Managing Legal Precedents with Case Retrieval Nets

Marco Costa*, Orlando Sousa** and José Neves*

**Universidade do Minho, Departamento de Informática, Portugal*
***Instituto Superior de Engenharia do Porto, Departmento de Engenharia Informática, Portugal*
marco@di-ia.uminho.pt, orlando@dei.isep.ipp.pt, jneves@di.uminho.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 system that establishes legal case-bases, allowing the users to find and retrieve information on cases similar to current ones. The system uses Case Retrieval Nets to index the cases, and Information Extraction techniques together with an ontology to semi-automate the conversion of legal texts into cases.

1 Introduction

One of the important aspects of the Law is its *open texture* nature (Hart 1961) that sometimes leads to conflicting interpretations of legal norms. The most 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. Therefore, 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 favour 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 objective is to provide the user with the most useful cases to support his current legal problems.

Marco Costa, Orlando Sousa and José Neves, 'Managing Legal Precedents with Case Retrieval Nets', in: H.Jaap van den Herik et al. (eds), *Legal Knowledge Based Systems, JURIX 1999, The Twelfth Conference*, Nijmegen: GNI, 1999, p. 13-22.

2 The Case-Base

Several aspects have to be considered in creating jurisprudential 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.

Figure 1 presents the overall structure of the system, described in the following sections.

2.1 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 relevant.

The case-base must be kept up-to-date, dealing with 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. The quality of the case-base is directly related to that of the cases it contains (the work of this committee was paramount for the success of the project).

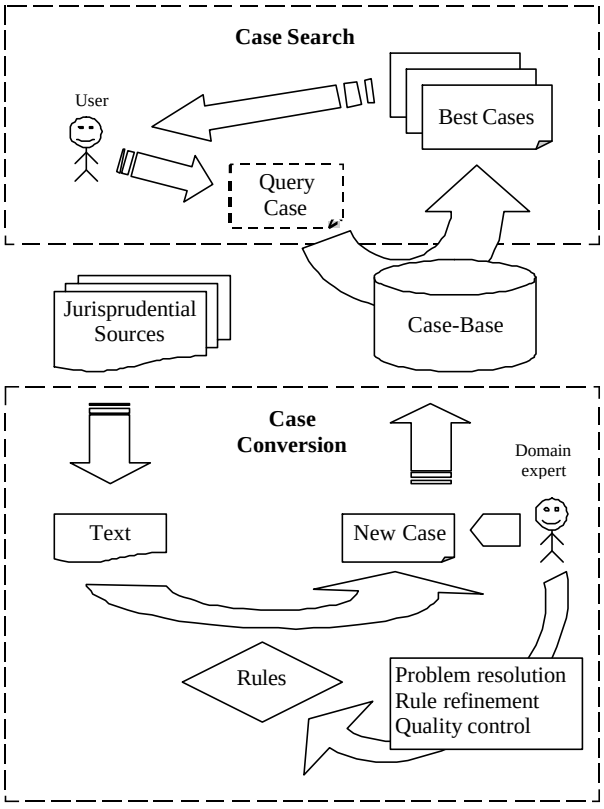

Figure 1 The overall structure of the system

2.2 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. Users are unwilling to use a system that is too slow.
- The retrieval process must be complete; i.e., all relevant cases must be 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-base 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 specialised in computer science (e.g. by 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 and viability of the system.

Case Retrieval Nets (CRNs) (Lenz & Burkhard 1996) fulfill all of the previous requisites.

The most fundamental items in the context of CRNs are the *Information Entities* (IEs). These may represent any item of basic knowledge, 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 measures, 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). Figure 2 presents an example of a CRN; the thickness of each line reflects the value assigned to the respective similarity or relevance arc. Some of the 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. One does not necessarily provide a fixed problem part to obtain a fixed solution part.
- The case-base can be tuned at run-time to express different similarity measures 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 of 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.

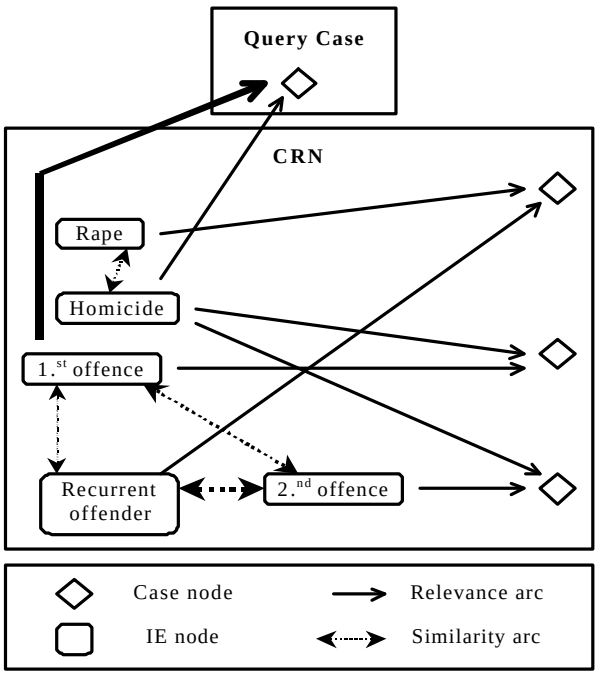

Figure 2 An example of a CRN

2.3 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 measures 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 conceptualisation; 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 full set of 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).

In its current state, the system is restricted to some specific areas of the legal domain, for which a simple conceptualisation was defined. This conceptualisation consists of the definition of IEs, similarity measures, and case scripts.

An IE encodes a specific piece of information, and has a specific data type (e.g. numeric, partially orderable, hierarchical). Figure 3 presents an example of a hierarchical attribute, the one that encodes the information about the crime committed. The dots in the figure represent branches of the tree not show.

Depending of the nature of the IE, the similarity measures may be implemented in different ways. One seeks to take advantage of the domain knowledge available (e.g. for the crime attribute, similarity is calculated using the knowledge about the position of each crime in the hierarchy). For a numerical attribute like age, there is no need to fully connect the IEs encoding all the possible values with similarity arcs. Instead, all the age IEs are connected to a reference IE, and the similarity measure between two age IEs is calculated by traversing the links between them (via the reference IE), and multiplying their difference by a user-defined constant (some expressive power is lost).

To guide the creation of cases, case scripts were created. A case script exists for each crime category, and defines the information considered relevant for cases related to those crimes. The information is divided in two groups: essential information and optional information. Case scripts are organised in a hierarchy manner, with each script inheriting the attributes of the parent script (the script hierarchy mirrors the crime hierarchy).

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

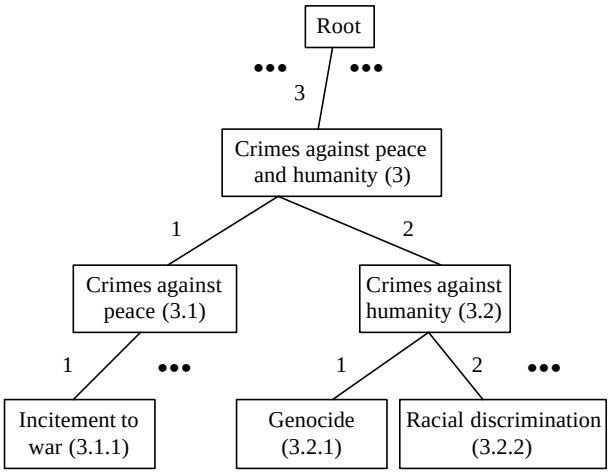

Figure 3 An hierarchical definition of crimes

2.4 From Texts to Cases

Cases describe legal texts, and are converted by the definition of IEs that describe and index the texts. The IEs that compose a case are selected in conformance with the ontology so that the IEs express all the information contained in the text. Information not relevant is discarded; e.g. the name of the offender is irrelevant, but its age may not, since it may be important to the reasoning associated with the case.

Ideally, a formal grammar would be applied to parse the texts with sophisticated techniques known from Natural Language Processing (NLP).

This approach, however, is not feasible because of the high volume of the data to process and because these NLP techniques are too complicated to be applied efficiently. Also, in domains like the Law, unknown terms may occur in any sentence. This is not compatible with NLP techniques, since they require a complete dictionary and would reject sentences in which just one unknown word is present. Due to these shortcomings, techniques developed within the Information Extraction community (Cunningham 1997) (Riloff & Lehnert 1996) are more practical. The basic idea is to perform a more shallow analysis of the structure of sentences than in NLP. This analysis is heavily pattern-driven in that typical phrases and stereotypes are used to identify the different roles of the various objects in a sentence.

In practice, the extraction of the IEs follows a two-stage process. The idea is to abstract from some of the less important components of the sentences, identify the main phrases, and recognise the existing structures between these. This approach is motivated by the observation that stereotypical expressions are very common. Very often this is due to the fact that a fairly small group of people is responsible for writing the documents and, hence, there is a natural tendency to use cut-and-paste techniques whenever possible.

The first stage of the conversion process is conducted automatically by using Natural Language Processing techniques (NLP). A set of rules was created after the domain experts (judges) have examined the texts with the objective of identifying the structure of the texts and the location of relevant pieces of information. The rules are defined using *concept nodes* (Lenz & Glintschert 1999). A concept node becomes active when a specific word is found. Once activated, a set of conditions is tested. These conditions define the required structure of a sentence. The matching process starts by splitting a given piece of text into sentences and adding part-of-speech information to each word (i.e., information that describes a lexical item in grammatical terms; e.g. singular common noun, comparative adjective, past participle). Sentences are then scanned for potential trigger words, and the conditions of the corresponding concept nodes are checked. When a matching concept node is detected, the relevant information present in that sentence is extracted and originates one or more IEs. Using case scripts, the system knows what IEs must be defined, and an error message is logged when an essential piece of information can not be extracted from the text.

At a second stage, a human expert examines the result of the automatic conversion. If some error occurred, a study is conducted to find the reason. If necessary, the rules are changed, with the possible addition of new ones, making it possible that similar situations be automatically dealt with successfully in the future. Gradually one obtains a tuned set of rules capable of handling most situations. The success rate depends mostly on how "normalised" are the documents. The more the legal texts are consistent in the terminology used and in their general format, the easier it is to obtain a set of rules capable of achieving a high success rate in the conversion process.

Even if no problems were signalled during the automatic conversion, the resulting cases may still be examined, trying to improve the conversion process by manually changing the result, and assuring a kind of quality control on the case-base.

2.5 Maintenance of the Case-Base

The maintenance of the case-base includes the tasks of inserting new cases, removing old ones, and making the necessary adjustments to reflect changes to the ontology used to define the CRN (it is expected that the ontology will stabilise after some initial tuning).

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 second instance courts.

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 extent 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 concerned with 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).

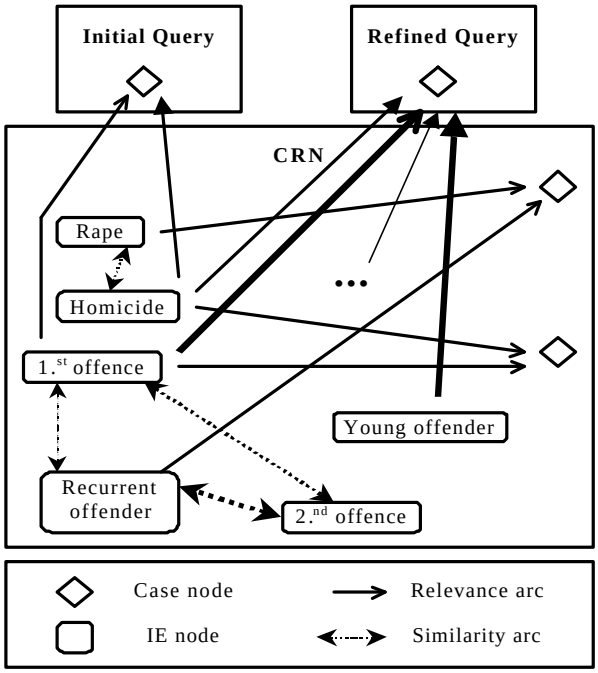

Figure 4 An example of a refined query case

2.6 Use of the Case-Base

The system was designed to make the task of using it 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 containing all the available IEs arranged in groups (e.g. facts about the crime, information about the people involved, final decision). As an example, the selection of the crime is done in the context of a tree presenting the crimes in a hierarchical manner that mirrors their definition in the penal code.

In order to improve the first results, the user may repeatedly add or remove IEs, adjust the similarity measures between IEs, or 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, lets consider a case of young offender that committed his first offence, a murder. His lawyer builds a query case as shown in Figure 4. After an analysis of the initial results, more IEs are added, providing more information and increasing the relevance of the IE that encodes the information that it was a first offence. This leads to results that better correspond to the expectations of the lawyer, who is trying to take advantage of specific particularities of the case at hand.

3 Related Work

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

The SALOMON project (Uyttendaele et al. 1996) tries to improve access to legal texts by automatically generating summaries of court decisions. Each summary contains the most relevant text units of the alleged offences and of the opinion of the court. SALOMON and one's system have similar objectives, differing mainly in the techniques used.

In (Brüninghaus & Ashley 1997), work is presented in automatically assigning factors to legal texts, thereby leading to a representation of cases in terms of factors. A discussion of different approaches for analysing text documents is also included.

SHYSTER is a system capable of providing advice based upon an examination of the similarities and differences between cases (Popple 1996). Both this project and SHYSTER require cases to operate (in SHYSTER, real and ideal cases), and the definition of the attributes (in one's 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. One's system only allows the selection and retrieval of cases, it does not try to produce a report; it was considered 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 one's own is described in (Weber-Lee et al. 1997), a Brazilian case-base reasoner for legal cases. The main difference is the use of CRNs and the development of an ontology in our system, which

allows the end users to take a more active part in the management of the system.

4 Conclusions and Future Work

In this paper it was 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 (the case-base); i.e., by 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 experiments conducted with the prototype built have shown that it is possible for the people from the legal domain to conduct the task of 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.

The availability of correct and up-to-date cases is crucial for the success of a CBR system. Therefore, in the system presented, the definition of the ontology and the semi-automatic extraction of cases from legal texts are two of the key aspects that in the medium and long term will dictate the failure or success of the project. The results so far are encouraging, but the work has still to be extended to other legal domains.

The CRNs allow a degree of flexibility far beyond common CBR systems. With CRNs, a user is never confronted with a situation in which he/she obtains an unsatisfactory response without being able to do something to improve the response of the system (in terms of the relevance and recall rates). In fact, with CRNs, the user has the power to significantly change the processing associated with the retrieval process by altering the input-case IEs, by adjusting their individual relevance, or ultimately, by changing the measures of similarity between the concept IEs. The high level of participation the user has in the retrieval process makes it necessary to evaluate CRNs differently from normal case retrieval systems. Instead of relevance and recall rates, one has to determine if the ontology is correct; i.e., if the concept IEs are correctly defined. Beyond CRNs, another part of the system that must be evaluated is the conversion of legal texts into cases (i.e., what IEs are derived from the texts). Here the concern is not with the final results of the conversion process, since the automatic part of the conversion process is complemented by human intervention. Instead, the question is how much and how correct is the work done in the automatic part of the processing of the texts.

In its present state, the system discussed in this paper does not involve much reasoning. As so, in terms of CBR, the focus is more in the retrieval of cases, and not in reasoning with them. This will be corrected in the future, which passes by the integration of the current system in a larger hybrid system. This larger system will use both case law and legislation as knowledge sources to help the user reason about a particular case. The CBR part of the system will be applied to alleviate the problems associated with the open-texture nature of legal norms mainly by assisting in decide about the degree of applicability of legal norms.

Acknowledgements

The first author was supported by a grant from the Portuguese Ministry of Science and Technology under the contract Praxis XXI/BD/13650/97.

References

Aamodt & Plaza 1994
A. Aamodt & E. Plaza, 'Case-Based Reasoning: Foundational issues, Methodological Variations, and System Approaches', in: *AI Communications*, 7(1), 1994, pp. 39-56.

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

Brüninghaus & Ashley 1997
S. Brüninghaus & K. Ashley, 'Finding Factors: Learning to Classify Case Opinions under Abstract Fact Categories', in: *Proceedings of the Sixth International Conference on Artificial Intelligence and Law*, Melboune: Association for Computing Machinery, 1997, pp. 123-131.

Cunningham 1997
H. Cunningham, *Information extraction – a user guide*, Research Memo CS-97-02, Sheffield: University of Sheffield, 1997.

Gruber 1993
T. Gruber, 'A translation approach to portable ontologies', in: *Knowledge Acquisition*, 5(2), 1993, pp. 199-299.

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

Lenz & Glintschert 1999
M. Lenz & A. Glintschert, 'On Texts, Cases, and Concepts', in: *Proceedings of XPS-99*, Berlin: Springer Verlag, LNAI, 1999.

Lenz & Burkhard 1996
M. Lenz and H.-D. Burkhard, *Case Retrieval Nets: Foundations, properties, implementation, and results*, Technical Report, Berlin: Humboldt University, 1996. URL:
http://www.informatik.huberlin.de/~lenz/Papers96/techRep.20March.ps.gz

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

Riloff & Lehnert 1994
E. Riloff & W. Lehnert, 'Information extraction as a basis for high-precision text classification', in: *ACM Transactions on Information Systems*, 12(3), 1994, pp. 296-333.

Uyttendaele et al. 1996
C. Uyttendaele, M.-F. Moens & J. Dumortier, 'SALOMON: Automatic Abstracting of Legal Cases for Effective Access to Court Decisions', in: *Proceedings of JURIX '96*, 1996.

Visser & Bench-Capon 1997
P. Visser & T. Bench-Capon, A Comparison of Two Legal Ontologies', in: *Proceedings of the First International Workshop on Legal Ontologies*, Melbourne, Australia, 1997, pp. 37-45.

Weber-Lee et al. 1997
R. Weber-Lee, R. Barcia, M. Costa, I. Rodrigues Filho, H. Hoeschl, T. Bueno, A. Martins & R. Pacheco, 'A Large Case-Based Reasoner for Legal Cases. Lecture Notes in David Leake, Enric Plaza (eds), *Artificial Intelligence: 2nd International Conference on CBR*, ICCBR97. Berlin: Springer, 1997.