The evaluation of legal knowledge based systems

Andrew Stranieri

School of Information Technology and Mathematical Sciences,
University of Ballarat, Ballarat, Victoria, Australia, 3353
61.3.53279440
a.stranieri@ballarat.edu.au

John Zeleznikow

Database Research Laboratory, Applied Computing Research Institute,
La Trobe University, Bundoora, Victoria, Australia, 3083
61.3.94791003
johnz@cs.latrobe.edu.au

ABSTRACT

Evaluation strategies to assess the effectiveness of legal knowledge based systems enable strengths and limitations of systems to be accurately articulated. This facilitates efforts in the research community to develop systems and also promotes the adoption of research prototypes in the commercial world. However, evaluation strategies for systems that operate in a domain as complex as law are difficult to specify. In this paper, we present an evaluation framework put forward by Reich and describe how this motivated the evaluation of our systems in Australian family law. Strategies surveyed include a comparison of linear regression with neural networks, user acceptance surveys, a comparison of system predictions with those from past cases, and a comparison of system outputs with those proposed by a panel of lawyers. Specific criteria for the evaluation of explanation facilities are also described.

Keywords

Evaluation, legal knowledge based systems.

1. INTRODUCTION

Reich [10] notes that the evaluation of intelligent systems raises difficult theoretical and pragmatic issues. Evaluations of knowledge based systems have often been performed in an ad-hoc manner without regard to theoretical concepts associated with the nature of measurement and the identification of appropriate evaluation strategies. Evaluation strategies for legal knowledge based systems (LKBS) need to apply to systems that represent knowledge in various ways including first order predicate logic based systems, rule based systems, case based reasoners, various hybrids or systems that represent knowledge in other ways. Evaluation strategies for legal knowledge based systems need also to apply to systems that perform various types of reasoning including systems that reason inductively, deductively, analogically or dialectically. Furthermore, evaluation strategies need to apply to systems developed for different purposes such as prediction, analysis, computer assisted learning, document retrieval or document drafting.

The need for effective evaluation strategies particularly arises in the context of the emerging trend toward knowledge discovery from databases (KDD) in law. Zeleznikow and Stranieri [20] discuss numerous key decisions that need to be made when applying data mining techniques to data collected from the legal domain. Decisions regarding outliers, missing values, the choice of data mining procedure, and estimation of generalisation performance can all impact significantly on systems that use knowledge discovered in this way and further highlights the need for some analysis of evaluation strategies.

The identification of a universal set of strategies for the evaluation of any legal knowledge based system is difficult given the breadth of type and use for these systems. Key features that distinguish the evaluation of legal systems from other knowledge based systems include:

- Jurisprudential theories can be used to evaluate a model of reasoning underpinning a LKBS
- The explanation facility is particularly important in an LKBS and requires fine grained evaluation
- LKBS that uses knowledge automatically derived from data require assumptions in the data pre-processing and data mining phases that impact on an evaluation.
- The manner in which disparities between expert predictions, particularly in a discretionary domain, are conceptualised impacts on evaluation criteria

Reich [10] has proposed a framework that is a useful starting point for the evaluation of knowledge based systems. A framework which facilitates a classification of strategies is a useful conceptual tool. The use of a framework can assist a developer in selecting an appropriate set of strategies and the same framework can help a reviewer accept or reject a system.

In the following section his framework is described. Following that we outline some strategies used to evaluate the Split Up system in order to demonstrate that the framework is a useful tool.

We then illustrate how the framework can lead to refinement of strategies for specific tasks such as the evaluation of explanation facilities.

2. AN EVALUATION FRAMEWORK

According to Reich [10], the determination of appropriate criteria for a comprehensive evaluation cannot be done without a conceptualisation of the nature of knowledge. He claims knowledge can be defined in two ways; structurally and functionally. In the structural definition, knowledge is a static entity that includes facts, rules and models that represent real world phenomena. This definition of knowledge enables the direct measurement of knowledge.

In the functional definition, knowledge has a purpose and cannot be measured directly but only indirectly by measuring the behaviour of a system that has knowledge. This view of knowledge has been preferred to the structural view by a number of researchers including Newell [9]. Defining knowledge as a structural entity alone is quite limiting. Reich [10] maintains that only with simple knowledge representation schemes and simple inference mechanisms can a prediction about the expected performance of a system be reasonable. Such a prediction becomes increasingly difficult once the knowledge schema becomes complex.

There are benefits inherent in the structural view of knowledge for the evaluation of intelligent systems despite limitations raised by [9]. Inspection of rules, facts and inferences made is certainly useful in the testing phase of rule based system development. Furthermore, the size of the rule set is often used as an indication of the performance of the system. A knowledge base consisting of 80,000 rules is very large and the system that infers with such a set is likely to make sophisticated inferences.

In addition to the structural/functional knowledge dichotomy, criteria for the evaluation of knowledge based systems can be based on quantitative or qualitative metrics. A count of the number of rules in a rule based expert system is quantitative while an end user opinion is qualitative. Furthermore, both qualitative and quantitative criteria can be discovered for knowledge that is structural and also for knowledge that is functional. This leads to four categories of evaluation criteria. Examples are summarised in Table 1.

The extent to which a knowledge base is readable or transparent to engineers is important because maintenance is difficult with knowledge bases that are overly complex or ill-structured. This is a qualitative measure and assumes the structural view of knowledge. Quantitative measures that assume functional knowledge are more commonly used in evaluating knowledge based systems. The Mycin system of Shortliffe et al [12] was evaluated by comparing diagnoses, with those of specialists. Buchanan et al [3] advocate user acceptance as an appropriate evaluation criteria for their intelligent system in the medical domain. User acceptance is a useful criteria because ultimately, the benefits of the system will only be realised to the extent that user's actually engage the system.

The extensive empirical study performed by Aleven and Ashley [1] evaluated the CATO system by comparing learning outcomes of a group of students that used CATO with one that did not. This strategy can be seen as one that assumes the functional view of knowledge using quantitative metrics. Outcomes were thus able to be assessed using statistical tests of significance.

Verification and validation are concepts that are traditionally invoked for the evaluation of any software system including knowledge based systems.[4] Verification refers to the process that involves checking for compliance with the system specifications and checking for syntactic and semantic errors in the knowledge base. Specification checking includes measures to ensure that the user interface, explanation facility, real time performance and security provisions reach the requirements specified in the system design. Checking for errors in the knowledge base depends on the knowledge representation. For rule based systems, this involves checking for redundant, conflicting, unreachable or missing rules. Verification methods assume a structural view of knowledge and use quantitative and qualitative metrics.

The concept of validation refers to the determination of the correctness of the system with respect to user needs and requirements. Validation criteria include comparisons with known results (eg past cases), comparison against exert performance, comparison against theoretical possibilities. Validation tests include the Turing Test, user acceptance surveys, direct comparison on random test cases between experts inferences and that of the system. Validation is a broad concept that covers structural and functional views of knowledge and qualitative and quantitative metrics. For example, comparison of results against theoretical possibilities assumes a structural view of knowledge and can be performed with quantitative metrics whereas the extent to which a user is satisfied by the system assumes the functional view of knowledge and may be measured qualitatively.

The framework suggested by Reich [10] avoids the explicit use of the verification and validation concepts but does not contradict their use. Further, the framework, of itself, does not suggest techniques for evaluation of knowledge based systems that have been used. However, it does provide a useful starting point for the design of an evaluation strategy for a particular knowledge based system. In order to illustrate this we describe evaluation studies of the Split Up system. This system uses rule sets and neural networks embedded in Toulmin argument frames for the prediction of property outcomes in the Family Court of Australia. Details regarding Split Up have been reported by Zeleznikow and Stranieri [20].

3. STRATEGIES USED TO EVALUATE SPLIT UP

Split Up was evaluated in each of the four categories outlined by Reich [10]. Each evaluation trial is listed in Table 2 and described in more detail in the following sections.

3.1 Structural / Qualitative

1. Domain expert assessment of the content and structure of knowledge base

Qualitative measures that focus on the structural view of knowledge involved feedback from experts and practitioners about the content of the knowledge base. A tree of relevant factors and argument structure used in family law were viewed positively by four family law practitioners outside the project.

	Qualitative metric	Quantitative metric
Structural knowledge	1. Readability of knowledge base to knowledge engineers	1. Number of rules 2. Correctness of inferences
Functional knowledge	1. Problem solving behaviour analysed qualitatively	1. Quantitative comparison with experts 2. Quantitative assessment of user acceptance of system

Table 1. Examples of qualitative and quantitative metrics for two types of knowledge

	Qualitative	**Quantitative**
Structural view of knowledge	1. Domain expert assessment of the content and structure of knowledge base. 2. Extent to which the knowledge base is ontologically specified	1. Extent to which a data mining technique has learnt patterns in training data 2. Extent to which a data mining technique generalises well from the training data
Functional view of knowledge	1. Domain expert assessment of the problem solving strategy adopted in Split Up 2. Extent to which the problem solving strategy is based on theoretical perspectives 3. Qualitative feedback from end users in different categories 4. Comparison of predictions made by Split Up with those reported in written judgements of cases	1. Comparison of predictions made by Split Up with those made by eight lawyers on facts from the same three cases

Table 2. Split Up evaluation studies categorised according to knowledge type and metric type

2. Extent to which the knowledge base is ontologically specified

Bench-Capon and Visser [2] argue that the conceptualisation of legal knowledge stored in a knowledge base ought to be explicitly specified with the use of an accepted ontological framework such as those developed by van Kralingen [5], Visser [17] or Valente [16]. The extent to which this has been achieved can be a strategy used to evaluate a knowledge base. This strategy would assume a structural view of knowledge, even if the ontology is that proposed by Valente [16] which focuses on the functional aspect of knowledge, because the strategy is measuring something about the knowledge base itself and not its use. The metric is far more likely to be qualitative because it is difficult to quantify the extent to which, or the accuracy with which a knowledge base has been specified. The conceptualisation of knowledge used in Split Up, for example, has not been ontologically specified to date.

3.2 Structural / Quantitative

1. Measurement of extent to which a data mining method has learnt patterns in training data

A measure of classifier performance typically used in classifier training in non legal domains is the number of examples correctly classified. As Weiss and Kulikowski [18] point out, this measure of machine learning model performance may not be adequate for all domains. They suggest a metric that includes the costs of predicting a positive outcome when the actual outcome was negative (called False positives) and the risks associated with predicting a negative outcome when the actual was positive (called False negatives). For example, a neural network, trained to discern the presence or absence of a disease will ideally, err far more times on the side of predicting a disease when there is none present than it will err in missing a disease which is actually present.

A False positive/False Negative analysis of errors is not warranted in family law because the direction of the error is rarely as critical as it is in medical diagnostic problems and other legal problems. Zeleznikow and Stranieri [20] describe an error metric that was more fine grained than simply counting the number of examples correctly classified in that the magnitude of an error was captured but stopped short of taking the direction of the error into account.

A strategy that measures the extent to which a data mining method has learnt patterns in training data can be used to compare data mining algorithms. However, Michie et al [7] describes considerable methodological difficulties inherent in comparisons of data mining methods. They report on an extensive study of a many data mining algorithms on a range of data sets that attempted to identify features of a data set that would suggest one data mining technique over another. Results indicate that comparative studies are difficult to formulate and general rules regarding any technique cannot easily be drawn. This experience was confirmed by our attempts to compare linear regression with neural networks on one small data set in Split Up. The difference between the means of the linear regression and neural outputs was not significant at the 0.05 level ($t = -0.25$, tcritical 2 tail = 1.98).

The comparison of linear regression and a neural network is based on the structural view of knowledge. In performing the comparison we are not concerned with the function of use of the system but are examining an aspect of the internal structure of knowledge. The comparison is quantitative because we exposed the linear regression formula and the neural network to all possible inputs and performed statistical tests of significance and explore descriptive statistics on the outputs.

2. Extent to which a neural network generalises well from the training data

Measuring the extent to which a network trained with a sample can perform on data drawn from the rest of the population is also a quantitative metric that assumes a structural view of knowledge. Reich and Barai [11] cite surveys of articles in leading machine learning journals that report less than 75% of articles clearly report the use of any technique that measures generalisation performance despite the importance of this estimate. Further, according to a recent empirical study by the same authors, the evaluation of generalisation performance depends markedly on the technique employed. They evaluated the same neural network trained with the same data using five re-sampling techniques used to estimate generalisation performance in data mining; re-substitution, hold out, leave one out, cross validation and bootstrapping. As described in Zeleznikow and Stranieri [20] the generalisation performance of Split Up networks were estimated using the resampling technique known as cross-validation. However, future research is planned to use other techniques, particularly bootstrapping in the light of the recent findings of Reich and Barai [11].

3.3 Functional / Qualitative

1. Domain expert assessment of the problem solving strategy adopted in Split Up

A qualitative analysis that presumes a functional view of knowledge involves the examination of the problem solving behaviour of the system by domain experts. This was performed constantly with the principal domain expert and occasionally with other experts. Experts expressed positive comments regarding the problem solving strategy with the exception of one domain expert. Dr Richard Ingleby advocated the implementation of a problem solving strategy that commenced at a 50/50 split of marital assets and inferred appropriate deviations from this starting point. The strategy implemented does not infer deviations from 50/50 starting point but rather, infers a percentage split.

2. Extent to which the problem solving strategy is based on theoretical perspectives

The extent to which the problem solving strategy used by a system is supported by a jurisprudential perspective is a criteria that has fuelled past debates on legal knowledge based systems. Certainly, a system that makes jurisprudential assumptions based on theory previously articulated is preferable to one that has no such basis. Zeleznikow and Hunter [19] survey debates surrounding the extent to which standard rule based reasoning makes positivist assumptions and the degree to which positivism is appropriate to model legal reasoning.

Stranieri and Zeleznikow [13] survey the use of argumentation as a means of representing knowledge in order to illustrate that most researchers that use the Toulmin [15] structure ultimately vary the original structure principally because it is based on an informal conceptualisation and not on a rigorous theory. The knowledge representation used in Split Up is a variation on the Toulmin structure. A knowledge representation frame based on a recognised argument structure is preferable to one that is developed for a particular problem but perhaps not as preferable to one based on more rigorous theory.

The data mining component in Split Up is supported by the jurisprudential movement known as legal realism. This adds support for the application of data mining in legal knowledge based systems but opens their use to criticisms that are directed generally to this jurisprudential perspective. Furthermore, in preparing data for neural networks some examples were eliminated from the training and test sets altogether because they were perceived to represent an exercise of judicial discretion that was quite inconsistent with other cases. The removal of some cases, legitimately tried, from the sample seems to infringe the legal concept of stare decisis where any case, whether decided in a higher Court or not can, conceivably be cited as a precedent. This concern is addressed in Stranieri and Zeleznikow [14] by drawing the distinction between traditional, local and personal stare decisis.

3. Qualitative feedback from users in different categories

End users of Split Up are Family Court judges, registrars and mediators, lawyers and other parties conversant with family law. Each category of Split Up user has different objectives and, as a consequence, the information needs of each user is different. Registrars of the Family Court are required to attempt to mediate a settlement before a dispute is tried by a judge in a forum known as an Order 24 conference. They use the Split Up system in order to convince parties to compromise. This involves informing litigants about the basics of family law and judicial heuristics.

Explanations for predictions made are used by registrars as a convenient way to convince parties that the predictions are accurate. However, the explanations are used to educate parties typically unfamiliar with family law about fundamental principles. Lawyers are less interested in educating their client but need to validate their own predictions and be reminded of cases and statutes that would strengthen (or weaken) their arguments. These needs derive from a lawyer's primary goal of achieving the best result for a client.

Judges are required to arrive at an equitable outcome in the shortest amount of time possible. They need to ensure adherence to local stare decisis; that their own judgement is consistent with that of other judges of the same Court. Judges would rarely be interested in explanations.

The Split Up system was demonstrated to three judges and five registrars of the Family Court and numerous other family law specialists. Opinions of the system were largely gathered informally by open ended discussion during and after a demonstration. An evaluation feedback form was developed and made available to some users before a demonstration to prompt discussion though we found all users to be forthright with their opinions. The evaluation feedback included questions that were loosely adapted from those on a user acceptance survey used by Buchanan et al [3] to evaluate their knowledge based system in the medical domain.

4. Comparison of Split Up predictions with past judgements in the Family Court

A comparison of Split Up predictions with those made by judges in past judgements assumes the functional view of knowledge but the metric is typically qualitative rather than quantitative. This is because a written judgement necessarily leaves many points implicit. A judgement that articulates every aspect of every inference would be exceedingly long and unreadable. Yet, leaving many factors implicit may presume too much of the reader. In comparing system predictions with written judgements, interim conclusions leading to the ultimate outcome must typically be read into most judgements. This introduces a degree of subjectivity which necessitates a qualitative assessment.

Split Up predictions were compared with five cases tried in the Family Court. Results suggest that predictions overall are quite good, though counter intuitive and often incorrect results sometimes occurred. Some Split Up departures from conclusions made in judgements can readily be explained by the small sample size (n=102). Other departures revolved around factors such as incest in one case that is not normally a relevant factor in property proceedings yet may have impacted on a judgement.

3.4 Functional / Quantitative

1. Comparison of predictions made by Split Up with those made by eight lawyers on facts from the same three cases.

A comparison of Split Up predictions with those made by a panel of lawyers assumes the functional view of knowledge. This metric was quantitative as actual predictions were empirically compared and contrasted. Eight family lawyers were asked to analyse three hypothetical cases and record predictions and explanations confidentially, in writing. The cases were hypothetical cases specifically constructed to explore a range of issues because the same breadth of issues could not be guaranteed with the selection of three cases at random. Lovegrove [6] found a similar problem in a study of sentencing behaviour of Victorian judges. He invented 40 cases for each of eight County Court judges to sentence and justified the use of fictitious cases on the grounds that a full range of issues can be explored if the cases were invented.

Results reported in Table 4 indicate that although predictions amongst lawyers were far from consistent, Split Up system fell within the range of those made by lawyers on all three cases.

The generation of an effective explanation for inferences drawn is particularly important for legal knowledge based systems. Issues concerning the evaluation of explication facilities that are more fine grained than the Reich framework are described in the next section.

	Case A	Case B	Case C
Split Up	50%	40%	55%
Lawyer 1	50%	35%	55-60%
Lawyer 2	50%	35-40%	55%
Lawyer 3	50%	40%	50-55%
Lawyer 4	50%	50%	45%
Lawyer 5	50%	40%	45-50%
Lawyer 6	50%	35%	20-25%
Lawyer 7	50%	35%	45-50%
Lawyer 8	50%	40%	50%

Table 4. Split Up prediction compared with Lawyers prediction

4. Evaluation of explanations

Moore [8] provides a thorough review of computer based systems that generate explanations. She identifies a set of requirements that are intended to act as criteria for the evaluation of explication systems. According to Moore [8] explanations should have the qualities listed below. The Reichian category that we believe to be most appropriate appears alongside each of the criteria. By superimposing the Reichian classification on Moore's criteria we see that the latter are more fine grained. This suggests that the framework suggested by Reich is a very general one that can be used to motivate more fine grained analysis of criteria specific to different tasks.

- Naturalness. Explanations should appear natural to the user. Explanations that are not structured according to standard patterns of human discourse often obscure critical elements of an explanation. (Functional/Qualitative)
- Responsiveness. An explanation facility must have the ability to accept feedback from the user and to answer follow-up question. (Functional/Qualitative)
- Flexibility. An explanation facility must be able to offer an explanation in more than one way in order to accommodate differences knowledge and abilities of users. (Functional/Qualitative)
- Sensitivity. An explication system should take into account the user's goals the problem solving situation and the previous explanatory dialogue.
- Fidelity. An explication system must accurately reflect the system's knowledge and reasoning. (Structural/Quantitative)
- Sufficiency. An explication system should be able to answer a range of questions users wish to ask and not be limited to those questions predicted by developers. (Functional/Qualitative)
- Extensibility. It should be easy to extend the explanation system to accommodate questions not conceived of at design time. (Structural/Qualitative)

Early explanation systems use pre-written text attached to knowledge chunks or applied simple transformations to produce explanations from program code. This type of explanation is very simple to generate and is natural. However, it is far from responsive, flexible, sensitive or sufficient. Furthermore, this type of explanation may not ensure fidelity in that text associated with a rule set may not necessarily be appropriate for the rule set because of knowledge engineering errors.

Second generation explanation systems engage the user in an explanatory dialogue. The criteria outlined above can only be realised within a dialogue. According to Moore [8], the generation of a dialogue requires a natural language generation facility that can generate multi sentential text. This requires knowledge of the domain, knowledge of the rhetorical structure of human discourse, and thematic knowledge that governs the shift in focus of attention within an explanatory dialogue. Currently, few explanations systems can claim to be second generation systems.

The explanation facility in Split Up has been well received by users according to user acceptance surveys though it does not engage the user in a dialogue. The facility is intimately linked with the Toulmin structure and is natural, responsive, and has fidelity to some extent, regardless of whether inferences were neural or rule based . However, the explication facility does not meet any other of the criteria set out by Moore.

5. Conclusion

We have described a framework proposed by Reich [10] that is useful for classifying evaluation strategies for legal knowledge based systems. The need for an analytic discussion about evaluation strategies particularly arises in order to evaluate LKBS that use knowledge bases derived from the knowledge discovery from database (KDD) process. Given, the diversity of legal knowledge based systems, a universal set of strategies that can be used to definitely evaluate any system is unlikely to emerge. Rather, a framework that facilitates the classification of strategies is useful for the creation of new strategies or the appropriate re-use of old ones. We have classified different strategies used to evaluate the Split Up system according Reich's framework. Future work aims to extend work begun on Structural/Quantitative types of strategies and, in particular to compare predictions made by neural networks with those made by other procedures such as logistic regression, fuzzy logic and induction algorithms.

6. References

[1] Alevan, V., and Ashley, K. D., Evaluating a learning environment for Case-Based Argumentation Skills *Proceedings of the Sixth International Conference on Artificial Intelligence and Law.* Melbourne. ACM Press. New York. pp170-179

[2] Bench-Capon, T., and Visser, P., 1997., Ontologies in Legal Information Systems; The Need for Explicit Specifications of Domain Conceptualisations. *Proceedings of the Sixth International Conference on Artificial Intelligence and Law.* Melbourne. ACM Press. New York. pp132-141

[3] Buchanan, B., Moore, J., Forsythe, D., Carenini, G., Ohlsson, S. and Banks, G., 1995. An Intelligent interactive

system for delivering individualised information to patients. *Artificial Intelligence in Medicine*. 7 (1995) 117-154

[4] Gonzalez, A, J., and Dankel, D, D., 1993. *The Engineering of Knowledge-based Systems. Theory and Practice.* Prentice-Hall. NJ.

[5] Kralingen, R. W. van., 1995. *Frame-based Conceptual Models of Statute Law*, Computer/Law Series, Kluwer Law International, The Hague. The Netherlands.

[6] Lovegrove, A. 1997. *The framework of judicial sentencing : a study in legal decision Making*. Cambridge University Press, Cambridge.

[7] Michie, D., Spiegelhalter, D. J and Taylor, C.C., (Eds) 1994. *Machine Learning, Neural and Statistical Classification.* Ellis Horwood. West Sussex. England.

[8] Moore, J. 1995. *Participating in explanatory dialogues: interpreting and responding to questions in context* MIT Press. Cambridge MA.

[9] Newell, A. 1982. The knowledge level. *Artificial Intelligence*. 18(1). Pp87-127

[10] Reich, Y., 1995. Measuring the value of knowledge. *International Journal of Human-Computer Studies*. Volume 42. No. 1. pp3-30.

[11] Reich, Y., and Barai, S. V., 1999. Evaluating Machine Learning Models for Engineering Problems. *to appear in Artificial Intelligence in Engineering, 1999*

[12] Shortliffe, E. H. 1976. *Computer based medical consultations: MYCIN.* New York: Elsevier.

[13] Stranieri, A., and Zeleznikow, J., 1999. A survey of argumentation structures for intelligent decision support. *to appear in Information Systems and Decision Support Systems, ISDSS99.* Melbourne. July 18-21, 99

[14] Stranieri, A., and Zeleznikow, J., 1998. Assumptions regarding the concepts of open texture and stare decisis for data mining in a discretionary domain. *Legal Knowledge Based Systems. Foundations of Legal Knowledge Systems. Jurix98. The Netherlands*

[15] Toulmin, S. 1958. *The Uses of Arguments*. Cambridge University Press. Cambridge

[16] Valente. *Legal Knowledge Engineering; A modelling Approach.* University of Amsterdam. IOS Press.

[17] Visser, P. 1995. *Knowledge Specification for Multiple Legal Tasks; A Case Study of the Interaction Problem in the Legal Domain.* Computer/Law Series, No 17, Kluwer Law International, The Hague, The Netherlands.

[18] Weiss, S and Kulikowski, C. 1992. *Computer Systems That Learn: classification and Prediction Methods from Statistics, Neural Nets, Machine Learning and Expert Systems*. Morgan Kaufman

[19] Zeleznikow, J., and Hunter, D., 1994. . *Building Intelligent Legal Information Systems*. Kluwer Law and Taxation. Deventer.

[20] Zeleznikow, J., and Stranieri, A., 1997. Knowledge Discovery in the Split Up project. Proceedings of the Sixth *International Conference on Artificial Intelligence and Law.* Melbourne. ACM Press. New York. pp89-97