotiation process that includes an *offer* of a *consideration* (something of value, an asset) in exchange for another asset, such as money. The negotiation terminates with *acceptance* and *formation* (signing). The execution (*performance*) of a contract may be suspended, subcontracted, successfully or unsuccessfully terminated, renegotiated, or renewed.

Contracts can be understood as *legal processes*, prescriptions of allowable contract executions [4,8]. Compared to their business process cousins, contracts are outcome-oriented, focusing on ‘what’ the obligations and powers of different stakeholders are, leaving the ‘how’ to the parties responsible. In addition, contracts fundamentally differ from business processes in that they can change during their execution through the exertion of powers. For a pizza delivery example, if delivery is not completed within 30 minutes, the customer has the power to terminate the contract, or pay half of the agreed upon price.

2.2. An Ontology for Contracts

An ontology consists of a collection of concepts and relationships for conceptualizing a domain, in our case contracts. The ontology we adopt has been proposed in [9] for the Symboleo contract specification language and is shown in Figure 1. The ontology is based on the Core Legal Ontology UFO-L [10] and was developed in consultation with legal experts incorporating Hohfeld’s theory of legal positions [11] but without some of its shortcomings to address contract specific elements. In addition, it has been tested with dozens of contracts to confirm with legal experts its ability to capture legal discourse and its implementation is available to the public.

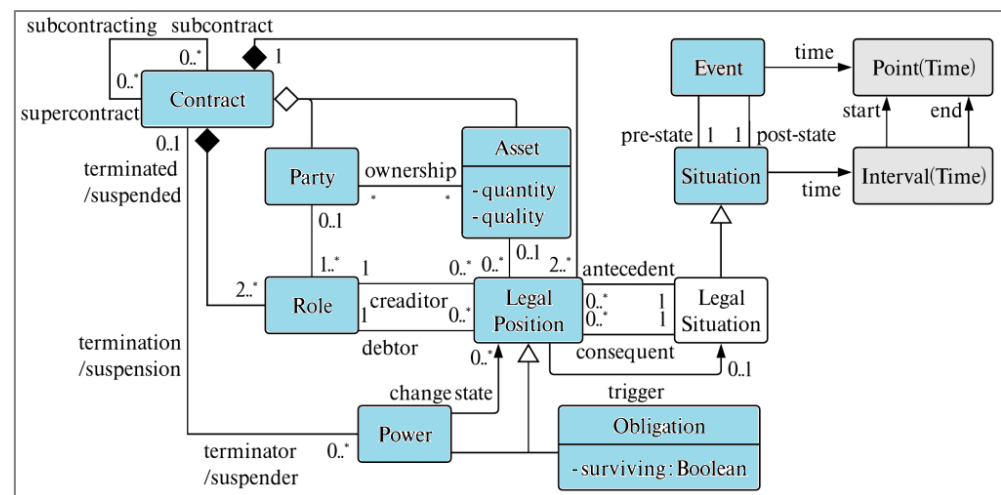


Figure 1. Contract ontology adopted from Symboleo [9] using standard UML notation for class diagrams.

The concepts included in the ontology are as follows:

Contract: a collection of *obligations* and *powers* between two or more roles, which are assigned to parties during contract execution, and are concerned with two or more *assets* respectively associated with each role. Contracts may involve subcontracting that assigns to third parties the responsibility for the fulfilment of obligations.

Asset: an owned (tangible or intangible) item of value that serves as contractual consideration [12]. Assets constitute the benefits contracting parties get out of a contract. Asset quantity and quality constraints are typically specified in contracts or are defined as contract parameters that vary from execution to execution.

Legal Position: a legal relationship between two roles. We consider only two such relationships: *obligation* and *power* [11], since these can account for all eight proposed by Hohfeld, according to our study of legal contracts.

Obligation: the legal duty of a *debtor* towards a *creditor* to bring about a certain legal situation (*consequent*) when another legal situation (*antecedent*) holds. *Surviving* obligations remain in effect after the termination of the contract, such as a 6-month nondisclosure obligation after the end of a contract. Obligations usually concern assets and are instantiated by *triggers*. When debtor A has an obligation towards creditor B, then B has a *right* towards A.

Power: the right of a party, the creditor, to create, change, suspend, or cancel legal positions. A power is instantiated by a trigger and has an antecedent (legal situation) that must be met for it to come into effect, as well as a consequent that the creditor can make true.

Legal Situation: a type of situation associated with a contract, obligation, or power instance. Situations are states-of-affairs and are comprised of entities and relationships. A situation *occurs* during a time interval T and *holds* during any subinterval of T.

Event: a happening that occurs at a time point and cannot change. Events have pre-state and post-state situations. For example, delivered is an event whose pre-state is the situation ‘being in transit’ and post-state is the situation ‘being at its destination point’.

Role: a placeholder for a party participating in obligations and powers that assign to it responsibilities and rights.

Party: a legal agent (person or institution) who owns assets and who is assigned roles in contracts.

2.3. The GaiusT Semantic Annotation Platform

GaiusT is a web-based platform intended to build annotation tools in the legal domain [5]. GaiusT uses patterns defined by eBNF grammars to annotate legal text with structural and semantic tags. Structural analysis identifies legal document structure—such as title, chapters, sections and clauses—as well as cross-references to other sections of the same document (i.e., internal references) or external ones, such as applicable laws. The purpose of semantic annotation is to identify the footprint of the main constituents of a legal document in text, such as rights and obligations. Using an ontology of legal concepts, GaiusT supports the identification of lexical patterns that guide the annotation process.

The core components of GaiusT are the annotation schema generator and the annotation engine. The annotation schema generator takes as input an annotation ontology expressed in XML Metadata Interchange (XMI), RDF or OWL format that describes the concepts to be annotated in text and supports the generation of an annotation schema that describes allowable nesting for annotations. In addition, the generator takes as input lexical indicators that describe instances of the annotation ontology and supports the generation of lexical patterns used by the annotation engine. The generator also supports interfaces to lexical resources such as WordNet, Thesaurus and Wikipedia.

The annotation engine takes as inputs an annotation schema, as well as lexical patterns and a structural grammar to annotate text using the concepts in the schema, (a) extracts plain text from files in a variety of formats (including Microsoft Word, rdf, pdf, HTML); (b) normalizes the plain text by removing unprintable characters and produces a text document where each line represents a phrase; (c) annotates text fragments with tags for structure and cross-references; and (d) annotates text fragments with semantic tags present in the annotation schema. Its output is semantically and structurally annotated text in XML format.

3. ContrattoA 1.0

The development of ContrattoA 1.0 required building a tool on the GaiusT platform to perform semantic annotation of legal contracts. As indicated in Section 2.3, the following three inputs are required by GaiusT to build an annotation tool: (a) a semantic annotation ontology, as introduced in Section 2.2; (b) a structural grammar for input texts; and (c) lexical patterns for each concept on the semantic annotation ontology. Since our annotation ontology was adopted from Symboleo, in this section we discuss the last two items, as well as the process by which they were obtained.

3.1. A Structural Model for Contracts

Contracts often come with syntactic structure that identifies title, contracting parties, contracting issues (delivery, payment, termination), clauses that define obligations and

powers. The identification of such structure can facilitate the semantic annotation process, but also later stages of the transformation from NL to a formal specification.

The syntactic structure of contracts is domain- and jurisdiction-dependent. For instance, rental contracts are differently structured than retail ones, and rental contracts are differently structured in Italy as compared to Canada.

Our structural model was constructed by reviewing a small number of sample contracts from the Vaccine Procurement domain in the USA. A fragment of a structured contract (<https://www.hhs.gov/sites/default/files/pfizer-inc-COVID-19-vaccine-contract.pdf> (accessed on 7 August 2022)) is shown in Figure 2. The identification of structural elements—such as title, preamble, issue and clause—has been performed manually.

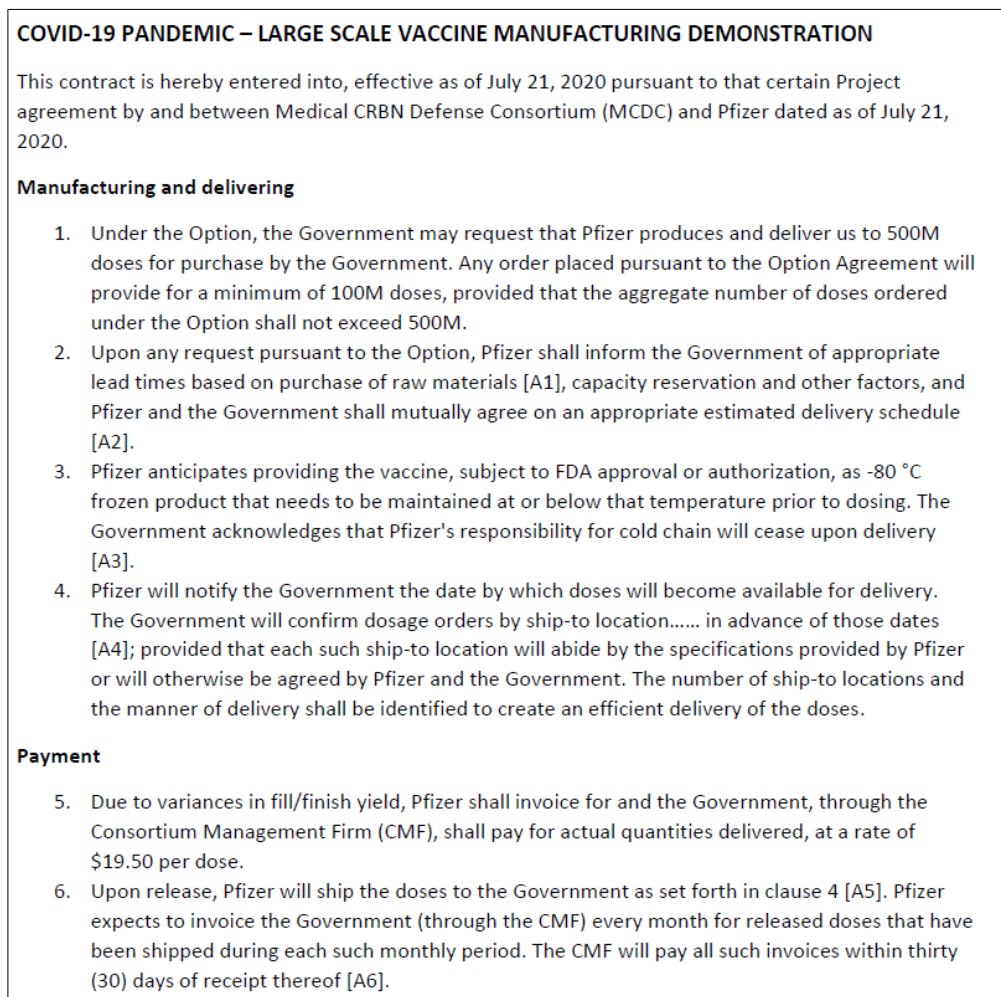


Figure 2. Excerpt from a sample contract.

Afterwards, recurring patterns were identified among sample contracts, resulting in an eBNF grammar, as shown in Figure 3.

It should be added that manually determining a structural model for a set of contracts is a simple-to-conduct process that can be repeated for different collections of contracts coming from new jurisdictions.


```

<Contract> ::= <Ident> <Title> <CLRF>
| [(Preamble) <CLRF>]
| [(Definitions)]
| <Issues>
| <Signatures>

<Ident> ::= ['Contract n.'] ((Digit) |(Letter))+ [(CLRF)]
<Title> ::= <String>
<Preamble> ::= ['Preamble'] <String>
<Definitions> ::= ['Definitions'] [(CLRF)] ((Definition) <CLRF>)+
<Definition> ::= <Name> ('means'|'refers to') <Text>
<Issues> ::= ((Issue) <CLRF>)+
<Issue> ::= <Name> <CLRF> ((Clause)<CLRF>)+
<Signatures> ::= ((PartyName) <Date>)+ /* at least two parties */
<CLRF> refers to carriage return or new line symbols
<String> includes any sequence of letters, numbers and special symbols
<Name> includes a word, sequence of letters and possibly periods ('.') with no
blanks
<PartyName> includes one, two or three words
<Date> includes dates in several different formats

```

Figure 3. eBNF grammar for contract structure.

3.2. Lexical Patterns for Contract Concepts

Lexical patterns define the combinations of words that identify instances of a concept in the text. To define such patterns, we use an eBNF with a further extension, the & (and) operator. If pattern <Simple> is defined as 'a' & 'b', then a text fragment x will be tagged as Simple, in XML <Simple> x </Simple> if it contains 'a' and 'b' in any order. Note that this is different from the lexical pattern <Simpler> defined as 'a' 'b' that will tag a piece of text as Simpler if it contains the string 'ab'. Likewise, the more elaborate pattern

```

<Obligation> ::= (<Party> | <Role>) & [<Party> | <Role>] & (<NecessityM> <Action> <Asset>)
& <Situation> & [<Situation>] & [<Situation>]

```

will tag a text fragment as an obligation if it includes a text fragment tagged Party or Role, and an optional second party or role (sometimes the obligations described in a contract only mention the debtor), and a fragment tagged NecessityM, Action and Asset in sequence and also includes one necessary situation (the consequent of the obligation), and two optional ones, for the trigger and antecedent respectively.

Lexical patterns were identified with support from GaiusT that conducts statistical analysis on sample contracts using frequency measures and n-grams to determine a Word Frequency List (WFL). Subsequently, the WFL is enriched with input obtained from lexical databases such as WordNet, Thesaurus, Google Ngrams and Wikipedia. The output is integrated with manual refinement to make sure that the most recurring patterns are included. The resulting patterns for tags corresponding to the elements of the annotation ontology are presented in Figure 4.

```

<Asset> ::= ( <Resource> | <Amount> )
<Party> ::= 'Mr' | 'Ms' | 'Mister' | 'Misses' | 'Inc' | 'Ltd'
          | 'Spa' | 'Srl' ... <Actor> <String>
<Role> ::= 'supplier' | 'buyer' | 'landlord' | 'lessee' | 'seller'
          | 'lessor' | 'manufacturer' | 'carrier' ... /* domain specific roles */
<CurrencyUnit> ::= '$' | '£' | '€' | 'USD' | 'GBP' | 'EUR' ...
<Unit> ::= <CurrencyUnit> | <WeightUnit> | ...
<Connector> ::= 'not' | 'any' | 'either' | 'all' <expr>
<Condition> ::= 'if' | 'then' | 'else' | 'or' <expr>
<Possibility> ::= 'can' | 'could' | 'may' | 'permit' | 'right' ...
<Necessity> ::= 'have to' | 'must' | 'shall' | <NegConnector>
              | <PositiveModal>
<Amount> ::= <Unit> <Number> | <Number> <Unit>
<TemporalExpression> ::= ((0[1-9])|(1[0-2]))((0[1-9])|(1<Digit>)|(2<Digit>))
                        | ((3[0-1]))(<Digit>4)
<StartDate> ::= <TemporalItem>
<TypeOfDate> ::= <StartDate> | <EndDate> ...
<Obligation> ::= ( <Party> | <Role> ){1,2} <Action>{0,1} <NegativeModal>
                | <Action>{1,2} <Asset>{1,2} <TemporalExpression> {0,1}.
<Power> ::= ( <Party> | <Role> ){1,2} <Action>{0,1} <PositiveModal>
            | <Action>{1,2} <Asset>{1,2} <TemporalExpression> {0,1}.
<Exception> ::= 'only' | 'but' | 'except' | 'without' | 'limit'
               | 'restrict' | 'exclusion' | 'other than' | 'unless' ... .

```

Figure 4. eBNF grammar for lexical patterns.

3.3. The Annotation Process

The contract ontology, along with the structural and semantic grammars represent the inputs required to perform semantic annotation with ContrattoA. The two grammars are manually imported into ContrattoA.

Contracts to be annotated are imported in 'Files' and preprocessed to generate a WFL. Subsequently, the annotation process can be run from the 'annotation' tab (Figure 5).

The result of the process is annotated text that marks text fragments of the contract description using tags derived from the semantic annotation ontology, as in the following examples.

<Obligation> The <Role> Company </Role> hereby <Situation> employs the <Role> Employee </Role> as its CFO </Situation> </Obligation>

<Power> <Role> Company </Role> may <Situation> terminate employment at any time for cause. </Situation> </Power>

In addition, ContrattoA provides the list of all annotated chunks of text classified under each concept of the ontology. Finally, the annotated contracts with markup can be downloaded in XML or HTML format.

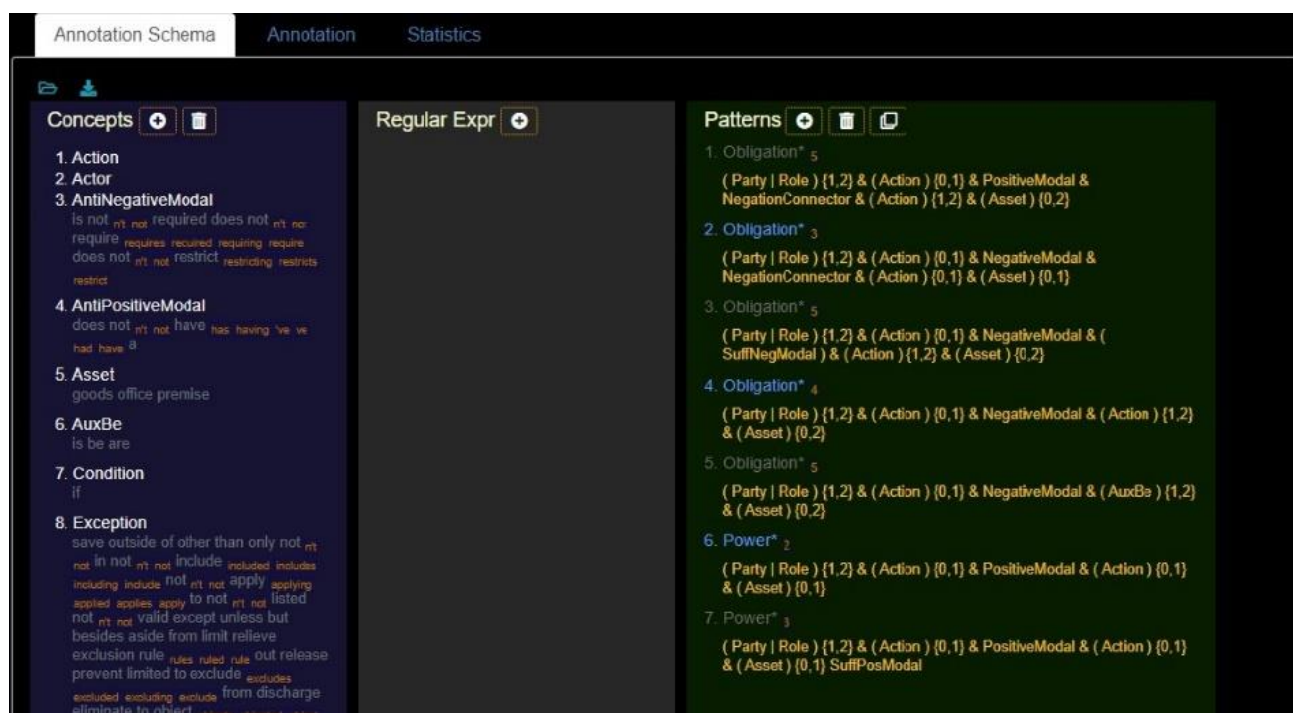


Figure 5. Interface for concepts and patterns input for semantic annotation in ContrattoA.

4. Experimental Evaluation of ContrattoA 1.0

The evaluation covers its performance in terms of annotation precision and recall, as well as an empirical evaluation that assesses how much ContrattoA helps human annotators carry out the annotation task. Precision and recall constitute the standard metrics for evaluating the quality of semantic annotation [13].

4.1. The Experimental Setup

The experiment was conducted with two real-life business contracts involving a Freight and a Rental agreement. The contracts were selected to have an approximately equal number of clauses. Six persons attending a course on Public Law at the University of Trento, all novice annotators, agreed to participate in the experiment. The participants were divided into two groups, and one was asked to individually annotate the contracts, while the other edited the output of ContrattoA. To mitigate the impact of learning effect, in the first round, one group received the original text of the contract whereas the other group received the same text augmented with annotations generated by ContrattoA. In the second round, the tasks were reversed using a different contract. The participants were requested to annotate the following concepts: Party, Role, Asset, Obligation, Power, Situation.

Prior to the experiment, the participants were provided with the definition of annotation concepts. To support insertion and modification of semantic tags in the input documents, the participants were provided with a user-friendly web-based tool.

4.2. Experimental Results

The results of the experiment have been compared to the manual annotation performed by an expert, a PhD student in Law, which served as gold standard. This resulted in a total of 44 legal concepts annotated in the Rental agreement and 39 in the Freight agreement. The involvement of one expert annotator has been deemed sufficient given the simplicity and limited length of the contracts. In addition, the expert's annotation has been checked by the authors who are experts in annotation and have some background in Law. The subjects' annotations were then compared to the expert annotation for recall and precision in performing ContrattoA-supported and manual annotation (Tables 1 and 2) as defined

in [13] (details on dealing with annotation agreement are given in [5]). The statistical results of the study consist of recall and precision values for the two subject groups relative to the gold standard. In the experiment, precision measures the fraction of concept instances correctly annotated out of the total number of concept instances identified by a subject. Recall, on the other hand, measures the fraction of concept instances correctly annotated by a participant out of the total number of concept instances in the gold standard. One of the problems in evaluating annotation results concerns the different granularity, i.e., length of marked up text, for different concepts. For simple concepts, such as role, party or asset, few-word patterns are sufficient. For deontic concepts, such as obligation or power, lexical patterns used to annotate them can be quite complex depending on how domain-specific these patterns are. An annotation of a concept instance was considered correct when the annotated text included all word sequences in the concept's lexical pattern.

Table 1. Comparison of manual and assisted annotation of Rental agreement and Freight agreement contracts.

Measure	Rental Agreement		Freight Agreement	
	Manual	Assisted	Manual	Assisted
Recall	0.47	0.75	0.31	0.40
Precision	0.57	0.73	0.34	0.41
Fallout	0.40	0.53	0.46	0.49
Accuracy	0.51	0.71	0.40	0.47
Error	0.49	0.29	0.60	0.53
F-Measure	0.52	0.74	0.33	0.41

Table 2. Average Recall and Precision for concept annotation of both contracts.

Measure	Temporal Condition		Power		Obligation		Situation	
	Manual	Assisted	Manual	Assisted	Manual	Assisted	Manual	Assisted
Precision	0.79	1.00	0.75	0.63	0.76	0.90	0.41	0.78
Recall	0.52	0.69	0.45	0.52	0.48	0.80	0.17	0.36
F-Measure	0.66	0.84	0.60	0.57	0.62	0.85	0.29	0.57

The results of the experiment did not identify a reduction in annotation time when supported by ContrattoA, as the average annotation time was 23:41 min for ContrattoA-assisted annotation compared to 22:16 min for manual annotation. ContrattoA support was found to be more effective for annotation quality for both contracts and most of the concepts. Except for precision in the annotation of power, the support of ContrattoA was found to improve performance in most cases, while being comparable for the rest. Generally, precision results were higher than recall ones for all concepts. Moreover, the results suggest a high degree of variability for different concepts. That is not surprising, as some concepts, such as power, are more difficult to grasp than others, and therefore more difficult to identify in text.

Temporal conditions and obligations are the concepts for which the annotation tool performed best. Conversely, powers and situations were identified to have the highest level of difficulty and variability among subjects with the two concepts frequently interchanged or overlapping. In addition, seldom recurring elements such as internal and external references, were less frequently identified both in assisted and manual annotation. By comparison, ContrattoA, performed better than manual annotation, but also exhibited similar difficulties and strengths. Annotation results concerning assets and parties have

been excluded from the statistics in Tables 1 and 2 as they are all composed of a few alternative words and were annotated correctly by all participants as well as ContrattoA. For such concepts, the tool helps in reducing annotation time and manual effort. In addition, it was useful in generating tags for all concepts in the annotation ontology, rather than a subset. Moreover, the lower performance in manual and assisted annotation for the Freight agreement compared to the Rental agreement suggests that ContrattoA could be more useful for contracts with a higher degree of complexity, a significant number of parties involved, and lower familiarity of the annotators for the domain, noting that the Freight agreement refers to a contract mostly used in a B2B context. We consider recall to be more important than precision for practitioners in the annotation of legal contracts as it speeds up the annotation process by reducing false positives. For future development, the results suggest using ContrattoA as support to human annotators for more complex, laborious or recurring contracts (e.g., rental agreements for a rental agency) and for which annotation fatigue may lead to a significant decrease in quality of manual annotation.

4.3. Threats to Validity

The results of the experiment have been influenced by several factors revealed by the subjects in interviews after the experiment. The generation of lexical patterns was based on a handful of contracts, and as a result, precision and recall were generally quite low. Two participants admitted being reluctant to using a web-based tool for the annotation, resulting in longer annotation times. One participant confused core contract concepts (obligation and power). The misinterpretation significantly influenced the results of the experiment since a discrepancy between a human annotator and ContrattoA was always resolved in favor of the human annotator. Another significant threat to validity of the experiment is that concepts of the annotation have a significant level of ambiguity and are frequently interpreted in different ways by different annotators. Moreover, even though the gold standard has been double-checked, an element of subjectivity in semantic annotation remains [5]. The annotation of ContrattoA is based on the recognition of specific lexical indicators (e.g., modal verbs for obligations) but the vocabulary used may diverge in different domains, leading to a decrease in the performance of automatic annotation. The eBNFs for structure identification are derived from a heuristic approach, thus the accuracy is bound to the types of contracts. As such, an assessment of the performance of the tool is highly variable depending on the type of contract, and on the subjects of an experiment, including the annotator(s) defining the gold standard.

External validity of our study is concerned with the generalizability of the results to other contracts. Given the limited availability and access to law students, the sample size of participants is insufficient to obtain statistical significance. The results of our investigation are encouraging but preliminary, so they need to be confirmed by other experiments including a larger set of participants, both expert and non-expert, and other types of contracts adopted from other domains.

Internal validity—factors affecting subject performance during the study—is also critical. The skills of the subjects involved in the experiments were appropriate to the objectives of our preliminary investigation, though more pre-experiment tutoring might have helped improve annotation results. Moreover, there was no bias of the subjects towards the topics covered by the contracts used for the experiments.

5. Refinement into ContrattoA 2.0

This section contains an evaluation of ContrattoA 2.0, a refinement of ContrattoA 1.0 intended to improve the quality of annotation for powers and obligations. The source of the difficulty in distinguishing between powers and obligations is primarily conceptual. A power is a special case of right, which is a dual concept to an obligation. If A has an obligation towards B, then B has a right towards A. Moreover, If B has a power towards A, then B has the right to alter (cancel, suspend, etc.) an obligation or power of A. At the

same time, the ability to identify obligations and powers is a key element of the contract annotation process as these concepts define the terms and conditions of a contract.

5.1. Experimental Setup

To improve the annotation process, we obtained 10 additional standard contracts from the Internet with a diverse range of styles and content. Such contracts have been annotated for obligations and powers; the annotation process led to the identification of 147 obligations and 62 powers. These were double-checked by two experts in annotation with no significant differences identified. Subsequently, we analyzed the contracts to identify significant words and patterns to be included in a new eBNF grammar for obligations and powers. The analysis performed suggested including verbs such as ‘entitle’, ‘terminate’ and ‘suspend’ to recognize powers; or verbs such as ‘agree’ and ‘acknowledge’ to define obligations. Moreover, we considered the possibility of complementing concept annotation with structural annotation to increase the precision and recall for the two concepts. For example, obligations and powers are often presented as lists of clauses with formatting tags (end-of-line, bullet, etc.). Accordingly, such structural tags have been used to improve annotation performance.

5.2. Experimental Results

Testing of ContrattoA 2.0 required the annotation of 10 contracts using the tool. In a first step, the annotation was performed relying on the grammar previously used for the experiment described in Section 4.2, to subsequently compare the results with the annotation performed using the improved ContrattoA 2.0 annotation grammar. The refined lexical patterns have been tested through an iterative process to isolate the impact of each change in grammar on the quality of annotation (Table 3). Of course, there were tradeoffs in that improvements in performance for obligations led to inferior performance for powers and conversely. The grammar of ContrattoA 1.0 led to excellent precision—93% for obligations and 84% for powers—and poor recall—25% for obligations and 43% for powers. For ContrattoA 2.0, there was comparable precision with significantly increased recall, specifically, 78% for obligations and 74% for powers. As such, the refined lexical pattern based on a larger number of sample contracts, allowed similar precision for obligations and a substantial increase (+22%) for recall. In general, recall is more important than precision for large contracts, since high recall allows the annotator to focus on what has been annotated by the tool, as opposed to the text of the legal contract.

Table 3. Comparison of performance measures for powers and obligations.

Measure	Obligations		Powers	
	ContrattoA 1.0	ContrattoA 2.0	ContrattoA 1.0	ContrattoA 2.0
Precision	93%	94%	84%	90%
Recall	25%	78%	43%	74%
F-measure	39%	85%	57%	81%

Despite the effort required for refinement, the results are encouraging as they lend support to the possibility of efficiently annotating a diverse set of contracts using a common grammar for powers and obligations. It remains to be seen whether similar results can be obtained with different sets of contracts without further refinements of the tool. The identification of lexical patterns could be improved with the use of Machine Learning (ML) techniques and the use of such tools is a topic for future work. For further development and experimentation, the time and effort required to refine the tool could be accounted for in the evaluation of tool support to human annotation.

The work in [5] includes an empirical study where annotators with expertise in Law were asked to annotate legal text and the resulting annotations were compared. The study concluded that there is subjectivity in annotating legal text even among experts such that an

expert's annotation judged with another expert's annotation as gold standard only scores in the range of 0.80–0.90 for precision and recall. From these results, we conclude that ContrattoA 2.0 scores for precision and recall for obligations and powers approach the performance of human experts.

5.3. Threats to Validity

ContrattoA 2.0 suggests that good quality annotations can be generated for a broad range of contracts with minor refinements to the two eBNF grammars for semantic and structural annotation. However, such a result needs to be further evaluated using a larger number of contracts as input, and experiments with a larger number of subjects to attain statistical significance. Secondly, the influence of subjectivity in determining a gold standard, as discussed in Section 5.2, needs to be further explored. Thirdly, the annotation process has been performed by annotation experts, rather than legal ones. This was not deemed to be a significant threat to validity as annotators have had much practice in annotating contracts with consistent annotation results. Finally, differently from ContrattoA 1.0, the improvement of the tool has not been tested with subjects outside the project team. Considering the influencing factors reported by subjects of the ContrattoA 1.0 experiment, the precision and recall results of ContrattoA 2.0 may be lower when used by external participants compared to those reported in Table 3.

6. Related Work

The possibility of extracting requirements from texts has been investigated for decades [14]. The inception of Blockchain, Smart Contracts and LegalTech has resulted in an increasing number of projects and commercial applications for metadata extraction in legal documents and semi-automation in the drafting and execution of legal contracts.

Many projects—such as eBrevia (<https://ebrevia.com/> (accessed on 7 August 2022)), LawGeex (<https://www.lawgeex.com/> (accessed on 7 August 2022)), Prose (<https://tryprose.com/home> (accessed on 7 August 2022)), and Concord (<https://www.concordnow.com/> (accessed on 7 August 2022))—support the extraction of information from contracts, to speed up their review or to ensure compliance. Among them, a few concern Ricardian contracts, i.e., contracts that are readable by both human and machine (<http://webfunds.org/guide/ricardian.html> (accessed on 7 August 2022)). Similar initiatives have been developed in academia, such as the Computable Contract project (<http://compk.stanford.edu/> (accessed on 7 August 2022)), within the Stanford Project for Legal Informatics (<https://law.stanford.edu/codex-the-stanford-center-for-legal-informatics/> (accessed on 7 August 2022)), that aims to create a Contract Description Language (CDL) representing the terms and conditions of a contract in a machine-readable format. However, none of these projects addresses the structure and content analysis steps for transforming a contract written in NL into a formal specification.

Many academic studies refer to the extraction of requirements from legal documents that include legal contracts as well as laws, regulations, court orders, etc. Among them, a significant number of studies focus on structural annotation such as [15] that relies on an experimental approach for the extraction of contract elements combining ML and manually written post-processing rules. That paper's contribution is the idea that structural annotation rules can be learned from a benchmark, differently from ContrattoA where they are manually constructed into a structure model defined by a grammar. A combined manual and ML approach like the one proposed by [15] could improve the identification of contract structure elements, but training ML algorithms requires the availability of a large set of sample contracts. The approach proposed in [16] relies on normalized NL based on deontic logic and extracts normative statements and conditionals together, along with existing relationships among them. In the context of formal modelling legal documents [17], proposes Fact-Based Modelling (FBM) to annotate and generate diagrams that represent potential scenarios in the application of tax law and traceability. In the proposed approach, the re-elaboration of legal text is aimed at facilitating IT implementations. Similarly, FBM

is used by [18] together with Cognitation—a legal analysis software tool—to analyze legislation and identify its functional structure. The framework supports the identification of a domain model by evaluating rules and activities associated with such rules. However, the two approaches relying on FBM require significant manual support and are not aimed at the generation of formal specifications. Another approach identifies structural elements to define the most relevant parts of a text to summarize it by relying on NLP, rule-based and statistical methods [19]. The approach does not deal with relationships between legal elements, nor does it offer a graphical representation.

Studies on the identification of suitable concepts for a contract—such as obligations, assets, parties—are mostly based on subsets of the contract ontology used in ContrattoA. A limited number of studies report on annotation performance for the concepts identified in our ontology and that we experimentally tested. The approach of [20] relies on a Provision Automatic Classifier to detect and classify legal provisions. The approach obtains good results in the identification of obligations and a subset of rights or anti-rights, such as permission and prohibition. Breaux et al. [21] propose a process called Semantic Parametrization to discriminate between rights and obligations. Strategies are proposed to identify and resolve ambiguities based on the use of restricted natural language. The approach is not automated and does not account for powers. Wyner et al. [22] test in a case-study the annotation of a mix of structure and content based on an online legal annotator tool, General Architecture of Text Engineering (GATE). However, the annotation includes domain knowledge and is significantly specialized for the jurisdiction of a legal document. Kiyavitskaya et al. [23] use the Cerno framework to semantically annotate documents and is a precursor of the ContrattoA project. The framework, based on parsing and markup, relies on the structural transformation system TXL, was tested on two regulations, and proved useful in supporting human structural annotation but the results are less promising for semantic concepts. Cerno was subsequently refined into GaiusT [5], which was similarly tested on a large legal corpus. The tool proved useful in decreasing annotation time for human annotators, also good performance measure in terms of precision and recall. GaiusT did not include the concept of power, as it was intended for legal documents in general, and does require significant improvements in recall. Another approach focuses on the communication difficulties arising from the translation of contracts into different languages [24]. Consistency in use in contract terms is evaluated using the Herfindahl-Hirshman Index (HHI), a common measure for market concentration. The result suggests the potential use of HHI to verify contract term consistency. However, ContrattoA is currently built for semantic annotation of English.

To automatically identify concepts in a contract, we exploit an ontology for contracts. In [25] a legal ontology is obtained from a legal text relying on an SE platform and a linguistic analysis tool. A conceptual tree is obtained to identify relationships and obtain a contract ontology. Similarly, in [26] an ontology is extracted from a legal document written in simplified NL that identifies structural, lexical and domain elements, as well as their inter-relationships. However, such a process focuses on identifying a domain model for a particular contract, i.e., step (d) in the transformation process proposed in the introduction. Regarding the annotation of a contract, a few ontologies have been proposed—such as the Public Procurement Ontology (<http://contsem.unizar.es/def/sector-publico/pproc.html> (accessed on 7 August 2022)) and PROMS (<https://confluence.csiro.au/public/PROMS/proms-ontology> (accessed on 7 August 2022))—however, they are either too specialized or too broad relative to the ontology adopted for ContrattoA.

There are a number of upper ontologies for the legal domain, i.e., covering any kind of legal document. Among them, we note UFO-L, a foundational ontology for the legal domain [10]. UFO-L extends UFO, a descriptive foundational ontology, with concepts from the legal domain based on Alexy's Theory of Constitutional Rights [27]. The ontologies that share more commonalities with ours are those used for GaiusT [5] and NómoS [6]. However, these ontologies are missing concepts that are specific to legal contracts, notably the concept of power.

An increasing number of works focus on service contracts. Nardi et al. [28] propose a core reference ontology called UFO-S for services. This ontology is grounded on UFO and aims to provide a domain independent conceptual model for services. Griffo et al. [29] rely on UFO-S and extend the ArchiMate enterprise architecture language to develop the UFO-S ontology. However, the process of extracting contract elements from text is not supported. Similarly, concerning service contracts, Griffo et al. [30] explore an approach bridging the gap between contract languages—for formal representations of contracts—and other approaches, such as ArchiMate that does not support the representation of rights and obligations. Another conceptual model to support the automatic extraction of software requirements from legal documents is proposed in [31] and attempts to harmonize the variety of semantic legal metadata proposed in requirements engineering to derive extraction rules based on constituency and dependency parsing. However, the proposed approach is not specialized to legal contracts and does not address structural analysis.

Several formal specification languages have been considered as the target output of the five-step translation process. One of the first efforts to transform a contract in NL into a formal specification exploits the Business Contract Language [32] and is based on Propositional Logic in an event-driven language. Different legal modalities are supported, such as obligations, permission, prohibition and violation although the notions of power, termination and suspension are not supported. RuleML [33] and Oasis LegalRuleML [34] offer an XML-based language with a holistic approach to legal document management and is based on deontic logic that supports obligations, prohibitions and permissions.

LegalRuleML allows the modelling of constitutive and prescriptive rules. However, it offers a lower level of formality compared to other approaches and normative monitoring, runtime changes are not specifically addressed. CL specification language [35] combines deontic logic and Propositional Dynamic Logic (PDL) where modalities are applied only over structured actions. CL does not grant a high level of expressiveness but allows the performance of automatic analysis of properties of the contract. The logic proposed does not support powers and temporal constraints. Formal Contract Language (FCL) [36] has been proposed to provide specific semantics for NL that enables the transformation of legal text into a formal specification. As such, it allows contract templating, parametrization and monitoring but does not provide runtime flexibility. In Azzurra [37], contracts are modelled as business processes through a graphical notation and represent the relations between social concepts such as roles and agents based on the execution of a set of commitments. However, Azzurra has limited expressiveness and does not support the notion of power.

7. Conclusions

We have proposed a process for transforming legal contracts expressed in natural language into formal specifications. Focusing on the second step of the process, we use GaiusT, a semantic annotation platform, to build ContrattoA, a tool for semantically annotating contracts. ContrattoA was evaluated in an empirical study that suggested that the performance of inexperienced human annotators can be substantially improved when supported by the tool. Good quality annotations can be generated for a broad range of contracts with minor refinements to the two eBNF grammars although domain knowledge is required.

The development of the prototype and the experiments we conducted highlight a variety of critical issues to be addressed in refining ContrattoA in our future work. The lexical patterns for several concepts, notably those for situations, need to be improved. Moreover, further evaluation needs to be conducted with human subjects, including both novice and expert annotators. We also plan to experiment with domain-specific models for contract structure and content to study their impact on the performance of ContrattoA.

Our longer-term plans include tackling other steps of the transformation process with Symboleo as the target specification language, with the aim of creating an environment composed of several tools to facilitate the transformation for a legal contract in NL [38].

Author Contributions: Conceptualization, J.M. and L.M.; methodology, L.M. and N.Z.; software, N.Z.; writing—original draft preparation, M.S., N.Z. and L.M.; writing—review and editing, L.M. and J.M.; experiment design, L.M., N.Z. and M.S.; experiment execution, M.S., N.Z. and L.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the experiment.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Clarke, E.M.; Grumberg, O.; Peled, D.A. *Model Checking*; MIT Press: Cambridge, UK, 2018.
- Sebastiani, R.; Trentin, P. OptiMathSAT: A tool for optimization modulo theories. *J. Autom. Reas.* **2020**, *64*, 423–460. [\[CrossRef\]](#)
- Szabo, N. Smart contracts: Building blocks for digital markets. *EXTROPY J. Transhumanist Thought* **2020**, *18*, 2.
- Governatori, G.; Idelberger, F.; Milosevic, Z.; Riveret, R.; Sartor, G.; Xu, X. On legal contracts, imperative and declarative smart contracts, and blockchain systems. *Artif. Intell. Law* **2018**, *26*, 377–409. [\[CrossRef\]](#)
- Zeni, N.; Kiyavitskaya, N.; Mich, L.; Cordy, J.R.; Mylopoulos, J. GaiusT: Supporting the extraction of rights and obligations for regulatory compliance. *Requir. Eng.* **2015**, *20*, 1–22. [\[CrossRef\]](#)
- Zeni, N.; Seid, E.; Engiel, P.; Mylopoulos, J. N6mosT building large models of law with a tool-supported process. *Data Knowl. Eng.* **2018**, *117*, 407–418. [\[CrossRef\]](#)
- Soavi, M.; Zeni, N.; Mylopoulos, J.; Mich, L. Contract—from Legal Contracts to Formal Specifications: Preliminary Results. In Proceedings of the IFIP Working Conference on the Practice of Enterprise Modeling, Riga, Latvia, 25–27 November 2020; Springer: Cham, Switzerland, 2020; pp. 124–137.
- Daskalopulu, A. Modelling legal contracts as processes. In Proceedings of the 11th International Workshop on Database and Expert Systems Applications, London, UK, 4–8 September 2000; pp. 1074–1079.
- Parvizimosaed, A.; Sharifi, S.; Amyot, D.; Logrippo, L.; Roveri, M.; Roudak, A.; Rasti, A.; Mylopoulos, J. Specification and Analysis of Legal Contracts with Symboleo. *Softw. Syst. Model.* **2022**, *in press*.
- Griffo, C.; Almeida, J.P.A.; Guizzardi, G. Towards a legal core ontology based on Alexy’s theory of fundamental rights. In Proceedings of the Multilingual Workshop on Artificial Intelligence and Law, ICAIL, San Diego, CA, USA, 8–12 June 2015.
- Hohfeld, W.N. Some fundamental legal conceptions as applied in judicial reasoning. *Yale Law J.* **1913**, *23*, 28–59. [\[CrossRef\]](#)
- Bix, B.H. *Contract Law: Rules, Theory, and Context*; Cambridge University Press: Cambridge, UK, 2012.
- Manning, C.; Schutze, H. *Foundations of Statistical Natural Language Processing*; MIT Press: Cambridge, MA, USA, 1999.
- Allen, T.; Widdison, R. Can computers make contracts? *Harvard J. Law Technol.* **1996**, *9*, 25–52.
- Chalkidis, I.; Androutsopoulos, I.; Michos, A. Extracting contract elements. In Proceedings of the 16th edition of the International Conference on Artificial Intelligence and Law, London, UK, 12–16 June 2017; ACM: New York, NY, USA, 2017; pp. 19–28.
- Libal, T.; Pascucci, M. Automated reasoning in normative detachment structures with ideal conditions. In Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law, Montreal, QC, Canada, 17–21 June 2019; pp. 63–72.
- Nijssen, S.; Dulfer, D.; Bollen, P.; Rozendaal, J. Meaning Based Structured Legal Code. In Proceedings of the OTM Confederated International Conferences “On the Move to Meaningful Internet Systems”, Rhodes, Greece, 23–28 October 2017; pp. 193–201.
- Bulles, J.; Cartigny, B.; Bollen, P. Analyzing the new 2019 dutch environment and planning act. In Proceedings of the OTM Confederated International Conferences “On the Move to Meaningful Internet Systems”, Rhodes, Greece, 23–28 October 2017; pp. 163–172.
- Grover, C.; Hachey, B.; Hughson, I.; Korycinski, C. Automatic summarisation of legal documents. In Proceedings of the 9th International Conference on Artificial Intelligence and Law, Edinburgh, UK, 24–28 June 2003; pp. 243–251.
- Biagioli, C.; Francesconi, E.; Passerini, A.; Montemagni, S.; Soria, C. Automatic semantics extraction in law documents. In Proceedings of the 10th International Conference on Artificial Intelligence and Law, Bologna, Italy, 6–11 June 2005; pp. 133–140.
- Breaux, T.D.; Vail, M.W.; Anton, A.I. Towards regulatory compliance: Extracting rights and obligations to align requirements with regulations. In Proceedings of the 14th IEEE International Requirements Engineering Conference (RE’06), St. Paul, MN, USA, 11–15 September 2006; pp. 49–58.
- Wyner, A.Z.; Peters, W.; Katz, D. A Case Study on Legal Case Annotation. In Proceedings of the International Conference on Legal Knowledge and Information Systems (JURIX), Bologna, Italy, 11–13 December 2013; pp. 165–174.
- Kiyavitskaya, N.; Zeni, N.; Breaux, T.D.; Ant6n, A.I.; Cordy, J.R.; Mich, L.; Mylopoulos, J. Automating the extraction of rights and obligations for regulatory compliance. In Proceedings of the International Conference on Conceptual Modeling, Barcelona, Spain, 20–24 October 2008; pp. 154–168.
- Gašpar, A.; Seljan, S.; Kučič, V. Measuring Terminology Consistency in Translated Corpora: Implementation of the Herfindahl-Hirshman Index. *Information* **2022**, *13*, 43. [\[CrossRef\]](#)

25. Amardeilh, F.; Laublet, P.; Minel, J.L. Document annotation and ontology population from linguistic extractions. In Proceedings of the 3rd International Conference on Knowledge Capture, Banff, AB, Canada, 2–5 October 2005; pp. 161–168.
26. Amato, F.; Mazzeo, A.; Penta, A.; Picariello, A. Building RDF Ontologies from Semi-Structured Legal Documents. In Proceedings of the International Conference on Complex, Intelligent and Software Intensive Systems, Barcelona, Spain, 4–7 March 2008; pp. 997–1002.
27. Alexy, R.; Rivers, J.; Rivers, P. *A Theory of Constitutional Rights*; Oxford University Press: Oxford, UK, 2002. Available online: <https://books.google.it/books?id=4G1U8JMSJDcC> (accessed on 15 July 2022).
28. Nardi, J.C.; de Almeida Falbo, R.; Almeida, J.P.A.; Guizzardi, G.; Pires, L.F.; van Sinderen, M.J.; Fonseca, C.M. A commitment-based reference ontology for services. *Inf. Syst.* **2015**, *54*, 263–288. [[CrossRef](#)]
29. Griffo, C.; Almeida, J.P.A.; Guizzardi, G.; Nardi, J.C. Service contract modeling in enterprise architecture: An ontology-based approach. *Inf. Syst.* **2021**, *101*, 101454. [[CrossRef](#)]
30. Griffo, C.; Almeida, J.P.A.; Guizzardi, G.; Nardi, J.C. From an ontology of service contracts to contract modeling in enterprise architecture. In Proceedings of the International Enterprise Distributed Object Computing Conference (EDOC), Quebec City, QC, Canada, 10–13 October 2017; pp. 40–49.
31. Sleimi, A.; Sannier, N.; Sabetzadeh, M.; Briand, L.; Dann, J. Automated extraction of semantic legal metadata using natural language processing. In Proceedings of the International Requirements Engineering Conference (RE'18), Banff, AB, Canada, 20–24 August 2018; pp. 124–135. [[CrossRef](#)]
32. Governatori, G.; Milosevic, Z. A formal analysis of a business contract language. *Int. J. Coop. Inf. Syst.* **2006**, *15*, 659–685. [[CrossRef](#)]
33. Athan, T.; Governatori, G.; Palmirani, M.; Paschke, A.; Wyner, A. LegalRuleML: Design principles and foundations. In Proceedings of the Reasoning Web International Summer School, Berlin, Germany, 31 July–4 August 2015; pp. 151–188.
34. Governatori, G. Representing business contracts in RuleML. *Int. J. Coop. Inf. Syst.* **2005**, *14*, 181–216. [[CrossRef](#)]
35. Prisacariu, C.; Schneider, G. Abstract specification of legal contracts. In Proceedings of the 12th International Conference on Artificial Intelligence and Law, Barcelona, Spain, 8–12 June 2009; pp. 218–219.
36. Farmer, W.M.; Hu, Q. FCL: A formal language for writing contracts. In Proceedings of the 4th IEEE International Workshop on Formal Methods Integration—The IEEE FMI 2016, Pittsburgh, PA, USA, 28–30 July 2016; pp. 190–208.
37. Dalpiaz, F.; Cardoso, E.; Canobbio, G.; Giorgini, P.; Mylopoulos, J. Social specifications of business processes with Azurra. In Proceedings of the International Conference on Research Challenges in Information Science (RCIS), Athens, Greece, 13–15 May 2015; pp. 7–18.
38. Soavi, M.; Zeni, N.; Mylopoulos, J.; Mich, L. Contratto—A method for transforming legal contracts into formal specifications. In Proceedings of the International Conference on Research Challenges in Information Science (RCIS), Barcelona, Spain, 17–20 May 2022; pp. 338–353.