Legal knowledge based systems
JURIX '95
Telecommunication and AI & Law

The Foundation for Legal Knowledge Systems
Editors:
J.C. Hage
T.J.M. Bench-Capon
M.J. Cohen
H.J. van den Herik

Paolo Guidotti and Fabrizio Turchi, Knowledge-based systems and distributed processing models, in: J.C. Hage, T.J.M. Bench-Capon, M.J. Cohen, H.J. van den Herik (eds.), *Legal knowledge based systems JURIX '95: Telecommunication and AI & Law*, Lelystad: Koninklijke Vermande, 1995, 43-52, ISBN 90 5458 252 9.

More information about the JURIX foundation and its activities can be obtained by contacting the JURIX secretariat:

Mr. C.N.J. de Vey Mestdagh
University of Groningen, Faculty of Law
Oude Kijk in 't Jatstraat 26
P.O. Box 716
9700 AS Groningen
Tel: +31 50 3635790/5433
Fax: +31 50 3635603
Email: sesam@rechten.rug.nl

KNOWLEDGE-BASED SYSTEMS AND DISTRIBUTED PROCESSING MODELS

Paolo Guidotti and Fabrizio Turchi
Istituto per la Documentazione Giuridica, IDG
via Panciatichi 56/16, 50127, Florence, Italy
email: {guidotti, turchi}@idg.fi.cnr.it

Abstract

This paper deals with problems encountered when adopting distributed-processing models in developing knowledge-based systems. Our reference is the experience matured at the IDG of Florence. We suggest that the adoption of these models does not introduce unsolvable difficulties, whereas the use of these enables a much broader user access. After a brief description of the client/server architecture, an interface for intermediary simulation in information retrieval is presented as an example of knowledge-based systems expressly designed to manage network connections. A concise discussion follows on the possibilities and problems arising from the adoption of network technologies in developing decision-support systems, also drawing attention to the possible developments that a wider use can entail. Finally, a short presentation is given to a project, currently in the phase of feasibility study, for the computerisation of the Court of Bologna.

1 Introduction

The past years have witnessed the break-through of network technologies. Combined with the Internet technology, they have assumed the dimensions of a mass phenomenon. At the basis of the most immediate and usable applications, such as WWW (World Wide Web), we find the utilisation of the distributed-processing models developed in the 1980s. Indeed, the development of LANs (Local Area Networks) has made it possible to substitute centralised processing with forms of distributed processing.

At the same time, approximately in the mid 1980s, the first legal expert systems made their appearance. These systems can be divided into two large categories: rule-based systems and case-based reasoning (CBR) systems. In legal expert systems using rule-based formalisms to represent statutes, the legal decision-making process is mechanically reproduced by an inference procedure (Sergot *et al.*, 1986). The CBR systems make it possible to establish relations with judicial decision precedents by means of a matching process (e.g., Rissland *et al.*, 1987).

In the meantime, neither of these systems has been developed for wide-scale use and instead both encounter difficulties in finding fields of application which enable a real industrial development. Indeed, few of the applications developed even leave the research laboratory or university. In these systems, only a limited segment of a normative system can be modelled by a consistent set of logic sentences representing a selected unambiguous interpretation. We feel that this should be attributed essentially to three factors:

- the difficulty in building a knowledge base, both for problems in interpreting norms and for problems of formalisation;
- the simplicity of inferential engines: in the legal field, it appears necessary to develop more complex reasoning methods than the methods based on the resolution rules only;
- the mistrust of operators to utilise the tools provided in their work; this mistrust is not only based on the preceding conditions.

Knowledge-based systems facing this impasse, we now witness the birth and the development of systems based on knowledge with the purpose of providing the user with support in particular tasks and no longer focusing on decision-making only. Among these systems, let us mention the intermediary simulation in information retrieval (Rissland and Daniels, 1995; Matthijssen, 1995) and the assistance in document preparation (Bench-Capon *et al.*, 1995). These systems seem destined to have a real development, thus confirming an actual trend in the entire sector of AIs: knowledge moves towards improving or modifying a traditional product.

In this panorama, the development of networks poses the necessity of considering distributed-processing models to be adopted on networks, in order to verify what problems and advantages can derive from their use.

This paper deals with the interactions between knowledge-based systems and networks in light of the experiences matured at the Instituto per la Documentazione Giuridica (IDG). We suggest that the adoption of these models does not introduce unsolvable difficulties, whereas the use of these models enables a much broader access use.

After a brief description of the client/server architecture, an interface is presented to simulate the intermediary in information retrieval, as example of knowledge-based systems expressly studied to deal with network operations. Finally, a concise discussion is opened on the possibilities and problems that arise from integrating decision-support systems with network technologies, drawing attention to the possible developments that a broader use can entail. In this part, the questions doubtlessly outnumber the answers, hypotheses outnumber realisations. To conclude, we shall briefly discuss a project, currently in the phase of feasibility study, for the computerisation of the Court of Bologna.

2 Client/server architecture

The client/server architecture became an essential feature after the arrival of LANs. In this architecture, the server offers the other computers of the network (clients) several information resources: the server provides them with the application software and related data, and performs the relevant processing.

The supplying of data and programmes is often termed file-serving. It was the first function performed by the network server and enabled the integration of information resources in organisations utilising several computers. By supplying application software, a range of computers can use a single programme, thus reducing the purchasing costs of software and the disk space occupied. The applications are stored on the server to which the client goes to when he needs to run one of these applications. The server provides the client with the software requested, which is running just as though it were on his own local disk unit. Making data available this way makes it possible to avoid duplications when the information is of interest to users. The information is stored only on the server which offers it to those clients who need it. Among the numerous applications that can be brought back to this model, we cite the currently very popular WWW.

Shared processing between client and server, denoted with the term client/server, has enabled greater speed of execution, especially when data stored in the server must be processed. In this model, the higher-powered server not only stores data but also performs part of the processing. The task of the client (the user's computer) is to present the results.

The most significant example in this regard is provided by the applications which utilise SQL server functions. In these cases, the server manages a data bank (for example, the client's archives of a firm or of a law office) which can be interrogated using the SQL language (a standard for accessing relational data bases). The software for managing and querying the data bank in part resides with the client and in part on the server. The client-resident portion of the programme enables the user to formulate his query and also sees to sending this query on to the server. The server examines the query, extracts the data requested and sends them to the client. The client visualises, orders and prints the information received. In this type of application, the user must therefore only formulate his question without worrying about the localisation of the data.

Furthermore, this means of processing enables each user to modify or extend the shared data. (In these cases, however, mechanisms must be introduced to avoid conflicts between operations performed almost simultaneously by several users.) This means of processing is used, for example, by the WAIS network application.

3 User-support knowledge-based systems

As opposed to decision-support knowledge-based systems, we see that user-support knowledge-based systems are becoming a real field of application. In the domain of law, two large application areas can be distinguished:

- assistance in drafting documents;
- intermediary simulation in information retrieval.

Systems that assist the drafting of documents can present various levels of intelligence, ranging from simple help in drafting texts (Biagioli, 1992) to help in preparing legal argumentations (Bench-Capon and Staniford, 1995). Depending on the level of depth of the knowledge accessibility, these systems can achieve results which are even better than results by decision-support systems. The integration in the network depends on the accessibility of the knowledge required. Hence, we shift from simple integrated systems to cases in which the knowledge base is particularly large and essential. The problems encountered are the same as those faced by decision-support expert systems (see section 4).

Intermediary simulation has arisen from the need to assist the final user in consulting generally remote data bases: the application is specifically designed to manage network operations. The systems of this type are presented as “intelligent” interfaces possessing knowledge on a limited domain and able to manage communication with the data base. The processing model is that of client/server type in the strict sense: the information retrieval system performs the search on the server, while the interface (the client) assists in creating the query and interpreting the results.

In these systems, the knowledge required is not deep but, instead, segmented into small portions relative to the tasks to be performed. To point out the required functions of these systems, it is useful to analyse the task carried out by the human intermediary who helps the user satisfy his information needs. In brief, he performs the following tasks:

- He understands the user’s request: he talks to the user until any ambiguity in the request is eliminated.
- He translates the user’s request into the formal data-base query language and defines a search strategy.
- He analyses the results.

The knowledge of a computerised intermediary system having the above capabilities could be classified as follows:

System knowledge: on command language and tools of information retrieval systems;
Search knowledge: concerning the strategy used in the search process;
Domain knowledge: on the specific area in which the current search is made;
User knowledge: for each user, including the previous searches.

The various segments of knowledge possessed by the system can be codified with different formalisms. As an example of a system of this kind, we briefly present the CABALA project for intermediary simulation in information retrieval in the environmental domain.

3.1 The CABALA project

The aim of the CABALA project, supported by the Environment Committee of the CNR (National Research Council), is to build a system that facilitates user/computer interaction while searching environmental data bases.

The situation regarding information retrieval systems in this sector in Italy cannot be considered satisfactory. While there are many data bases with a large total number of documents on line, such as those in the Italgiure System data bases, environmental law information is currently spread over many different data bases and almost always organised under the classifications found in standard legal sources (national legislation, regional legislation, civil "massime" - summary of the principles in a case - of the Italian Supreme Court of Appeal, criminal "massime" of the Supreme Court of Appeal, etc.) and stored in different documentation centres (Italgiure, Celex, IUCN, etc.).

To overcome search problems and enable the user to find documents stored in different data banks in a logically uniform manner, a system has been proposed called CABALA (Consultazione Assistita di Basi Dati di Leggi Ambientali) (Guidotti *et al*., 1990; Guidotti, 1994).

CABALA can be divided into three principal logically independent parts which communicate only by exchanging messages:

- Query Generator
- Data Base Questioner
- Data Bases

The system permits the user to formulate a query by means of visualising the conceptual structure of the documents contained in the information-retrieval systems connected. The conceptual structure of the information retrieval system's contents (*domain knowledge*) is represented by a *semantic network.* One of the most important aspects in defining a knowledge-based interface to retrieval systems is to decide the level of depth on which knowledge must be represented. This means identifying all the concepts to ensure the completeness of retrieval, necessary and sufficient to avoid transforming the system into a decision-support system. In this sense, the semantic network has the advantage of providing a great deal of elasticity in defining the level of depth.

Figure 1 provides an overview of the CABALA architecture. Three main parts are distinguished: the Query Generator, the Database Questioner, the Information Retrieval System.

For practical purposes, a thesaurus was created by extracting a series of significant terms from legal texts, representing the nodes in the network. In any event, the relations used are semantically enriched, making the thesaurus into a semantic network.

The heart of the system is formed by the semantic network and Query Generator. The Query Generator interacts with the user by means of the interface which enables the user to select several nodes on the semantic network, thus identifying a route on the semantic network itself. This path represents the query which the user submits to the data base. Starting from this path, the Query Generator forms the conceptual query by using an algorithm explained in Guidotti *et al.* (1990).

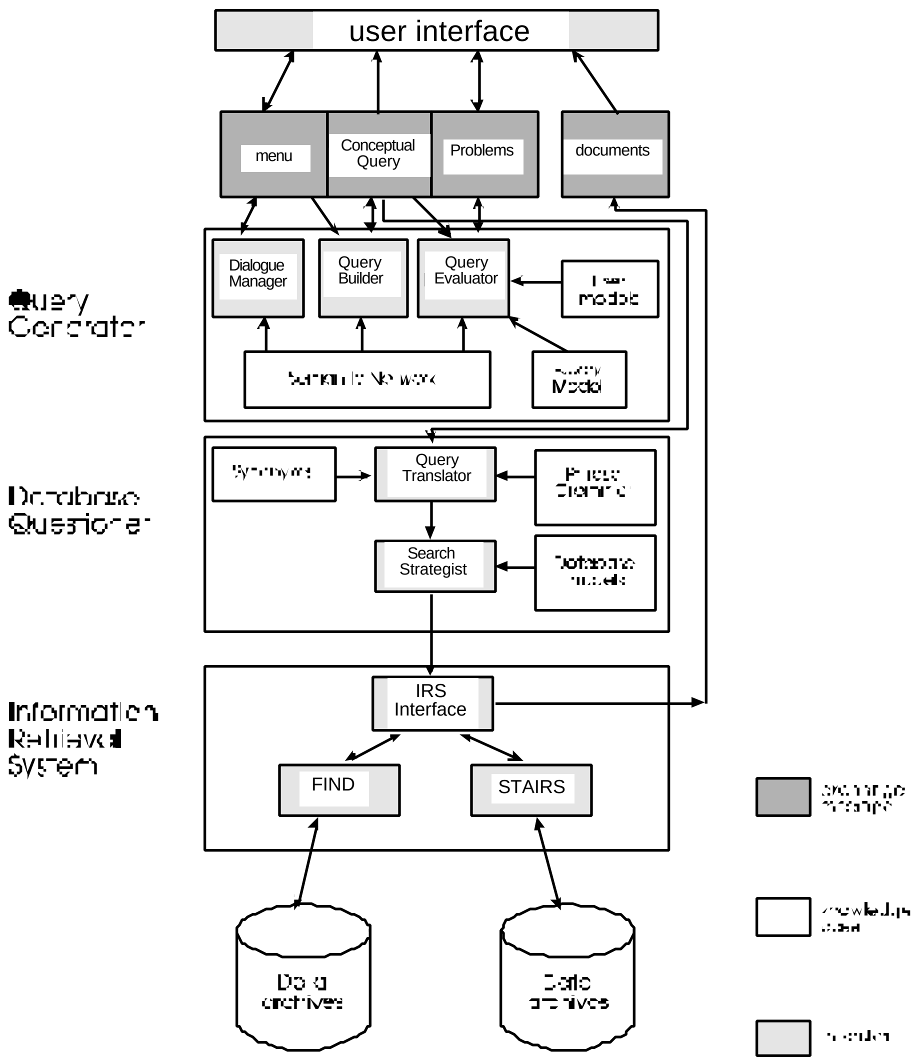

Figure 1: An overview of the CABALA architecture

The system must possess intrinsic knowledge of the overall contents and structure of the information retrieval systems in order to suggest changes to the user's query on the conceptual level. In this manner, other types of knowledge from the user or from other possible queries are used: *user and query models (user knowledge)*.

Once the conceptual query has been formulated, the Database Questioner transforms it into a query suited to being submitted to the available information retrieval systems. The system uses rules to convert the concepts into the phrase (*system knowledge*). The conversion rules have probability values which are modified by analysis of search results.

The Data Base Questioner translates the conceptual query produced by the Query Generator, written in the intermediate language, into the corresponding query for the data base, written in the data-base language. There are numerous equivalent real queries for each conceptual query. Two queries are said to be equivalent if they retrieve the same set of documents. The way real queries retrieve documents changes, however, with each query (*search knowledge*), as does the efficiency of the retrieval process. Therefore, in order to translate a query from the intermediate language into the data-base language, not only must we find an equivalent real query, but also select, from among the possible alternatives, the one that minimises the retrieval time. The set of real queries is then provided to the string search system.

The system uses parallel processes during conceptual query construction, while it uses sequential processes in the successive phases (translation into real query and data-base interrogation).

In Figure 1, we see three different types of modules representing (from the lightest to the darkest) a message exchange unit, the system modules and the knowledge base. The arrows show the type of knowledge used for each module. The user interface has access to the entire contents of the message exchange unit. In this manner, the system can show the user, at any time he makes a selection, the partially-conceptual query built up so far and the related problems, if any.

Our system architecture is based on the blackboard architecture, with the message exchange unit being the blackboard.

4 Decision-support systems

Decision-support systems are distinguished from the preceding systems by the dimensions of the knowledge bases available and the importance the latter have assigned in completely functioning systems.

In order to discuss the possibilities of adopting a distributed-processing model, we have to analyse the structure of these systems. The following fundamental parts can be distinguished:

- knowledge base (KB)
- inferential engine (IE)
- interface (I)

By distinguishing these components, the flow of information between them must be analysed; especially with respect to data travel on line. Data flows occur between the KB and the IE and between the IE and the I (see Figure 2).

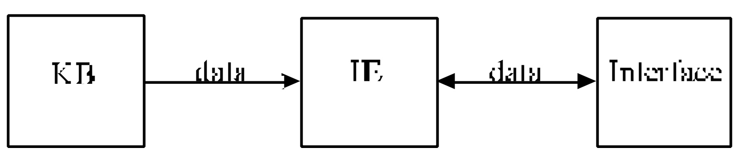

Figure 2: The data flows between KB and IE, and between IE and I

The KB is stored on disk and loaded in the memory where the IE resides to execute the query formulated by the user. While processing, data exchange between KB and IE is very intensive, both in rule-based systems and in CBR systems. In the former, the resolution method requires a continuous access to the KB in order to find the facts that permit the demonstration, while in the latter, the KB must be examined until a precedent is found which enables a match with the case under examination. Once a reply to the query is obtained, the interface presents the user the results of processing, visualising the justification of the conclusions obtained.

From the preceding exposition, it therefore appears possible to separate the system in a section of data flows from the KB to IE, and of data flows between the IE and the I.

4.1 Data flows between KB and IE

The data flow from KB to IE appears possible by loading the KB from a remote server at the moment the processing starts. Indeed, given the increased exchange of information from KB to IE, it is not possible to foresee access to the KB in the phase of execution.

To section the system on this level hypothesises a functioning similar to the file server or, to cite a popular application, to the WWW. Indeed, the data are managed by a remote server which does not co-operate in the processing of the data themselves. This solution can prove profitable in the presence of several remote archives. Finally, we must bear in mind that at the present line speeds, the remote KBs cannot have large dimensions, and thus the need to segment the KB arises. In rule-based systems, segmentation of the knowledge base may create anomalous situations unless the segments are disjointed. In particular, we mention

- redundancy, where a norm is nothing but a duplication of another norm or when a norm subsumes or is subsumed by an existing norm, and
- conflict between norms.

The theory of argumentation (e.g., Gordon, 1993; Prakken, 1993; Sartor, 1993) makes it possible to overcome these problems, obtaining justified conclusions from inconsistent sets of premises. This process requires an ordering to be defined over that set, since the ordering over the premises can be translated into an ordering over competing arguments. This fact is particularly relevant for legal reasoning, since lawyers effectively solve normative conflicts by using ordering relations. A lawyer solves normative conflicts by applying the following criteria: a hierarchical criterion (lex superior legi inferiori derogat), a temporal criterion (lex posterior legi anteriori derogat) and a specificity criterion (lex specialis legi generali derogat).

This suggests the convenience of segmenting the knowledge base into segments, *formalising a single normative text*, ordered according to the hierarchical criterion (constitutional laws, ordinary legislation, regional laws) and according to the temporal criterion. This type of approach to knowledge base segmentation is found in Guidotti *et al*. (1992). Here, information relative to the ordering introduced by hierarchical and temporal criteria was associated to every single segment of the KB which is more convenient than associating this information to every single rule. The information relative to the ordering of segments can therefore be maintained in the IE, which can use it to decide on access to the knowledge bases. This hypothesis suggests systems in which the KBs are maintained by different institutions. In particular, as an operative outlet to the prototype system presented in Guidotti *et al.* (1990), it was considered to propose to the Regional Councils, which have legislative power in matters of the environment, to create and maintain KBs at their Centres of calculus, where the legislation in matters of the environment of their competence would be formalised. In reality, we feel this hypothesis must be considered with extreme caution because, though the environment field is not wide, the building of several KBs, formalised in a non-homogeneous manner could prove unusable.

4.2 Data flows between IE and I

The data flow between IE and I appears less problematic. Indeed, the information which these components exchange is more limited than that of the previous case. In this model, the heaviest work is performed by the server, where the KB is also stored, while the presentation of the processing results, for example, the presentation of justification, is managed by the client. This solution is common to applications, such as the SQL server, and seems particularly profitable in the case of a server used by several users in a work group. Naturally, the server must be adequately powered in order to satisfy the requests of the single users in acceptable time; indeed, the solving of a query is a longer process.

The processing model makes it possible to foresee applications with several servers, i.e., KB plus IE. An example of such a system is a consultancy system determining pensions. This topic is at present quite pressing in Italy. Indeed, approval of the pension-system reform will modify and overlap the regulations, one different from the other, of numerous pension funds. Workers wish to understand how this reform will modify their position. It thus appears profitable to provide consultancy to citizens who require it by means of a decision-support system. Though the regulations are quite complex, the domain is limited and in this sector there are applications that have been developed successfully: the problem is certainly within the reach of today's technology. The solution advances separate servers for each pension fund. The personnel is then responsible for updating the KB. The servers have access to the archives of each fund in order to determine the work history of each employee, a necessary prerequisite for a reliable consultancy. As the system permits access to private information, we can expect an interface (client) in the peripheral offices, located throughout Italy, of the most important pension fund, INPS. Though in this case, the tool can be considered as suited to the task, the funds have difficulties in offering consultancy using automatic tools.

5 A project on the level of feasibility study

In this section, we present and discuss the project for automating the civil section of the Court of Bologna. Alongside office computerisation, recording and reducing sentences to massively employing automatic procedures was planned. This is in line with facilitating obtaining knowledge for the orientation of offices. We mention: the Bar, in view of articulating a defence without improper explorative expenses, as a condition to ensure a retrievability and usability of its product by colleagues and the Bar. Three areas have been singled out on which to experiment the recording activity:

- the area of ordinary contentious jurisdiction, with reference to the only sentences worthy of "qualified attention" for the nature of contentious jurisdiction and for the quality of their drafting;
- the area of corporate matters, a sector in which the judicial offices of Bologna have not yet expressed forms of orientation;
- the area of precautionary procedures, given the relevance of the topic dealt with.

For these areas, a documentary system will be created containing the integral text of sentences and permitting search both by textual words and by essential elements (data, number, judge's name, names of the parties, names of councils for the defence).

We cautiously hypothesise, alongside this search method, an indexing of sentences on the basis of a schematic representation of their rationes decidendi (or, in any event, of their content). This indeed would permit an advanced treatment of sentences, by means of forms of automatic reasoning. In this perspective, functions could be developed to make it possible to find only sentences in favour or against a given solution, ordered according to their pertinence, and to establish when a given sentence is implicitly overruled, distinguished, etc. In particular, as a subproduct of sentence drafting, performed by the single judge, a schematic representation of the sentence can be created to use in an expert system (CBR).

The information system shall be accessible both from inside the Court and from outside, both by means of direct access via modem and by means of a telecommunication network. The system based on a PC LAN must integrate the offices of Justices of the Peace, of the Magistrate's Court and of the Court. These three different organs of first degree magistrature have different competences based on the importance of the cases. In defining the project, the demand arose, expressed by magistrates, that the sentences archives of these three organs should be separate. Moreover, both the internal and external user should see a single consultation system. The solution that seems practical is that of a single expert system (CBR) with access to the three different KBs. This is dictated by the fact that the primary objective is the construction of the information retrieval system and that this solution enables experimentation from an isolated machine. Please note that the three KBs, in this case, are in fact disjoint.

6 Conclusions

Considering the foregoing items, we state that the adoption of distributed-processing models on knowledge-based systems presents no particular problems. Indeed, the widespread diffusion of networks inclines us to adopt these models, making a more extensive use possible. In particular, intelligent interfaces and integrated systems of office automation can find a wide use. The authors, however, believe that decision-support systems and, in particular, rule-based systems are, in any event, difficult to introduce to a wider, nonprofessional class of users.

References

Bench-Capon, T.J.M., and Staniford, G. (1995). PLAID - Proactive Legal Assistance, *Proceedings of the Fifth International Conference on Artificial Intelligence and Law*. ACM Press, Boston.

Biagioli, C. (1992). Law Making Environment. *Proceedings of the International Conference "Legal Knowledge and Legal Reasoning System"*, Tokyo.

Biagioli, C., and Mercatali P. (1990). Per la redazione automatica di testi legislativi. *Informatica e Diritto.* Firenze, XVI, 1, pp. 111-130.

Gordon, T.F. (1993). The Pleadings Game - Formalising Procedural Justice. *Proceedings of the Fourth International Conference on Artificial Intelligence and Law.* ACM Press, Amsterdam.

Guidotti, P. (1994). La simulazione dell'intermediario. *Proceeding of the conference 'Verso un Sistema Giuridico Integrale'*, Florence.

Guidotti, P., Lucchesi, L., Mariani, P., Ragona, M., and Tiscornia, D. (1990). A Simple Intelligent Interface to Data Bases on Environmental Law. *Proceedings of the International Conference on Database and Expert Systems Applications (Dexa 90).* Springer-Verlag, Wien.

Guidotti, P., Mariani, P., Sardu, G., and Tiscornia, D. (1992). Meta-level Reasoning. The Design of a System to Handle Legal Knowledge Bases. *Proceeding of the 7th Italian Conference on Logic Programming (GULP92).* Città Studi Editore, Milano.

Matthijssen, L.J. (1995). An Intelligent Interface for Legal Databases. *Proceedings of the Fifth International Conference on Artificial Intelligence and Law*. ACM Press, Boston.

Prakken, H. (1993). *Logical Tools for Modelling Legal Argument.* Ph.D. Thesis, Vrije Universiteit Amsterdam, Amsterdam.

Rissland, E.L., and Daniels, J.J. (1995). A Hybrid CBR-IR Approach to Legal Information Retrieval.. *Proceedings of the Fifth International Conference on Artificial Intelligence and Law.* ACM Press, Boston.

Rissland, E.L., and Ashley, K.D. (1987). A Case-Based System for Trade Secrets Law. *Proceedings First International Conference on AI and Law (ICAIL-87).* ACM Press, Boston.

Routen, T., and Bench-Capon T.J.M. (1991). Hierarchical formalizations. *International Journal Man-Machine Studies,* No. 35, pp. 69-93.

Sartor, G. (1993). A Simple Computational Model for Non-Monotonic and Adversarial Legal Reasoning. *Proceedings of the Fourth International Conference on Artificial Intelligence and Law.* ACM Press, Amsterdam.

Sergot, M.J., Cory, H.T., Hammond, P., Kowalski, R.A., Kriwaczek, F., and Sadri, F. (1986). Formallization of the British Nationality Act. *Yearbook of Law, Computers and Technology* (ed. C. Arnold). Butterworth, London.