On the Need for Standard Document Formats for eGovernance Applications of Legal Knowledge-Based Systems

Thomas F. Gordon,
Fraunhofer FOKUS, Kaiserin-Augusta-Allee 31, 10589 Berlin, Germany

Abstract. *eGovernance* is the use of Information and Communications Technology (ICT) to support the entire life cycle of law, regulations and other norms. Legal Knowledge-Based Systems have a great deal of potential in this application domain. Although a number of products for developing and using legal knowledge-based systems have been successfully used in deployed applications, the growth of the eGovernance market is inhibited by the lack appropriate industry standards. Starting from a process model for eGovernance, a number of document types required for e-governance applications of legal knowledge-based systems have been identified. Relevant existing or emerging document standards, such as MetaLex, RuleML and OWL, are assessed with regard to their suitability for use in this domain.

1. The eGovernance Market

eGovernance is the use of Information and Communications Technology (ICT) to support the entire life cycle of law, regulations and other norms [Reinermann 2002]. eGovernance is related to eGovernment, but has a different focus. Whereas eGovernment focuses on the use of ICT to support all the tasks of public adminstration, eGovernance applies ICT to support all actors involved in the production and use of norms.

Figure 1 shows the big governance picture, the life cycle of legislation, and elaborates on the diagram of [Macintosh 2003]. This is a kind of cybernetic control loop consisting of the following tasks or phases: agenda setting, policy analysis, policy creation, implementation and monitoring. The kinds of actors participating in the process are displayed around these tasks in the figure. The location of the actors is meant to suggest the tasks in which the actors primarily participate. You will notice that many of these actors are from civil society and the private sector, emphasizing once again that the governance "market" is not restricted to public agencies.

All of the tasks in this control loop deal with legislation, cases, regulations or other sources of norms in some way. The agenda setting task calls existing norms into question and puts issues in need of discussion on the political agenda. The task of policy analysis is to try to understand the issues, the interests of the persons and groups affected by the norms and to elicit high-level legislative proposals for resolving the issues and arguments pro and contra the alternative proposals. In the policy creation phase, the responsible legislative body, typically working in committees, makes use of the results of the analysis phase to formulate a policy and draft appropriate legislation, which then by some presumably democratic decision-making process gets enacted, perhaps after some revisions. The implementation task puts the policy into practice by producing regulations interpreting and operationalizing the legislation, reorganizing the responsible agencies as necessary to

execute the legislation, and designing and implementing the forms, computer systems and other artifacts required to support the work flows of the process. Finally, the purpose of the monitoring task is observe the effects of the implemented policy, to help critically evaluate whether it is meeting its goals, and to cope with any problems or conflicts which may arise. Thus the courts have a role to play in this phase, to the extent these conflicts are resolved by law suits.

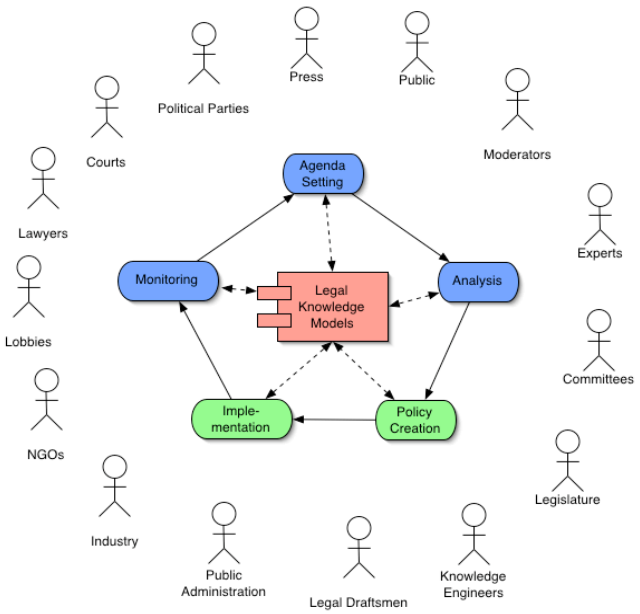

Figure 1. The Life Cycle of Legislation

Up until now we have been talking about governance in general. What makes this process **eGovernance** is the use of information and communications technology to support all phases of the process. We postulate that optimal IT support requires the use of appropriate computer models of the norms, integrated with and linked to the hypertext versions of relevant natural language texts, such as legislation, regulations and cases. These models, called "legal knowledge models" in Figure 1, can be of different levels of granularity or formality, depending on the requirements of the task. One focus point of our research is on finding ways to use multiple models together, in an integrated way, to work on some task. As will be explained in more detail in subsequent sections below, our current approach to solving this integration problem is founded on argumentation theory.

We have divided the five tasks or phases of the legislation life cycle into two groups, the top three and bottom two tasks in Figure 1, respectively. The top three tasks are those requiring a high level of support for interaction and collaboration between various stakeholders and interest groups. Thus, these tasks present opportunities to apply technologies for Computer Supported Collaborative Work (CSCW). More specifically, systems which have been developed for the Web for eParticipation or eDemocracy, such as Zeno [Gordon 2002] or DEMOS [Lührs 2003] could find application. The two tasks at the bottom, policy creation and implementation, provide greater opportunities to apply methods from Artificial Intelligence and Law, such as Legal Knowledge-Based Systems. Although we will focus on the use of these latter tasks in this paper, a comprehensive analysis of eGovernance would have to address collaboration support and eParticipation.

2. eGovernance Application Scenarios

We have developed a graduated series of eGovernance application scenarios for legal knowledge-based systems. Due to space restrictions, only the simplest and most complex of these scenarios will be presented here. We start with the scenario illustrated in Figure 2, which is entitled, somewhat ironically, "Conventional" Electronic Service Delivery. Of course, there is nothing conventional about this scenario at all, since it assumes that legal knowledge-based systems are being used. Nonetheless, even if this is not yet common practice, a variety of successful systems have been implemented and companies like Softlaw[1], RuleWise[2] and KnowledgeTools[3] have been building their businesses around this scenario.

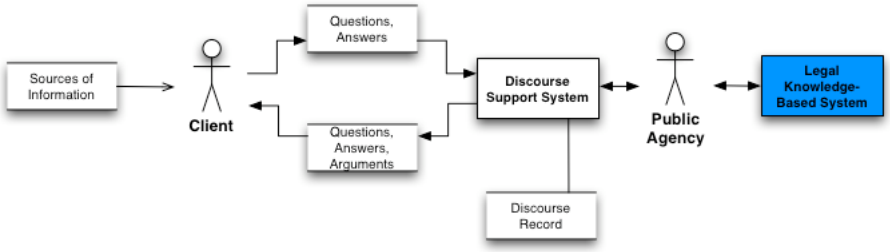

Figure 2. "Conventional" Electronic Service Delivery

The thing to notice about this scenario is that the legal knowledge-based system is used by the public agency to support its dialog with the client, who typically represents himself without the assistance of an advisor. Indeed, the public agency assumes multiple roles by both advising the client and deciding the case, for example by deciding whether a citizen is entitled to some social benefit. Clearly, these roles can conflict in concrete instances. Moreover, the public agency is in a position of relative power compared to the client, not only because has expertise in public administration and the relevant substantive law, but also because only it has direct access to powerful ICT systems including, in this scenario, legal knowledge-based systems.

The role of the client in this scenario is quite restricted. He or she comes to the public agency with a question. The agency asks questions of its own to learn about the facts of the case, perhaps by having the client complete one or more forms. After this data has been collected, the public agency consults its legal knowledge-based system, and perhaps other sources, and makes its decision. The amount of explanation or justification, in the form of arguments, will vary, but in general the client does not have access to the legal knowledge-based system to check the results or to experiment with alternative analyses of the case using other facts. (The legal consequences of filling out forms a particular way may not have been clear and there may have been some room to interpret the fact situation differently.)

Another limitation of this scenario is that the public agency uses only a single model of the norms for each kind of transaction or process. This means that, contrary to the "pure teachings" of Artificial Intelligence, control knowledge and domain knowledge are usually mixed together in the models. For example, only those parts of the domain (or domains) which are important for a particular, pre-defined, task are included in the model. This makes it difficult to reuse and maintain domain models of legislation, increasing overall

[1] http://www.softlaw.co.uk
[2] http://www.rulewise.com
[3] http://www.knowledgetools.de

costs and further inhibiting the adoption and spread of legal knowledge-based systems. They also make the models much more "brittle", since they cannot handle cases outside of the restricted set of intended tasks.

Lifting these restrictions bring us to the subsequent scenarios. Due to space restrictions, we will now jump immediately to the most advanced scenario, which we call the "parity" scenario, shown in Figure 3.

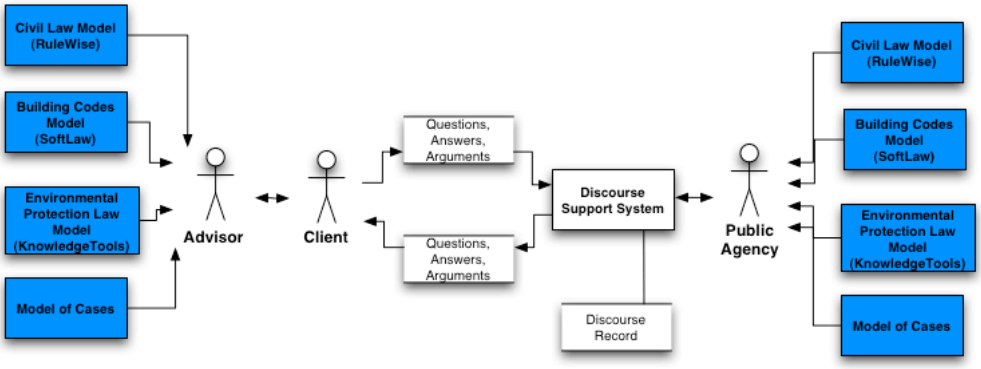

Figure 3. Parity Scenario

In this scenario, the client is empowered with the assistance of an expert advisor and advance ICT support, including a multitude of legal knowledge-based systems, indeed the same ICT support and models available to the public agency. We call this the parity scenario, since it strengthens the client's position in the dialog vis a vis the public agency. Indeed, since this scenario is as much about empowering citizens and increasing the transparency of government decisions as it is about improving the efficiency of bureaucracy, this scenario could be considered a contribution to eDemocracy as well as eGovernance.

This scenario enables clients to assess their legal situation on their own, prior to or during the dialog with the public authority. This is what Peter Johnson of Softlaw calls "self-assessment" [Johnson 2000]. Software for helping people assess their legal rights and obligations is not unheard of. One need only think of the popular packages for completing tax forms or preparing wills.

More novel in this scenario is the ability of both clients and public agencies to make use of multiple models to analyze a single case, in an integrated fashion. These may be competing models (products) from different companies, analogous to self-help books by different authors, or models of different legal domains relevant to the case. Figure 3 illustrates this idea with several models about different legal domains relevant for obtaining a building permit, including civil law, the building codes, and environmental protection law. Moreover, we imagine each of these models having been built using a different modeling tool and methodology, from different companies.

3. Use Case Analysis

This section summarizes [Gordon 2003], which provides a high-level analysis of use cases covering the e-governance scenarios discussed in Section 2, with the goal of identifying document types which, if standardized, would enable a modular set of interoperable components for building legal knowledge-based systems. This analysis is based in part on prior work by the POWER project [Engers 2001, Engers 2003] and the survey of AI

methods and tools for supporting legal argumentation by Bench-Capon, et. al., in [Bench-Capon 2003]. Use cases are a software engineering modeling and diagramming technique, standardized as part of the Unified Modeling Language. Use cases are very high level models showing how different kinds of actors interact with a system to carry out tasks. They are useful for helping to get an overview of the desired functionality of the complete system and are an "essential tool" for identifying requirements [Fowler 2000].

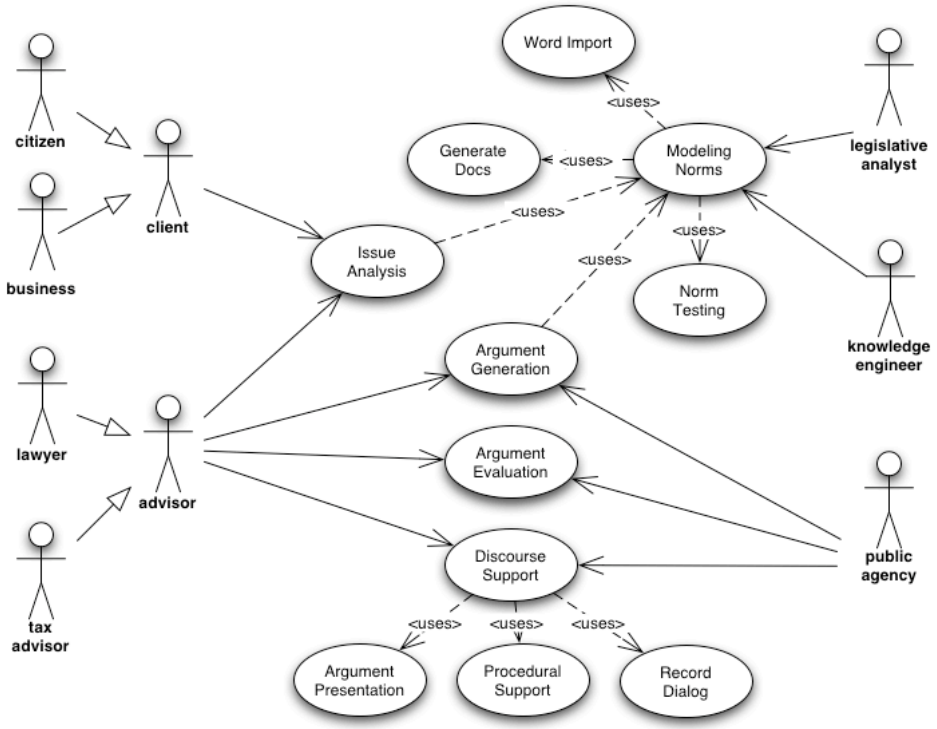

Figure 4. Legal Reasoning Uses Cases

Figure 4 is a UML diagram illustrating the use cases analyzed. There are five main kinds of actors participating in these use cases: *legal analysts* (lawyers) and *knowledge engineers* (computers scientists) work together to build formal, computer models of legislation; *clients* (citizens or businesses) consult *advisors* (such as lawyers or tax consultants) take part in a structured dialog, for example to apply for social benefits or a building permit, with a *public agency*. Both the advisor (or the client) and the public agency use these models and other resources to analyze the issues of a case and to generate arguments.

Rather than taking a technology-driven approach, by starting with existing legal knowledge-based systems and looking for common document types, a requirements-driven approach was taken, starting with e-governance use cases. The following document types were identified as candidates for standardization:

- A Microsoft Word template for marking up legislation and other sources of norms using Word;
- A markup language and exchange format for legislation and other sources of norms;
- A knowledge representation language and exchange format for formal models of norms;
- An exchange format for legal cases, including facts, issues and decisions;
- An exchange format for arguments and argument graphs; and
- Possibly a representation language and exchange format for argument schemas, if this is not part of the language for norm models.

4. Document Types of the Modular Architecture

Our analysis of the various use cases of legal knowledge-based systems for eGovernance scenarios identified a number of processes and document types used or produced by these processes. In this section, existing or emerging standards are discussed which might be relevant as a basis for realizing standards for these document types for the legal knowledge-based systems industry.

Bench-Capon, et. al., performed a related analysis which we have found very useful [Bench-Capon 2003]. However, one important difference between these two studies is that Bench-Capon analyzed only research prototypes, while we focus on industry standards. This is because we are interested in transferring research results into practice and need an overview of the supported and mature standards which can be used to build real applications.

We will not discuss the Word template further here, except to say that such templates are in use by governments and that it presumably will be a difficult task to get governments at the municipal, state and federal levels to migrate to a standard, not to mention European or international standards.

The markup language for legislation and other source text of norms should be based on XML. The Dutch MetaLex XML schema [Boer 2002] was designed precisely for this purpose, in the European ePower project, and has been validated in some applications in the Netherlands. Some publications refer to MetaLex as a "standard", however it is not clear whether it is a de facto or de jur standard and whether it is accepted as a standard anywhere outside of the Netherlands. However, whatever its normative status, MetaLex should be taken into consideration as a starting point for any international standardization activity on this topic, at least within Europe.

A standard document format for formally modeling norms is much more problematical. There are a number of efforts to create standards for first-order predicate logic (FOL) or well-defined subsets of FOL. In the United States, there is the Knowledge Interchange Format[4] (KIF), which uses Lisp s-expressions. KIF has the advantage of being both machine-readable and (relatively) easy for humans to read and write. Probably more promising from an international and also commercial perspective are the emerging standards resulting from the Semantic Web initiative of the World Wide Web Consortium (W3C), in particular RDF, OWL and RuleML.

RDF, the Resource Description Language[5], is an XML Schema for representing arbitrary directed, labeled graphs as a set of subject-verb-object triples. Interestingly, RDF has been given a logical, model theoretic semantics. OWL, the Web Ontology Language, is a "candidate" W3C standard for defining "ontologies" using Description Logic[6]. OWL is an XML schema built on top of RDF. That is, OWL documents are also RDF documents. Actually, there are three version of OWL – OWL Lite, OWL DL (for "Description Logic") and OWL Full – with somewhat different semantics and computational properties. However, they are upwards compatible, so that, e.g., OWL Lite documents are also OWL Full documents and retain their semantics.

The interesting thing about OWL is that it is a standard version (or rather versions) of Description Logic. Description Logic is a cleaned-up, logical version of several of the "scruffy" approaches to knowledge representation, such as frames and semantic nets. Description Logic is a fairly expressive subset of FOL. Its expressiveness overlaps that of Horn clause logic, the logical foundation of Prolog. That is, some things can be expressed

[4] http://logic.stanford.edu/kif/dpans.html
[5] http://www.w3.org/RDF/
[6] http://www.w3.org/TR/owl-ref/

in Description Logic that cannot be expressed in Horn clause logic, and vice versa. For example, Description Logic has a "real" negation operator which does not depend on the closed-world assumption, which is of questionable applicability for legal reasoning, where typically not all the relevant facts are known.

Description Logics were designed primarily for describing and defining concepts. Thus, they are well suited for defining terminology and constructing "ontologies". Indeed their primary intended application in the context of the semantic web is to construct cataloging schemes ("metadata") for describing and indexing resources on the Web, to help people find things.

For knowledge representation purposes, ontologies are sometimes considered to be useful only for declaring the predicates which can be used to formulate rules, but not sufficient as a complete knowledge representation in its own right. But this view may be mistaken when Description Logics like OWL are used to build ontologies. Description Logic is expressive enough that no further rule language is necessary to model many interesting and complex relationships. However, we are aware of no prior work which has attempted to model significant portions of legislation or other sources of norms using only Description Logic, so it suitability for this purpose remains to be empirically validated.

That said, there are still many adherents of rules for knowledge representation. The RuleML intiative[7] is an attempt to create an XML standard for rules. This emerging standard is also targeted at the semantic web. The intention of the initiative is to submit the specification to the W3C for approval, but this has not yet happened. RuleML is very ambitious, as it aims to provide a neutral exchange format for just about every kind of rule formalism, including SQL queries and views, Horn clause logic, Prolog rules, and production rules. It also aims to be interoperable with OWL. But these various rule languages have very different semantics and it is unclear how the RuleML initiative will be able to overcome these semantic differences in order to be able to exchange "rule bases" between any pair of rule-based systems. For example, how could a set of production rules or Prolog rules, both of which have procedural, nondeclarative semantics, be converted into first-order logic? Or how can a set of sentences in full first-order logic be converted into the Horn clause subset? The purpose and goals of RuleML are still somewhat unclear, at least to this author.

However exciting OWL and RuleML may be, they were not designed specifically for legal reasoning. Much work in Artificial Intelligence and Law has revealed the special requirements of legal reasoning and argumentation, also from a knowledge representation perspective. Just about every kind of logic or reasoning schema has been shown to be applicable to some kind of legal reasoning task, including defeasible or nonmonotonic logic, case-based and analogical reasoning, higher-order or metalevel reasoning and, last but not least, various deontic logics. In recent years work on defeasible legal reasoning and case-based reasoning has begun to converge. But the field is still quite a long way from achieving a synthesis or consensus on how to represent and reason with all kinds of norms.

In the near and middle term, the legal knowledge-based systems industry probably will have to continue to take a pragmatic approach. The leading legal knowledge-based system companies in Europe use different modeling techniques. Softlaw's tool, Statute Expert, uses production rules. KnowledgeTools models rules using an and/or tree representation of propositional logic. And RuleWise adopts the ePower method, which uses UML/OCL, the software engineering standard, to model the concepts and rules of legislation as UML classes and OCL constraints, respectively. Once again it is difficult to imagine a standard

[7] http://www.dfki.uni-kl.de/ruleml/

format for exchanging domain models between these systems. The semantics and expressivity of the formalisms used are quite different.

More promising as an approach for integrating these and other systems and making them interoperable is the idea of extending them to generate arguments justifying or explaining their reasoning. If the format for arguments were standardized, several such knowledge-based systems could be used together to work on legal problems having issues from several legal domains, as discussed in the section on application scenarios. However, there are currently no standards or emerging standards for this purpose. The closest thing I am aware of is the XML argumentation Document Type Definition (DTD) developed in the Araucaria project [Rowe 2003], which can represent both argument graphs and argument schemas.[8] There is also a new European project on this topic, called Argumentation Service Platform with Integrated Components (ASPIC). The goals of ASPIC include consolidating the many research results on argumentation and "proposing a set of open pre-standards representing the current state of the art in argumentation languages and methods, developing a suite of open source software components embodying these proposals, and demonstrating their use in demonstrator applications."[9] The ASPIC consortium includes some of the best computer science research groups on argumentation in Europe, including two prominent members of the AI and Law community, Trevor Bench-Capon at the University of Liverpool and Henry Praaken of Utrecht University. Unfortunately, none of the industrial partners have any experience in the field of legal knowledge-based systems.

This leaves the issue of a standard format for legal cases. MetaLex is intended to be suitable for marking up cases and well as legislation, but what is of interest here is not such much marking up the structure of the natural language text of the published case "opinion" as a standard for exchanging formal models of cases as required for using case-based reasoning systems from Artificial Intelligence and Law. At the Evaluation of Legal Reasoning and Problem-Solving Systems workshop at ICAIL 2003 in Edinburgh, the lack of a shared repository of cases for use by the research community was addressed. So there is at least some awareness in the community of the need to find a way to exchange case data. But it remains open whether some kind of exchange format for cases will eventually result from this discussion. As yet no working group has been set up to follow-up on this issue.

Perhaps the OWL Description Logic standard could be used to model cases. I am aware of no work on using Description Logic to model cases for use by a case-based reasoner, but the similarities between Description Logic "roles" and the factors and factor hierarchies of legal case-based reasoners in the HYPO tradition [Ashley 1990; Aleven 1997] are quite striking. Also, Karl Branting's GREBE system used semantic nets to model cases [Branting, 1999]. Since Description Logic grew out of the research on semantic nets, among other influences, this is another indication that Description Logic, and thus the OWL standard, might be capable of modeling legal cases.

5. Conclusions

Is the glass half-full or half-empty? On the one hand legal knowledge-based system have found their way into practice and have been quite successfully applied in a number of eGovernance applications. Several companies have emerged which produce legal-knowledge-based system tools and applications. On the other hand, industry standards for document types of critical importance to eGovernance scenarios, such as legal domain

[8] http://www.computing.dundee.ac.uk/staff/creed/araucaria/

[9] The quotation is from Part B of the ASPIC proposal.

models, case bases and argumentation graphs, do not yet exist. Such industry standards are essential for helping to make these systems interoperable and reduce barriers to the continued growth of the eGovernance market. Although standards emerging in the context of the semantic web initiative of the World Wide Web Consortium, such as OWL and perhaps later RuleML, will help to alleviate this situation, they are not sufficient. Standards are needed which take into consideration the specific requirements of modeling legal knowledge and legal arguments, as identified during over twenty years of research in the AI and Law community.

References

Aleven, Vincent. "Teaching Case-Based Argumentation through a Model and Examples." Ph.D. University of Pittsburgh, 1997.

Ashley, Kevin D. Modeling Legal Argument: Reasoning with Cases and Hypotheticals. Artificial Intelligence and Legal Reasoning Series: MIT Press, Bradford Books, 1990.

Bench-Capon, Trevor, et al. "Computational Models, Argumentation Theories and Legal Practice." Argumentation Machines; New Frontiers in Argument and Computation. Eds. Chris Reed and Timothy J. Norman: Kluwer Academic Publishers, 2003. 85-120.

Boer, Alexander, Rinke Hoekstra, and Radboud Winkels. "Metalex: Legislation in XML." Legal Knowledge and Information Systems (Jurix 2002). Eds. T. Bench-Capon, A. Daskalopulu and R. Winkels: IOS Press, 2002. 1–10.

Branting, L. Karl. Reasoning with Rules and Precedents: A Computational Model of Legal Analysis. Dordrecht: Kluwer Academic Publishers, 1999.

Branting, L. Karl. "An Agenda for Empirical Research in Ai and Law." ICAIL Workshop on the Evaluation of Legal Reasoning and Problem-Solving Systems. Eds. Karl Branting and Steffi Brüninghaus. Edinburgh: International Association of Artificial Intelligence and Law, 2003. 28–35.

Engers, Tom M. van, et al. "Power: Using UML/OCL for Modeling Legislation – an Application Report." International Conference on Artificial Intelligence and Law (ICAIL 2001). Ed. Henry Prakken. St. Louis, 2001. 157–67.

Engers, Tom M. van, and Margherita R. Boekenoogen. "Improving Legal Quality – an Application Report." International Conference on Artificial Intelligence and Law (ICAIL 2003). Ed. Giovanni Sartor. Edinburgh: ACM, 2003. 284–92.

Fowler, Martin, and Kendall Scott. UML Distilled – a Brief Guide to the Standard Object Modeling Language. 2nd ed: Addison Wesley Longman, Inc., 2000.

Gordon, Thomas F., and Gernot Richter. "Discourse Support Systems for Deliberative Democracy." eGovernment: State of the Art and Perspectives (EGOV02). Eds. Roland Traunmüller and Klaus Lenk. Aix-en-Provence: Springer Verlag, 2002. 248-55.

Gordon, Thomas F. "A Use Case Analysis of Legal Knowledge-Based Systems." Legal Knowledge and Information Systems (JURIX 2003). Utrecht: IOS Press, Amsterdam, 2003.

Johnson, Peter. Legal Knowledge-Based Systems in Administrative Practice and Electronic Service Delivery (E-Government), 2000.

Lührs, Rolf, et al. "How to Grow? Online Consultation About Growth in the City of Hamburg: Methods, Techniques, Success Factors." EGOV 2003. Ed. Roland Traunmüller. Prague: Springer Verlag, 2003. 79–84.

Macintosh, Ann; International Teledemocracy Centre; Napier University; Edinburgh; personal communication; 2003.

Reinermann, Heinrich, and Jörn von Lucke. "Speyerer Definition Von Electronic Governance." Electronic Government in Deustchland. Eds. Heinrich Reinermann and Jörn von Lucke. Speyer: Forschungsint. für Öffentliche Verwaltung, 2002. 9-19.

Rowe, G. W. A., C. A. Reed, and J. Katzav. "Araucaria: Marking up Argument." European Conference on Computing and Philosophy. Glasgow, 2003.