Case Frames as Contextual Mappings to Case Law in BestPortal

Rinke HOEKSTRA a,b, Arno LODDER c, Frank VAN HARMELEN a

a AI Department, VU University Amsterdam, The Netherlands
b Faculty of Law, University of Amsterdam, The Netherlands
c Computer Law Institute, Faculty of Law, VU University Amsterdam, The Netherlands

Abstract. This paper introduces case frames as a way to provide a more meaningful structure to vocabulary mappings used to bridge the gap between laymen and legal descriptions of court proceedings. Case frames both reduce the ambiguity of queries, and improve the ability of users to formulate good quality queries. We extend the BestMap ontology with a formalisation of case frame based mappings in OWL 2, present a new version of BestPortal, and show how case frames impact retrieval results compared to simple contextual mappings and a direct fulltext search.

1. Introduction

Explicit metadata in the form of Linked Data allows for a flexible way of accessing and searching the information available online. Raw data is given a context by embedding it in a network of interlinked datasets. By connecting a metadata vocabulary to those of other parties, information can become disclosed through a multitude of different metadata standards. However, publishing metadata as Linked Data does not automatically accommodate the fact that different communities may, and most often will, use different vocabularies to describe the same data. This problem frequently arises in Law, where government, citizens and businesses form different speech communities but have a shared interest. Not only are lawyers stereotypically accused of “contagious verbosity” [7], especially in case law everyday words often refer to technical legal concepts [2]. For instance, the legal concept of ‘person’ includes both legal persons (organisations, businesses etc.) and natural persons.

This divergence in vocabulary is problematic when one wants to improve access to legal information for people from a different speech community. In fact, this holds both for laymen and for lawyers. A layman will be unable to phrase his query in such a way that relevant documents are retrieved. Consider a situation where a small child is bitten by a dog. A search on dog, child and bite may retrieve the one case in which a dog bit a child, but not the plethora of other cases in which the owner of an animal can be held liable for any damages caused to a third party. A lawyer will also have to deal with the disparateness of terminology between the written law – legislation – that defines a legal concept and the case law that applies it [9].

In previous work we therefore argued for a translation approach to legal information disclosure: queries to a corpus of legal documents are translated from ordinary lan-
In this paper, we introduce the use of case frames to provide more meaningful structure to the translations used in BestPortal. We first characterise the general approach (in section 2), we then introduce the case frames (Section 3) and describe how they both reduce ambiguity and improve the ability of users to adequately formulate their query. In section 4, we discuss the implementation of these mappings in BestPortal. Section 5 presents the latest incarnation of the system, concluding with a discussion and initial evaluation results in section 6.

2. Translation Approach

In the translation approach to legal information retrieval [9,5] the queries posed by laymen are translated to the queries a legal professional would use to find relevant information. This sets this work apart from initiatives that aim to provide a formal definition of types of liability for the purpose of (semi) automatic liability attribution [12,10]. Translating layman descriptions to a legal characterisation of a case is not trivial, and has been the object of study for quite some time, e.g. [2,13].

In [5] we discussed two alternate ways for ‘lifting’ everyday descriptions to the legal level. The first approach is to provide formal definitions of legal concepts using commonsense notions, as in [9,6] and others. The problem with this approach is that it assumes reducibility of legal concepts to commonsense notions via subsumption. Also, proper definitions of legal concepts require a grounding in legal theory, which is not necessarily instrumental in a relatively lightweight task such as case retrieval. The second approach, proposed in [5], is to specify mappings between everyday descriptions of cases and the terminology a legal professional would use to find relevant case law pertaining to that case. These mappings are contextual: they focus on the co-occurrence of concepts in the description rather than one to one correspondence between everyday and legal concepts.

We use the BestMap ontology [4] as a framework for specifying context-aware mappings between two (or more) vocabularies in OWL [2]. Users of BestPortal express their query by selecting everyday ‘laymen’ concepts. This query is an OWL individual with :d_about relations to relevant concepts in a commonsense SKOS vocabulary. A stan-
standard OWL reasoner is then used to determine whether the individual matches a mapping defined in the ontology. If this is the case, any restrictions on that class using the legal vocabulary will be inferred to hold for the individual as well. We applied this methodology in the development of mappings for the legal area of unlawful acts.

Example In the dog-bites-child case used in the preceding section, we can define a mapping that translates from (generic) commonsense vocabulary to the legal tort vocabulary as follows:

\[
\text{ex:Animal\_Map} \equiv \text{:about value cv:animal} \sqcap \text{:about value cv:dangerous\_action} \sqcap \text{:about value cv:person} \\
\sqcap \text{:d\_about value lv:animal\_owner} \sqcap \text{:d\_about value lv:strict\_liability\_for\_animals} \sqcap \text{:d\_about value lv:volatility\_of\_animal\_behavior} \sqcap \text{:d\_about value lv:damage\_caused\_by\_animal}
\]

Since the concepts cv:animal and cv:person respectively have skos:narrower relations with cv:dog and cv:child, and :about is transitive over skos:narrower [4], a case described using those concepts will be classified as an instance of the ex:Animal\_Map class, leading to new :d\_about relations with animal\_owner, etc.

Once we obtain a description expressed using legal concepts, the lexical manifestations of these concepts are used as a query on an indexed corpus of case law. The lexical manifestation of a concept is a combination of weighted phrases, a vector or fingerprint. The fingerprint includes indicative phrases for the concept in our corpus. For instance, a query for “strict liability for animals” will include the phrase “the animal’s own energy”, which indicates that the volatility of the animal is discussed in a document (See Figure 1). This distinguishes our approach from more traditional concept-based search (c.f. [1]) where search is performed on the basis of the word senses of synonyms of the concept label.

This translation approach solves two problems. First, the reducibility problem between layman and legal concepts, and secondly, the problem of fulltext search where a single concept can occur in the corpus using a wide range of lexical manifestations. In the following section, we introduce an extension to the contextual mappings of BestMap based on case frames [3]. Section 4 describes our experiences in implementing this approach in BestPortal.

\[5\]
is its transitive counterpart and is furthermore transitive over the skos:broader relation. See [4] and the OWL definitions of the ontology for more details.

\[6\]In the case of BestPortal we use OWLIM, see http://www.ontotext.com/owlim/.

\[6\]This example uses a variant of the standard Manchester Syntax, see http://www.w3.org/TR/owl2-manchester-syntax/. In this example, cv is the prefix for the commonsense vocabulary, an lv is the prefix for the legal vocabulary.

\[7\]The equivalence relation indicates a necessary and sufficient condition for class membership, where the subclass relation is only a necessary condition. By using these relations in this way, the definition of a mapping operates as a rule with the commonsense description as premise, and the legal description as conclusion.
3. Case Frames as Contextual Mappings

We discussed that the co-occurrence of commonsense concepts in a case description determines the applicability of legal concepts. However, the ex:Animal_Map mapping does not take into account what roles these concepts play in the description. To see what this leads to, consider a case where a fourteen year old child is bitten by a dog. The above mapping will not be able to distinguish that case with the converse case where the dog was bitten by the child. Where the former is a case of strict liability that holds the owner of the dog liable for any damage caused, the latter is a potential case of fault concerning the parents of the child.

To what extent would making these roles explicit interfere with our intention to take a lightweight approach? Including concept roles increases expressiveness, but it does not necessarily increase the effort required in specifying mappings. Why is this the case? The vocabularies used in specifying the mappings are taxonomies with very distinct branches that correspond to the important aspects of a case. For instance, the commonsense vocabulary that was adopted from [9] contains branches for e.g. actions, actors, consequences, means and so on. The taxonomic nature of vocabularies is the source of their power. Not taking this structure into account in the definition of mappings leads to information loss that can easily be avoided. Making the roles explicit will help the modeler in determining what aspects of the case, i.e. what branches of the taxonomy, are of potential importance.

Case Frames Explicit roles have another advantage as a means to enable structured case descriptions that guide users of the portal to provide enough information about their case. If left unguided, a prototypical user will not know what and how much information is necessary for an adequate interpretation of her case. This insight led Winkels and de Bruijn [13] to develop a set of pre-structured templates for case descriptions in labour law, case frames. These frames were used to guide volunteers of a labour union in selecting the proper legal interpretation for frequently occurring commonsense terms.

The domain of unlawful acts is too diverse to rely on the presence of a small number of important commonsense concepts. We therefore looked for a more general pattern that could be applied in the description of any case in the domain of unlawful acts. Incidentally, the term ‘case frame’ is also term used in linguistics as a set of named slots for a verb [3] – the case frames of [13] bear more resemblance to the frames in frame-based knowledge representation languages, such as used in [11].

The unlawfulness of an action is determined by identifying the individual participants in the action, the situation in which the case takes place, and its consequences (typically some damage). These thematic relations describe the semantic relation of a filler with the verb. Typical thematic relations are actor, object and patient. We adopted the linguistic notion of case frame as a template for case descriptions by laymen. Thematic relations hold between a description and concepts in the vocabulary, rather than between an action and its participants directly. We selected a minimal subset of thematic relations required for these descriptions:

**Action** Relates a description to the action that took place (e.g. ‘hitting’, ‘threatening’).

**Actor** Relates a description to the agent that performed the action (e.g. ‘child’, ‘dog’).

---

8See e.g. http://en.wikipedia.org/wiki/Thematic_relation for an overview of important thematic relations.
Object Relates a description to the thing on which the action was performed (e.g. ‘house’, ‘agreement’). 9
Location Relates a description to the location where it took place (e.g. ‘farmyard’, ‘factory’).
Result Relates a description to the result of the action (e.g. ‘financial damage’).
Situation Relates a description to a special situation in which the action took place (e.g. ‘under duress’, ‘by accident’).

These relations were integrated with the BestMap ontology by specifying them as sub properties of the :d_about property. We follow the same procedure for defining a limited set of properties used in the legal description (based on [8]):

Ground for Unlawfulness Relates the description to the grounds on the basis of which the act is deemed unlawful, e.g. ‘violation of right’. Applicable to almost all cases.
Duty of Care Relates the description to a possible violation of duty of care, e.g. knowledge of a concrete danger.
Ground for Justification Relates the description to possible excuse or grounds for justification, e.g. the right of self-defence. Applicable only in cases of guilt.
Ground for Reduced Compensation Relates the description to a possible reduction of compensation, e.g. in case of an accepted risk. Applicable only in rare cases.
Ground for Exclusion of Guilt Relates the description to a possible exclusion of guilt, e.g. when the actor is a child under 14 years of age.
Ground for Attribution Relates the description to the grounds on which the act is attributed to the liable person, e.g. on the grounds of guilt, on grounds of a specific law. 10
Criterion for Strict Liability Relates the description to criteria that may hold in the case of strict liability, e.g. the volatility of animal behaviour.
Liberating Circumstance Relates the description to certain circumstances that may lift strict liability, e.g. when the damage was caused as a consequence of civil war.
Liable Person The person held liable on the basis of the description.

As such, these legal thematic relations do not improve search results – the query to the corpus is a simple weighted boolean query over all fingerprints of legal concepts applicable to the case. However, they can help in explaining what the most important aspects of a case are from a legal perspective.

Example We can now provide a more precise definition of the ex:Animal_Map mapping:

\[\text{ex:Animal_Map} \equiv \text{:actor value cv:animal} \land \text{:result value cv:damage} \land \text{:action value cv:dangerous_action} \land \text{:recipient value cv:person} \]
\[\land \text{:liable_person value lv:animal_owner} \land \text{:ground_for_attribution value lv:strict_liability_for_animals} \land \text{:criterion_for_strict_liability value lv:volatility_of_animal_behaviour} \land \text{:criterion_for_strict_liability value lv:damage_caused_by_animal}\]

This mapping will not match a case description where the child was the one who bit the dog. By adopting case frames and thematic relations into the framework of BestMap, we now have a powerful way to create more distinctive descriptions of cases without a

---

9 Strictly speaking this relation is a combination of the thematic relations patient and theme.
10 This relation is currently also used to express a liability attribution on grounds of strict liability, although this is not a proper use of the technical term ‘attribution’.
significant increase in the complexity of modelling: we only make use of distinctions already present in the vocabularies. Users can formulate their queries in a fashion that will improve their chances of intelligible results. It also increases precision for adequately formulated queries by diminishing the chances that distinct types of unlawful act have equivalent definitions.

4. Implementation

This section describes our findings in implementing the approach described in the previous section. We discuss several significant adaptations to the vocabularies and fingerprints used in BestPortal.

**Fingerprints**  The introduction of context aware mappings [4,5] required us to reconsider the way in which vocabularies and fingerprints interact in BestPortal. Until [9], the assumption was that a commonsense description would translate to a single concept: every query to the corpus was based on the fingerprint of one concept. As is evident in the example of the previous sections, context aware mappings are designed to result in descriptions using multiple concepts from the legal vocabulary. Fingerprints of several legal concepts are combined in a query to the corpus, which can have undesirable effects.

Since mappings are OWL classes, they can be positioned in a subsumption hierarchy. A mapping that characterises some case as describable by lv:strict_liability can subsume cases of more specific forms of strict liability, e.g. lv:strict_liability_for_animals. This is a very powerful mechanism that we use to modularise and structure the mappings. A case that is described by a specific form of strict liability is automatically also described as a case of strict liability. The portal will combine the fingerprints of both concepts in its query to the corpus. It therefore requires fingerprints to be more clear-cut lexical representations of the concepts.

Fingerprints for generic concepts such as lv:strict_liability were designed to be used as stand alone queries that are able to retrieve any and all subsumed cases. Fingerprints for these concepts therefore encode hierarchical information, and include search phrases for the more specific concepts. As a result, a query that combines fingerprints of generic and specific concepts becomes a query for all other concepts subsumed by the generic one. This is clearly undesirable. Combined queries lead to more possibilities for overlapping fingerprints. As argued in the previous section, every legal case description has a number of fixed ingredients, which in combination create a unique characterisation. The stand-alone fingerprints tended to accommodate this contextual co-occurrence and oftentimes contained search phrases of contextually related concepts. This results in a query bias towards search phrases occurring in the overlap of fingerprints in the query. We checked all fingerprints for potential undesirable overlap with other concepts. Hierarchical information was pruned away from the fingerprints of non-leaf concepts, moving search phrases to lower level concepts where necessary.

**Vocabulary Structure**  We had another careful look at the legal and commonsense vocabularies in light of the extension to case frames. The organisation along thematic roles was latent in the commonsense vocabulary; even though most roles were present as concepts, the vocabulary still contained various branches that were not unified under those categories. It also contained a fair number of lexical word-like concepts such as ‘he’ and
'she' that were left from an earlier knowledge elicitation exercise that assumed layman descriptions would be input as natural language texts. Lastly, we enriched the vocabulary with a number of additional skos:broader relations that allow categories to play multiple roles, e.g., animals can be both actor and object in an action.

The legal vocabulary was a large collection of concepts related to unlawful acts, and required more attention. The case frames depend on structured descriptions of legal cases, but the vocabulary was not designed with this structure in mind. Secondly, certain branches of the vocabulary were underdeveloped and had little or no hierarchical structure. Lastly, the vocabulary was the result of a translation from an OWL ontology to a SKOS vocabulary and did not take advantage of the lightweight semantics – and thus flexibility – of the skos:broader relation as compared to owl:subClassOf. Letting go of the strict semantics of class subsumption allows for a concise hierarchical structure that is more aligned to information retrieval.

On the basis of these insights, we developed a new legal vocabulary that is more aligned with the thematic roles of our legal case frame. The vocabulary is based on an introductory text [8] that gives a general overview of the domain of unlawful act. The number of top-level categories in the vocabulary was reduced from ten to four (liable person, liability, unlawfulness and attribution). Each lower level in the hierarchy is a palette of concepts that are to be considered before the higher level can be said to be applicable. For instance, to establish whether liability holds, it first needs to be determined what type of liability holds (guilt, fault or strict liability). Since all fingerprints were attached to the original legal vocabulary, the two vocabularies needed to be aligned. We used a simple two-way lexical alignment to determine concept similarity and bootstrap the alignment procedure. Each combination of concept labels that is the same, or has a suitable weighted Levenshtein distance, is added to an intersection of the concept labels. We then apply a Jaccard index to determine whether the two concepts are likely to be the same ($J > 0.25$): $J = \frac{|W_{\text{old}} \cap W_{\text{new}}|}{|W_{\text{old}} \cup W_{\text{new}}|}$

Running this comparison resulted in 440 out of 657 possible concept alignments (a coverage of 67%). After manual tuning, the final merged vocabulary contains 375 mappings, between 117 (out of 180) new concepts and 174 (out of 477) concepts of the old vocabulary. Although we did not perform a proper evaluation for these proposed mappings, they formed a significant bootstrap to the alignment process (given a total of over four thousand fingerprints). For each new concept without a mapping, initial fingerprints were created based on their label. The pruning described in the previous section is applied after this step.

5. BestPortal Revisited

The BestPortal was reimplemented using client-side JavaScript technology to make the user interface more responsive and allow more flexible interaction (see Figure 2). The layman case frame is presented to the left; for each thematic role in the frame, a combo box allows the selection of a concept from an indented list. The selected concepts are then positioned in the ‘Beschrijving gebeurtenis’ bar, the role of a concept is indicated between brackets. Once a selection is made, the system checks whether any mappings

---

Figure 2. The BestPortal interface

apply (by querying our Sesame/SwiftOWLIM repository),\textsuperscript{12} and displays the applicable legal concepts below the layman case description. Hovering over a question mark will give additional information about the concept. If one or more legal concepts apply, the system will use their fingerprints to produce a weighted boolean query which is sent to the Solr Lucene index, and displays a synopsis of the results in the center frame of the page. Attributes, such as the type of court, the place of court, and the legal area are clickable links that operate as filters on the search results. For instance, selecting the “Rechtbank Groningen” will filter the results for all verdicts by the court in Groningen.

Every result is annotated with RDFa attributes and can be scraped for RDF metadata.\textsuperscript{13} It also contains explicit deeplinks to an RDF/XML representation of the metadata (the blue RDF icon), a linked data browser (Marbles)\textsuperscript{14}, and a deeplink to the LJN index at the Dutch Council of the Judiciary.\textsuperscript{15} The timeline in the bottom right is a Simile widget, that displays all results according to publication date. Every Dutch court present in our court proceedings was linked to the DBpedia and Geonames linked data repositories, allowing us to present our results on a Google map (the dark red icons). We also scanned the texts of our corpus for Dutch place names taken from Geonames, and created links wherever a match was found. This allows us to position the places mentioned in court proceedings on the Google map. The entire repository is exposed via a browse-able

\textsuperscript{12}OpenRDF Sesame is a triplestore, see http://www.openrdf.org. SwiftOWLIM is a lightweight OWL reasoner that can operate on top of Sesame, see http://www.ontotext.com/owlim/.

\textsuperscript{13}RDFa is a way to transparently incorporate RDF descriptions within HTML, see http://www.w3.org/TR/rdfa-syntax/.

\textsuperscript{14}See http://www5.wiwiss.fu-berlin.de/marbles/

\textsuperscript{15}The LJN is the official identifier of every court proceeding.
6. Evaluation and Discussion

A first small evaluation shows that using case frames has a significant effect on retrieval quality over the simple contextual mappings. Table 1 compares results for our example queries, “Dog bites child, resulting in damage” and “Child bites dog, resulting in damage”, achieved via a naive full text query on Rechtspraak.nl, and a query on the same corpus via simple mappings and via case frames.\(^\text{18}\) Results were capped at a maximum of 20. As was expected, Rechtspraak.nl could only find a limited number of applicable cases (a total of four), and could not distinguish between the two queries. The one correct hit concerned a case where a woman was bitten by a dog. Using simple mappings improves matters somewhat – the total number of results rose to over 140 thousand – but still produces the same results for both queries, and generates high number of false positives. Case frames have a significant positive effect on the first query: false positives dropped to 30%, with no cases of animal molest. The top 6 results all concerned strict liability for animals. However, case frames do not perform as well for the second query, where most results concerned child custody cases\(^\text{19}\) and none of the results were about animals.

Although we are still in the process of improving and pruning our fingerprints (section 4), and the quality of search results can be much improved, even this very limited evaluation shows the potential for case frame based conceptual search as compared to the simple mappings. The difference with a traditional naive full text search on the same corpus is even more striking and shows that the current Rechtspraak.nl requires users to make an effort in coming up with the right formulation.

We are currently in the process of preparing a larger evaluation using two groups of students: law students and computer science students. They will use either Rechtspraak.nl or the BestPortal to solve a small number of cases. The performance of the systems will be evaluated on the basis of the students’ ability to correctly predict the judgement of a legal expert for each case.

This paper introduces case frames as a better way of structuring mappings between two distinct and incommensurable vocabularies that concern the same domain. Although the approach is deemed applicable to many domains, we have successfully applied case frame based mappings in a revamped version of BestPortal. We discussed how case

---

\(^{16}\) All URIs are dereferencable via http://linkeddata.few.vu.nl/rechtspraak

\(^{17}\) See http://ckan.net/package/rechtspraak.

\(^{18}\) Our full text search string was “dog bites child damage”. The evaluation in table 1 is at a general level: the three positive results on “Child bites dog” all refer to cases about responsibility for the behaviour of children.

\(^{19}\) These form the bulk of court proceedings that concern children.
frames improve retrieval results without requiring significant additional modelling effort. We indicated our expectation that it will be easier for users to formulate answerable queries if they use the structure provided by these frames, and shown some initial – but promising – evaluation results.

References