Case Law Retrieval by Concept Search and Visualization

Elisabeth M. Uijttenbroek[1] Michel C.A. Klein[2] Arno R. Lodder[1] Frank van Harmelen[2]
e.m.uijttenbroek@rechten.vu.nl michel.klein@cs.vu.nl lodder@cedire.org frank.van.harmelen@cs.vu.nl
[1]Faculty of Law / [2]Department of Artificial Intelligence, VU University Amsterdam, The Netherlands

1. INTRODUCTION

The BEST-project (BATNA Establishment using Semantic web Technology, `http://best-project.nl`) strives to provide disputing parties with information about their legal position in a liability case. Our assumption is that through intelligent disclosure of Dutch Tort Law cases, laymen can estimate their chances: information derived from previous court decisions can help to obtain insight into BATNAs (Best Alternative to a Negotiated Agreement), alternatives a party has if negotiation fails. Information BATNAs also contributes to determining the room left for negotiation.

Initial experiments in 2005 aimed to match layman terminology with legal ontologies [4]. Early 2007 we started to analyze layman input, and try to identify relevant parts in case law that in a next step can be linked to elements in the description by the layman. In 2006 we concentrated upon good retrieval of case law. At first we defined a search document for each relevant code section. Although retrieval results were satisfactory [2], we considered this approach too static for relating retrieval results to laymen input. We therefore decided to define search documents on the level of the code section concepts. This approach allowed us to use our thesaurus-based statistical retrieval techniques together with a visualization technique successfully used in [7]. An additional advantage is that the search documents can be used for concepts of several code sections and an extension of the domain can build upon concepts already defined.

Below we present a technique for the retrieval of case law in the domain of tort law and address the following research questions:
- Is the concept-based retrieval method together with our method for building search documents suitable for retrieving court decisions?
- Can we use the visualization method to query for code sections by combining the query results for individual legal concepts?

2. CONCEPT-BASED RETRIEVAL

For the retrieval of the relevant documents, we use a thesaurusbased statistical indexing method. This technique has been implemented in a commercially available software tool by Collexis. The main advantage of this technique is that, when compared to standard information retrieval techniques based on the vector space model, the indexing is guided by a thesaurus; in this way, only terms relevant in a specific domain are taken into account.

A list of the relevant concepts identified in a document is called a concept fingerprint or search document. For each identified concept a unique concept identifier is added to the fingerprint. This concept identifier is assigned a relevance score, based on term frequency and the specificity of the term in the thesaurus, and the lexical similarity of the term with the textual contents [1].

2.1 Determining essential concepts

For 15 Dutch tort law code sections, we determined essential legal concepts. Tort law doctrine, the relevance of the concept in light of case law and context have been used to determine how to split up a section into legal concepts. The following criteria have been used to divide a section into legal concepts.

A The legal concept should have a certain level of broadness to make it applicable to a large category of case law;
B The legal concept should be precise enough to be relevant in a particular factual context;
C Tort law doctrine is the leading guideline;
D Coherence.

2.2 Formulating search documents

The relevant sections of the Civil Code are split up into essential concepts. For each concept a search document is created. The search documents vary depending on the nature of the concept. We distinguish open textured concepts and clear concepts.

2.2.1 Open textured and clear concepts

Legal reasoning is indeterminate due to its open, procedural nature [5]. Open texture is the main reason to treat the legal domain as a specific domain of retrieval. An alternative method of retrieval for open textured concepts is proposed. Following Stranieri & Zeleznikow [6] six criteria are used to determine the open textured character of a concept: *Ambiguity, Granularity, Discretionary statutes, Jurisprudence, Socio-political environment, and Completeness of knowledge.*

Clear concepts do not have to be interpreted. An example of a clear concept is an act violating a statutory duty. In case of clear concepts the code text alone suffices to create good search documents. All the possible violations can be found in the Dutch code. In case law the reason for unlawfulness of the act, such as acting in conflict with the obligation to identify, or the relevant section can be mentioned.

ICAIL '07 June 4-8, Palo Alto, CA USA

2.2.2 Creation of search documents

The content of the search documents is determined by the following criteria.
A. *Either code- based or case-based*: Code text or terminology in court decisions is used to create the search document.
B. *Certain level of abstraction*: The content of the search document should be abstract enough to retrieve as much relevant court decision as possible.

2.3 Results

The results showed a relevance of approximately 70relevance score for concepts created for sections of the code that are not applied regularly were below average, while the relevance score for concepts of often applied sections were above average.

3. VISUALIZING OVERLAP BETWEEN CONCEPTS

Because each code section is split into several concepts and hence search documents, an intuitive assumption is the following:
The relevance of a retrieved case for a specific code section increases with the number of concepts of that code section for which this case is relevant.
Therefore, the intersection between the sets of retrieved cases for concepts of the same code section are probably the most relevant cases. For the visualization of the clustering of cases we use the clustermap viewer from Aduna [2]. This software creates Vennlike diagrams of objects and show if they belong to one or more sets. Each object, in our implementation a court decision, is represented as a sphere. All retrieved cases that contain a specific concept are clustered and visualized as amoeba-like shapes (blob shapes). If an object belongs to multiple clusters (which means that a court decision is relevant for more than one concept), the blob shapes overlap and the object is displayed in the overlap. In our implementation, we created direct links for the objects (spheres) to the online version of the verdicts. An interface has been implemented that connects the Collexis search software to the Aduna clustermap viewer. This interface allows formulating queries for sets of concepts. We use this interface to specify sets of legal concepts that together represent a section of the code.

3.1 Visualization experiments

We did some experiments with different combinations of the concepts for which we defined search documents. We chose the sets of concepts in such a way that we were able to visualize overlap between the cases for concepts that together establish a certain kind of liability. A retrieved court decision is determined relevant if it discusses the type of liability for which we created the *combination of concepts.*

3.2 Results

The results showed relatively high scores on precision for the sections that are often applied to establish liability and Poorer results for sections that are not often applied, such as the liability for representatives. The results showed an average of 60% relevance. The results for often applied sections show results up to 100%, while sections of the code that are rarely applied resulted in a relevance score of less than 40%. We hypothesize that the poor results for the clusters of concepts that resembled less applied sections of the code is possibly also due to the fact that the used database from www.rechtspraak.nl exists since 1999 and that few court decisions about certain types of liability are available.

To validate the recall, we used standard court decisions that contain the basic interpretation and argumentation for a liability section of the code. All these court decisions were from before the launching of `rechtspraak.nl` (1999), and therefore added to our database. We hypothesized that these basic court decisions would be displayed by the clustermap viewer. Poor results for the recall were obtained. The court decisions relevant for a specific category of liability were not displayed by the clustermap viewer. These poor results on the recall could possibly be explained by the use of different terminology in older court decisions that is not used in the fingerprints. Another possible explanation is the limited manually composed thesaurus and or the limited use of terminology for the manually created fingerprints.

The clustering results show that some concepts can be omitted. The redundancy of concepts could be explained through the neutral character of the terminology related to the concept of damage.

4. CONCLUDING REMARKS

We described our approach to case law retrieval that consists of two steps. First, code sections are split up into legal concepts and for these concepts search documents are created based on the "character" of the concept (open textured or clear). Second, relevant case law is selected based on a visualization of the overlap of the search results for the individual concepts.

5. ACKNOWLEDGEMENTS

This research is funded by the Netherlands Organisation for Scientific Research NWO under grant number 634.000.436.

6. REFERENCES

[1] C. van der Eijk, E.M. van Mulligen, J.A. Kors, B. Mons & J. van den Berg. Constructing an associative concept space for literature-based discovery. *J. Am. Soc. Inf. Sci. Technol.*, 55(5):436444, 2004.

[2] C. Fluit, F. van Harmelen, M. Sabou, Ontology-based Information Visualization: Towards Semantic Web Applications. In *Visualising the Semantic Web* (2nd edition), 2005, Springer Verlag.

[3] M.C.A. Klein, W. van Steenbergen, E.M. Uijttenbroek, A.R. Lodder & F. van Harmelen. Thesaurus-based Retrieval of Case Law. In: *Proceedings of the 19th International JURIX conference*, IOS Press, 2006

[4] R. van Laarschot, W. van Steenbergen, H. Stuckenschmidt, A.R. Lodder & F. van Harmelen. 'The Legal Concepts and the Layman's Terms', *Proceedings of the 18th International JURIX conference*, IOS Press, 2005.

[5] A.R. Lodder, 'Law, Logic, Rhetoric: a Procedural Model of Legal Argumentation' (chapter 26), in: S. Rahman & J. Symons (eds.)., *Logic, Epistemology, and the Unity of Science*, Kluwer Academic Publishers 2004.

[6] A. Stranieri and J. Zeleznikow. Knowledge Discovery from Legal Databases, Law and Philosophy Library, volume 69, Springer, 2005

[7] Stuckenschmidt, H., van Harmelen, F., de Waard, A., Scerri, T., Bhogal, R., van Buel, J., Crowlesmith, I., Fluit, Ch., Kampman, A., Broekstra, J., van Mulligen, E. Exploring Large Document Repositories with RDF Technology: The DOPE Project, *IEEE Intelligent Expert*, Vol. 19, No. 3, pp. 34-40.