Legal Knowledge Management

Marco Costa[†], Orlando Sousa[‡] & José Neves[†]

[†]Universidade do Minho, Departamento de Informática
Largo do Paço, 4719 Braga Codex, Portugal
marco@di-ia.uminho.pt, jneves@di.uminho.pt

[‡]Polytechnic Institute of Porto (IPP), Institute of Engineering, Department of Computer Engineering
Rua de S. Tomé, 4200 Porto, Portugal
orlando@dei.isep.ipp.pt

Abstract

Knowledge in the legal domain assumes two distinct forms: case law and legislation. Case law complements legislation, and refers to the use of cases decided in court to provide interpretations to subjective aspects of the legislation. These cases are used as guidance in future similar cases. This paper presents a Case-Based Reasoning (CBR) system that establishes legal case-bases, allowing the users to find and retrieve information on cases similar to current ones.

Keywords: The Law, Case-Based Reasoning, Knowledge Management.

Introduction

One of the important aspects of the Law is its *open texture* nature (Hart 1961) that sometimes leads to conflicting interpretation of legal norms. The common way of resolving these conflicts is to invoke past interpretations, especially if stated by hierarchically superior courts; given their privileged position, it is presumed that the jurisprudence established by them in their decisions is adopted in future similar cases. By studying rulings from superior courts concerning cases similar cases to their own, legal agents can gather elements to guide their actions; judges can make fundamented rulings, and lawyers can seek arguments that favor their cases.

Under a classical setting, if someone needs information about a particular case similar to their own, he/she needs to search for the information manually on whatever textbooks or jurisprudence archives are available. Those elements may not be at hand, may be incomplete, or may not contain all the information necessary. A manual search is a time consuming process, and relevant cases can be overlooked.

This paper presents a Case-Based Reasoning (CBR) (Aamodt & Plaza 1994) system that establishes legal case-bases. These legal case-bases are an effective form of knowledge disseminating and reuse in the legal domain, helping to overcome the mentioned difficulties in accessing legal jurisprudence.

The Case-base

Several aspects have to be considered in creating jurisprudencial case-bases:

- A selection of which cases to store must be done, following a criterion of usefulness and validity (i.e., cases must advance some knowledge, and that knowledge must have not been invalidated by posterior events).
- The format of the cases must be defined. Should one use flat files, attribute values, or some other form of encoding the information contained in cases?
- The vocabulary to be used must be defined (what attributes to define, or what terms to use). Similarity measures must also be defined.

The selection of cases

The selection of which cases to incorporate in the case-base is an important one. Different sources of cases must be considered, and from those, a collection of cases must be selected that contains all the knowledge considered important.

The case-base must be kept up-to-date, eliminating all the cases whose knowledge has been invalidated by legislative measures or by rulings from a superior court. To this end, a committee of judges has been formed, responsible for the evaluation and selection of cases contained in known sources of jurisprudence, and for maintaining the case-base up-to-date by inserting new cases and deleting invalid ones. The quality of the case-base is directly related to the quality of the cases it contains, so the work of this committee is very important for the success of this project.

Case-Retrieval Nets

The selection of the CBR technology to use was done on the basis of some characteristics considered important in a good CBR system, namely:

- The CBR system must support efficient case retrieval, since users are unwilling to use a system that it too slow.
- The retrieval process must be complete, i.e., all relevant cases must retrieved. Users would loose trust in the system if they detected that it was overlooking important cases.
- The retrieval process must be flexible, allowing the users to conduct a search with whatever information they possess at the moment. Forcing the users to always provide the same information to seed a search would be very restrictive, since different pieces of information are known at different times.
- The construction and maintenance of the case-case must be straightforward. The insertion and removal of cases must be a simple task. It should be an incremental process, with no need to rebuild the case-base each time a case is added or removed. This way, the case-base can be regularly updated by people not specialized in computer science (i.e., the judges themselves). Allowing the maintenance of the case-base to be done by domain experts, and not by computer science specialists is an important aspect for the acceptance of and viability of the system.

Case-Retrieval Nets (CRNs) (Lenz & Burkhard 1996) fulfill all of the previous requisites, so they were selected as the CBR technology to use.

The most fundamental item in the context of CRNs are the *Information Entities* (IEs). These may represent any basic knowledge item, such as a particular attribute-value-pair. A case consists of a set of IEs, and the case-base is a net with nodes for the IEs relevant to the domain and additional nodes denoting the particular cases. IE nodes may be connected by *similarity arcs*, and a case node is reachable from its constituent nodes via *relevance arcs*. A case retrieval is performed by activating the IEs given in the query case, propagating the activation according to similarity, and collecting the achieved activation to the associated case nodes. The output of the system is a list of cases ordered and classified by a relevance measure (the higher the relevance measure of a case, the more similar the case is to the query case).

Some of the important advantages of CRNs are:

- They can handle only partially specified queries.
- Any part of a case may be given; the retrieval algorithm will deliver the remaining part, thus completing the case (i.e., a case is not necessarily seen as being composed of a problem and a solution).
- The case-base can be tuned at run-time to express different similarities between IEs, or different relevance measures between IEs and case nodes.
- The attributes that define each case are flexible, since any set of IEs may be connected to a case node.
- The insertion/removal of cases, and even the addition of new IEs can be performed incrementally by adding the respective nodes and arcs.

The knowledge contained in the case-base is not limited to the cases themselves. In fact, three distinct forms of knowledge may be identified: the cases (defined by the relevance arcs between IEs and case nodes), the vocabulary or concepts used to define cases (IEs), and the similarity measure between concepts (similarity arcs between IEs). The management of these three forms of knowledge is the responsibility of the committee, but both the similarity and relevance arcs may be tuned at run-time to better suit the necessities of particular users. This adds more power to the system, making the searches for similar cases much more flexible, thus increasing the chances of user satisfaction.

The need for an Ontology

The three forms of knowledge previously mentioned must somehow be defined or collected. The collection and selection of legal cases is considered a straightforward task, but the definition of the terms to use and the similarity between different terms requires more attention. The creation of an ontology for the domain is the right way to perform the last two tasks, i.e., the ontology provides the support from which the available IEs and similarity measures are defined.

An ontology is an explicit specification of a conceptualization, i.e., the description of the concepts and relationships that define a domain (Gruber 1993). The main purpose of an ontology is to enable knowledge sharing and reuse; by committing to an ontology, an agreement is made to use a defined vocabulary. This avoids knowledge sharing problems that may arise when different agents use different vocabularies. More specifically, everyone using the system knows the IEs that compose the CRN, and therefore, knows how to express himself (define a query case), and how to interpret the knowledge provided by the system (understand the terms used to describe cases).

For the moment, our system operates only in some specific areas of the legal domain, for which a simple conceptualization was easily achieved. Our ontology limits itself to provide support for the definition of terms and similarity measures between related terms (e.g., for the crimes, hierarchical categories were defined in successive levels of specialization - Fig. 1).

More complete ontologies exist for the legal domain (e.g., Visser & Bench-Capon 1997).

Maintenance of the Case-Base

The maintenance of the case-base includes the tasks of inserting new cases, removing old ones, and make the necessary adjustments to reflect changes to the ontology used to define the CRN (we expect the ontology to stabilize after some initial tuning to reflect user suggestions).

The cases to insert in the case-base are, for now, extracted from a journal edited by the Ministry of Justice. This journal contains the most important jurisprudence from the Supreme Court of Justice and from the 2.nd

instance courts.

The processing of a real-life case is done by analyzing the text of the case and, in conformance with the ontology, obtain a set of IEs that expresses the relevant information contained in the case. Information not relevant is discarded; e.g., the name of the offender is irrelevant, but its age is not, since it may be important to the reasoning associated with the case. This processing is done manually, but it may become necessary in future to automate the process by using Natural Language Processing (NLP) techniques (Weber-Lee et al. 1997; Lenz & Glintschert 1999).

Due to the changing nature of the Law, the cases contained in the case-base may become obsolete, i.e., they would have been conducted differently in the current legal context. The decision of keeping or removing those obsolete cases depends on the extend to which the changes invalidate the knowledge contained in the cases, and on the availability of cases containing the same knowledge (the part not invalidated).

A problem similar to the previous one is what to do with contradictory rulings (at different times and/or from different courts). In the case of ruling from different courts, a hierarchy is defined, so it is up to the user to select the courts to consider (by means of IEs and relevance measures). Contradictory rulings from the same court are also maintained; when the Supreme Court of Justice detects that it has pronounced contradictory ruling, it provides a special ruling that overrules the others; this special rulings may be given priority by the user (again, by means of IEs and relevance measures).

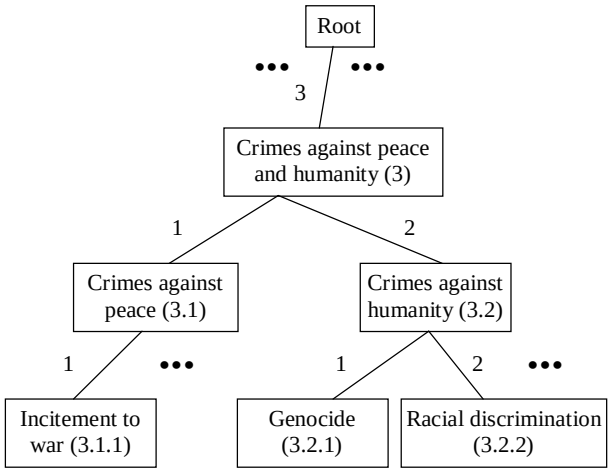

Figure 1 - Hierarchical definition of crimes

Use of the Case-Base

The system was designed to make the task of using easy, and to facilitate the understanding of its inner-workings, giving its users the ability to effectively make use of all of its capabilities.

To make a query, the user provides a set of initial IEs that expresses the facts that the cases must include. The IEs are selected from a palette that contains all the available IEs arranged by groups (e.g., facts about the crime, information about the people involved, final decision).

In order to improve the first results, the user may repeatedly add or remove IEs, adjust the similarity measures between IEs, and change the relevance of each IE.

As the final result of a query, the user obtains a list of cases ordered by relevance. For each case, it is possible to obtain its description in terms of the IEs that compose it, and also the full textual description from which the IEs were created.

As an example, a lawyer may search murder cases committed by young people, where the accused was found guilty, but had a light sentence. After an analysis of the initial results, more IEs are added, providing information about how the crime was committed. Also, the relevance of the IE that encodes the information that the offender was young is increased. This leads to results that better correspond to the expectations of the lawyer, trying to know how to better take advantage of the youth of his client to obtain a light sentence.

Related Work

HYPO (Ashely 1989) is considered the first precedent-based CBR system. It does adversarial reasoning with cases and hypotheticals in the legal domain.

A more recent system is SHYSTER (Popple 1996), capable of providing advice based upon an examination of, and an argument about, the similarities and differences between cases. Both our project and SHYSTER require cases to operate (in SHYSTER, real and ideal cases), and the definition of the attributes (in our case, IEs) that compose cases. But while our system is designed to be very dynamic, with an active maintenance of the case-base, SHYSTER operates using domain *specifications* produced by a legal expert. A specification defines a set of attributes and a small number of relevant cases; once defined, a specification is not meant to be changed.

Also, the retrieval of similar cases and the use given to them is very distinct in the two systems. SHYSTER uses statistics to find the “nearest” case, and then proceeds by formulating a report based on the similarities and differences between the cases. Our system only allows the selection and retrieval of cases, it does not try to produce a report; we consider more important and useful the ability to find relevant cases in a representative set of cases, than the ability to obtain a report of questionable quality.

A system very similar to our own is described in (Weber-Lee et al. 1997), a Brazilian case-base reasoner for legal cases. The main differences are in the use of CRNs in our system and on the focus given to the automatic conversion of legal texts into legal cases (motivated by the large number of cases).

Conclusion

In this paper we provided a sketch of a system capable of effectively distribute and give access to one of the most important forms of legal knowledge: legal precedents. It is not an ambitious system that tries to reason with that knowledge and produce as output advice or reports to the user. Instead, the idea is to create complete and easily maintainable case-bases. By being complete, i.e., containing all the cases considered relevant, the users will not be disappointed by finding that the system does not contain an important part of the domain knowledge.

The simplicity of the process makes it possible for the people from the legal domain to conduct the task managing the case-base. In this way, their responsibility for the quality of the system increases, and a separation is achieved between the work of defining and implementing the system, and the work of managing the knowledge it contains.

Acknowledgements

The first author received financial support from the Portuguese Ministry of Science and Technology, under the contract Praxis XXI/BD/13650/97.

References

Aamodt, A.; Plaza, E. 1994. Case-Based Reasoning: Foundational issues, Methodological Variations, and System Approaches, *AI Communications*, 7(i), 39-56.

Ashley, K. 1989. Toward a Computational Theory of Arguing with Precedents: Accommodating Multiple Interpretations of Cases. In Proceedings of the Second International Conference on Artificial Intelligence and Law, 93-102. Association for Computing Machinery, Vancouver.

Gruber, T. 1993. A translation approach to portable ontologies. *Knlowledge Aquisition*, 5(2), 199-299.

Hart, H. 1961. *The concepts of Law*. Clarendon Press, Oxford.

Lenz, M.; Glintschert, A. 1999. On Texts, Cases, and Concepts. In Proceedings of XPS-99, Springer Verlag, LNAI.

Lenz, M.; Burkhard, H.-D. 1996. Case-Retrieval Nets: Foundations, properties, implementation, and results, Technical Report, Humboldt University, Berlin.

Popple, J. 1996. *A Pragmatic Legal Expert System*. Applied Legal Philosophy Series. Dartmouth (Ashgate), Aldershot.

Visser, P; Bench-Capon, T. 1997. A Comparison of Two Legal Ontologies. In Proceedings of the First International Workshop on Legal Ontologies, 37-45. Melbourne, Australia.

Weber-Lee, R.; Barcia, R.; Costa, M.; Rodrigues Filho, I.; Hoeschl, H.; Bueno, T.; Martins, A.; Pacheco, R. 1997. A Large Case-Based Reasoner for Legal Cases. Lecture Notes in Artificial Intelligence: 2nd International Conference on CBR, ICCBR97. David Leake, Enric Plaza (ed.)-Berlin: Springer.