Acquiring and Modelling Legal Knowledge Using Patterns: an application for the Dutch Immigration and Naturalisation Service

Patries Kordelaar, Freek van Teeseling, Edwin Hoogland

Pharosius, Netherlands
patries@pharosius.nl

Be Informed, Linie 620, 7325 DZ Apeldoorn, Netherlands
{f.vanteeseling,e.hoogland}@beinformed.nl

Abstract. The Dutch Immigration and Naturalisation Service is replacing its existing paper based case system with a fully electronic system with integrated decision support based on ontologies. In nature a large proportion of the organisations knowledge is based on law and regulations. In acquiring this legal knowledge for modelling we faced the classic knowledge acquisition bottleneck due to communication problems between experts from different background. To overcome this bottleneck a methodology was developed to transfer these laws in semantic knowledge models, based on legal patterns. In this paper we describe different archetypes of law and how these are transformed in models with the help of patterns. The patterns are thereby used for both translation of the true meaning of the law and for inclusion in the semantic models for automatic execution.

Keywords: Legal Patterns, Knowledge Representation, Semantics, Acquisition, Methodology

1 Introduction

The Dutch Immigration and Naturalisation Service (IND) [8] decided, as a reaction to a review of its processes by the Dutch Court of Audit [3], to completely replace its existing (mostly) paper based case system and some form of decision support (decision trees). The new application had to be an electronic case system with integrated decision and process support for the internal knowledge worker, including a directly connected front office for clients [9]. The case system consists of a Siebel workflow system communicating within a SOA architecture with intelligent components for the decision and process support made with the knowledge-based platform Be Informed (www.beinformed.nl). The core functionality of Be Informed is a semantic modelling environment in which models can be made that are used by the

inferencer for classification, calculation and decisions tasks. The new application replaces the current operational cluster of systems, of which the ICT maintenance costs are way too high.

The IND is responsible for enforcing Immigration Law in the Netherlands and handles around 255.000 permit applications per year. Since the IND is an enforcement organisation the main body of knowledge to include in the new IND application consists of legal knowledge. In acquiring and modelling this legal knowledge we faced a knowledge acquisition bottleneck. Actors with different backgrounds, legal experts and knowledge engineers, had to work together to make sound models. However, while the legal experts could not see how a piece of text was not sufficient to be machine-interpreted, the semantic engineers could not understand the interpretations of the legal experts. One of the main challenges of the project was getting the two groups closer to each other. Therefore we decided to standardize the communication between legal experts and knowledge engineers by introducing legal patterns as a means for exchanging information. Thereby it was important for the IND that the patterns would be intuitive and easy to use by the legal experts, since they are the most scarce resource within the IND. As an answer we came up with legal patterns based on legal theory. Practice has shown that these patterns could be easily used by the legal experts and that, as a side-effect, they could also be used for finding defects in the legislation.

In this paper we describe the development of the legal patterns. First we show where in the modelling process the patterns play a role. Next we shortly position the legal patterns in the light of existing literature on patterns. Then we turn our attention towards legislation. What is the nature of legislation and how did we use that nature in constructing our legal patterns. We conclude this paper with some concluding remarks and plans for future research.

2 Legal patterns in the modelling process

The process of implementing legislation in the IND application is shown in figure 1.
Interpretation of legislation plays a role in the phases: analyzing the legal source texts, formulating the IND interpretation of these texts and modelling the interpretation. In these phases over 20 knowledge engineers, many domain experts and other stakeholders participated. The modelling domain consists of over 30.000 concepts and a multitude of relations. On that scale it is clear that there is a strong need to structure the modelling process.
Before the introduction of the legal patterns we encountered a major delay in the modelling process transforming the interpretation to semantic models. The IND interpretation of the legal source was laid down by the legal experts in a natural language representation, with a form of cross-linking between fragments, and the representation is accompanied by a reference to the original source text. For the knowledge engineers it was not straightforward to get from the interpretation in natural language to the much more restricted language of Be Informed (a restriction that is essential for the automatic execution of the model). The knowledge engineers

still had a lot of questions about the right interpretation of the text in natural language and many times had to disambiguate the text themselves before they could model it. The (free format) textual representation and cross-links proved not to be explicit nor structured enough. Not being legal experts the engineers turned to the legal experts for the answers to their questions. Like in many organizations, legal experts are a scarce resource and the modelling process was severely delayed by these extra iterations. In response to this we developed the solution to standardize the patterns used in the communication between legal experts and knowledge engineers. This resulted in the legal patterns described in section 3.3

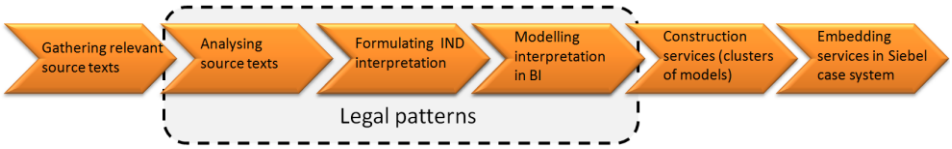

Fig. 1. Stages in the modelling chain

In the new process the developed legal patterns were introduced and used during the interpretation of the legislation. The patterns form the "glasses" with which the legal experts look at the legislation. When they recognize a norm in the legislation they make an IND interpretation of the norm (with accompanying definitions and fictions) and represent this in a structure according to a pattern. They highlight and name the different elements of the norm (and the different types of definitions, references and fictions) as agreed upon in the legal pattern. This representation is taken by the knowledge engineers to make a Be Informed model. Using the patterns, the legal interpretation was easily transformed into the semantic models. These models are used by the Be Informed inferencer for automatic reasoning tasks such as classification, calculation and deciding.

The patterns we describe in this paper have strong similarities to the patterns mapped to legal problems, like described by Gangemi [4]. His ODPs focus on legal tasks, one of which is knowledge extraction. The legal patterns we have developed for the IND can be seen as specific instances of patterns used within a broader pattern for knowledge extraction such as Gangemi's. The approach of Gangemi is towards automatic extraction. In contrast, our patterns have been used for acquisition purposes by humans, but as we will discuss in our concluding remarks our legal patterns certainly have the potential to be used for automatic knowledge extraction. Human readability however remains an important issue if the patterns are to be used in an acquisition task. Like most other ODP approaches Gangemi uses the patterns for ontology design rather than for supporting the knowledge acquisition process. This might be the reason that most formulations of ODP's are rather technical, making them unfit for use by legal experts. Graphs, and even more syntax like OWL or RDF or the Be Informed models appear to legal experts as "programming" language which they consider technical and very abstract and they found it very hard to map this to their text based world of legal knowledge. We therefore decided to turn to legal theory as the basis for the formulation of our legal patterns, instead of the final

ontology design, trying to stay as close as possible to the vocabulary of legal experts. In doing this, our approach resembles the lexico-syntactic branch of ODP development like described by Aguado et. al. [1], although they have a more didactical purpose.
In the next chapter we describe how we developed the legal patterns using legal theory. Following, examples of the patterns are given.

3 Developing legal patterns based on legal theory

3.1 Characterizing Legislation

Norms are the core element of legislation. A norm says which behaviour is allowed or forbidden. A norm can also say which situation may exist and which may not. Modelling legislation starts with acquiring and classifying the norms. In the following sections we will discuss a systemic analysis of legal norms, which forms the base of our patterns. Two key issues are important; form and typology.

First, a norm has a prototypical form. Take as an example the following sentence: "*As the sun sets car-drivers should turn on the lights of their vehicle*" [13]. This norm sentence bears all the constituting elements of the prototypical norm (see also figure 2):

- *Subject* or addressee of the norm: to whom is the rule directed (in the example "car-drivers");
- *Objective* or action part of the norm: What behaviour or situation does it allow or prohibit (in the example the phrase "turn on the lights of the vehicle")
- *Deontic Operator* of the norm: What is the normative type of the rule, is it a prohibition or a permission (in the example presented by the word "should")
- *Condition* of the norm: Under what circumstances does the rule apply (in the example the phrase "as the sun sets")

When the norm sentence is without flaws we can find all the elements mentioned above in the norm.

Since our legal patterns rely heavily on the prototypical norm structure, we must make sure that all norm elements are recognizable in the norm. This means that the legal experts have the task to make the norms complete and designate the different parts. The legal patterns help them to identify omissions in the norm, so that they can repair these. Given that we aim at making models that can be automatically executed, all other defects must be repaired. As a final resort some vagueness can be handed over for interpretation to the end-user (but we have to be sure then that the end-user has the competencies to make the interpretation).

Next to the typical form of norms, there are also several types of norms. Sartor [15] for instance distinguishes norms that forbid behaviour, that allow behaviour, that state that persons x should do y in situation z, state that people under certain conditions are entitled to z, etc. Literature contains a lot of different divisions in types of norms, [2,10, 12, 13, 14, 15]. For the development of our legal patterns we have used a combination of these theories. In this manner we could produce a typology of norms that was:

- well understood by the legal experts of the IND because they resemble closely the vocabulary they are normally using;
- detailed enough to be able to construct patterns that result in semantic models that can be used with the Be Informed inferencers.

As a first division we used the theory of Larenz [10] about independent and dependent legal rules. The independent rules are the norms as stated above. They can function as a separate rule by themselves. Dependent rules clarify parts of the norms. For instance a rule defining the concept of a car-driver in our example norm is a dependent rule in the sense of Larenz. Dependent rules cannot function as a separate rule but get meaning only in combination with an independent rule. This does not mean that dependent rules are not normative, but that we can only make sense of the dependent rule in the context of an independent rule. Larenz mentions as dependent rules definitions, references and fictions. As a result we made not only patterns for norms, but also for definitions, references and fictions[2].

Analyzing the literature we derived a major distinction in norm types between duties and permissions. Duties reflect the imposing side of the law. They state what is prohibited and what should be the case. Permissions on the other hand express what is permitted; we recognize dispensations and approvals (a right is a 'strong" approval). Another distinction that is important in our construction of legal patterns is that norms can see upon conduct and upon situations (tun-sollen and sein-sollen [14]). Taking all these distinctions into account we came up with legal patterns for conduct and situations for all kind of duties and permissions. In section 3.3 we give some examples of the legal patterns.

3.2 The representation language of Be Informed

Our source language is legislation where the goal representation is the semantic Be Informed representation language. All examples given in the following sections will thus be illustrated by graphically represented semantic models, native to Be Informed. The reader is trusted with the possibilities of using the very same patterns in any semantic standard, like for instance OWL and RDF(S). For clarity we briefly describe the elements of the Be Informed representation language.

[2] In this paper we only discuss the norm patterns. For research about patterns for definitions and references we refer to [11, 12].

The Be Informed knowledge representation formalism can be characterized as a triples-based semantic network. Concepts and strong typed relations are the main elements of the language. If needed, these concepts and relationships can be further specified with formula and conditions. The concepts that are important within a (e.g. legal) domain are represented in Be Informed concepts. Often these are nouns or noun-groups. A concept can typically have a Boolean value, but can also be a number, date, or string. A concept is defined once and only once in the model with relations to other concepts, formula and conditions. Once a concept is defined, this concept can be referred to from any other concept in the model. Among the relations the Be Informed reasoner reasons with are taxonomical relations (e.g. instance-of) as well as causal relations (e.g. requires). Typical tasks that the out-of-the-box inferencers of Be Informed automatically perform are classification, decision and calculation. The inferencers are based on propositional logic and use backward chaining mechanisms.

Before we turn to examples of the legal patterns we developed we have to take one last issue into account. An organization like the IND, and certainly employees of the IND with different roles, can apply norms from several distinctive viewpoints. A decision-maker for instance can apply the norms answering the question whether a person is compliant with the norms (only then the person gets the entrance permit) or whether a person violates the norms (this sanction is issued only if that person violates the rule). A legislator however might only be interested in what the norm is, without applying the norm (an informative position). In constructing the legal patterns we also have to take the viewpoint of the user into account.

3.3 Examples of norm patterns in Be Informed

We start with our fictitious exemplary norm: *As the sun sets car-drivers should turn on the lights of their vehicle.* This norm is a *command* stating which behaviour a car-driver must exhibit in case the condition(s) hold. Suppose we are interested in knowing whether someone has *violated* this rule (viewpoint). The representation in natural language, highlighting the different parts of the norm is shown in figure 2.

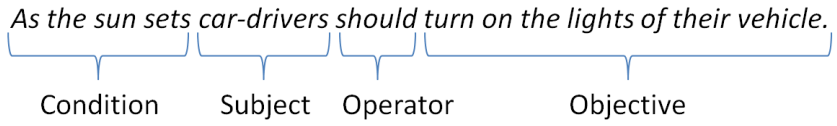

Fig. 2. Natural language representation of a command

The corresponding Be Informed model is shown in figure 3.

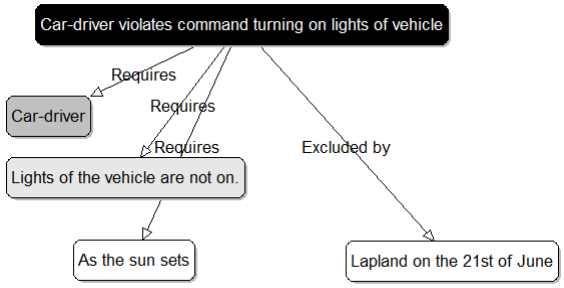

Fig. 3. Be Informed model representing a commanding norm

The central concept of the model (car-driver violates command "turning on the lights of the vehicle") reflects the question we're interested in. This concept also makes clear that the norm is a command. It is a Boolean concept, so in this case the question is either answered with a "yes" or "no". The other concepts determine the value of the central concept. The central concept becomes true when:

- the *subject* at hand really is a car-driver (behind this concept other models that define the notion of a car-driver can be placed),
- the *condition* " as the sun sets" is fulfilled; and when
- the negation of the *object* is true (the lights of the vehicle are *not* on).

In the final model we recognize all the elements of the norm that were already highlighted in the natural language representation.

If we extend this rule with an exception, for instance "the rule holds except in Lapland on the 21st of June", we add in the model an exception concept (see also figure 5). The relation between this concept and the central concept is another causal relation in Be Informed: "Excluded by".

The natural language representation of a second example, taken from the "*Wet tot vaststelling van bepalingen betreffende het opium en andere verdovende middelen*", 1928, section 2" (translation by the authors), is shown in figure 4.

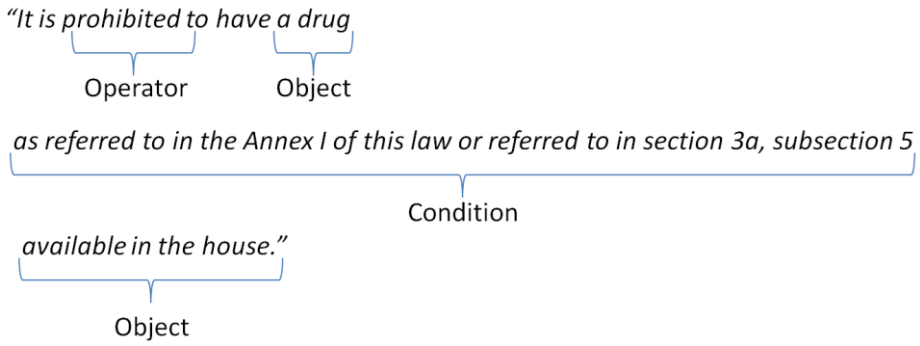

Fig. 4. Natural language representation of a prohibition[5].

We look at this norm now from an informative point of view (suppose you are a legislator interested in what the current norm is). The model is shown in figure 5.

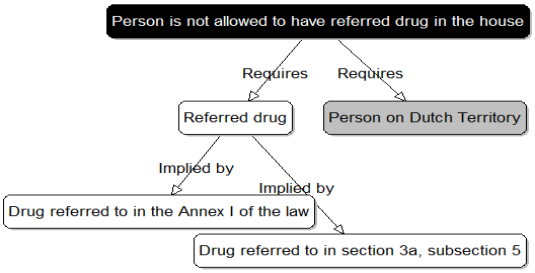

Fig. 5. Informative pattern

Notice that in this model the norm-object only plays a role in the central concept (more or less the "then" part of the rule), not in the conditional part of the model.
de model, ex lacks the norm-object completely, . This makes sense because from an informative perspective the user of the models is not interested in what happened in reality only in the "artificial" norm. Again we recognize in the resulting model the basic elements from the natural language representation of the norm.

3.4 Towards generic patterns

Figure 6 shows the abstract legal patterns of both "type of norm –perspective" combinations discussed in the examples. We see that in both cases with the help of the elements of a norm recognized in section 3.1 we can make a translation from the natural language text to a Be Informed model. We found that the same pattern can be

[5] Notice that the subject group is not mentioned in the rule (the norm is incomprehensively stated), we assume here that the group that is addressed consists of the persons that are located within the Dutch borders.

applied every time we encounter the same "type of norm-perspective" combination. As long as the elements of a norm can be recognized (and that's where the experts come in) it is not difficult to make the translation into models with the help of the patterns.

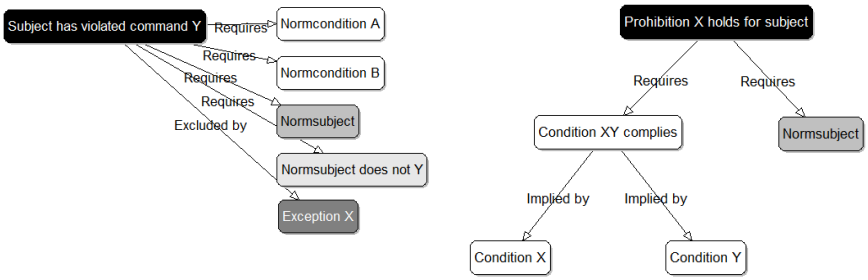

Fig. 6. Abstract legal patterns of the examples

We have constructed legal patterns for all "types of norms – perspective combinations". All legal patterns could be made using only the form and the type of the norm as well as the viewpoint of the user. In using the patterns we have found that the patterns cover all the norms we encounter so far in the Immigration legislation. At this moment legal experts as well as knowledge engineers are using the legal patterns in their acquisition and modelling processes.

4 Concluding remarks and future research

After the development of the legal patterns, workshops have been given to legal experts as well as knowledge engineers to learn them how to work with the legal patterns. At the moment about 10 legal experts and 20 knowledge engineers are using the patterns. Legal experts prove to find the patterns very intuitive. According to their evaluating statements the patterns represent the essence of legislation quite well. This is of course not surprising since the patterns find their origin in legal theory. Also the legal experts have better means to check their representation for comprehensiveness. Omissions which would lead to incomplete knowledge models are found at an early stage and can be repaired by making additional policy rules. The knowledge engineers find the representation more understandable now and also much more complete. For the knowledge engineers it is not difficult to use the patterns because they are stated in the elements they already know quite well, the Be Informed modelling elements. At this moment the first major delivery of the system is finalized and will be implemented, replacing part of the existing IND systems, in the second half of this year.

Currently, other legal based projects have started using the patterns with similar success. Mayor effort will be put in incorporating the legal patterns, used for acquisition, in the Be Informed Implementation methodology, which has already been put to the test in numerous projects, and is common practice in the company and her

partners. Adding legal knowledge acquisition as described here will benefit all these and future projects. At the same time we try to further speed up the modelling process by using the patterns for automated model generation. Another interesting approach is currently investigated by developing a text based knowledge editor, which will integrate with the graphical representations in Be Informed. This will allow (legal) experts to write down their interpretation in a specified (domain specific) syntax (possibly based on the legal patterns) which directly corresponds with the semantic models (in contrast to having to translate between the two representations) again eliminating another step in the process.

5 References

1. Aguado,G., Gómez-Pérez, A., Montiel-Ponsoda, E., Suárez-Figueroa, M.C., Natural Language-Based Approach for Helping in the Reuse of Ontology Design Patterns, in: A. Gangemi and J. Euzenat (Eds.): EKAW 2008, LNCS 5268, pp. 32–47, 2008
2. Boer, A, Legal Theory, sources of law and the semantic web, IOS press, Amsterdam, 2009, diss.
3. Dutch Court of Audit (Algemene Rekenkamer) Audit Report, http://www.courtofaudit.com/english/News/Audits/Bronnen/2005/09/Immigration_and_Naturalisation_Service
4. Gangemi, A. Introducing Pattern based Design for Legal Ontologies, in Breuker, J. et al., Law, Ontologies and the Semantic Web, IOS press 2009; doi:10.3233/978-1-58603-942-4-53, pp. 65-85, 2009
5. Heller, R., van Teeseling, F.: Knowledge Applications for Life Events, The Semantic Web: Research and Applications: 6th European Semantic Web Conference, ESWC 2009 Heraklion, Crete, Greece, May 31 June 4, 2009.
6. Heller, R., van Teeseling, F., Gülpers, M.: A Knowledge Infrastructure for the Dutch Immigration office., 7th Extended Semantic Web Conference, ESWC 2010 Heraklion, Crete, Greece, May 30 June 3, 2010.
7. Hoekstra, R, Ontology Representation: design patterns and ontologies that make sense, IOS press, Amsterdam, 2009, diss.
8. IND, Dutch Immigration and Naturalisation Service, http://www.ind.nl
9. IND Verblijfwijzer, http://www.verblijfwijzer.nl
10. Larenz, K., Methodenlehre der Rechtswissenschaft, Berlijn: Springer-Verlag, 1975
11. Maat, E. de, Winkels, R. and T. van Engers. Automated detection of reference structures in law. In T. M. van Engers, editor, Legal Knowledge and Information Systems. Jurix 2006: The Nineteenth Annual Conference, volume 152 of Frontiers in Artificial Intelligence and Applications, pp. 41-50. IOS Press, December 2006.
12. Maat, E. de and R. Winkels. Categorisation of norms. In A. R. Lodder and L. Mommers, editors, Legal Knowledge and Information Systems. Jurix 2007: The Twentieth Annual Conference Annual Conference, volume 165 of Frontiers in Artificial Intelligence and Applications, pp. 79-88. IOS Press, December 2007.
13. Ruiter, D.W.P, Bestuursrechtelijke Wetgevingsleer, Assen/Maastricht: Van Gorcum, 1987
14. Van Kralingen, R.W., Frame-Based Conceptual Models of Statute Law, Rijksuniversiteit Leiden, 1995, diss
15. Sartor, G., Fundamental Legal Concepts: A Formal and Teleological Characterisation, GS2006AILawNormativeOntologySpringer.tex, 2006