

ON-LINE: An Architecture for Modelling Legal Information

André Valente and Joost Breuker

Department of Computer Science and Law, University of Amsterdam

Kloveniersburgwal 72, 1012 CZ, Amsterdam, The Netherlands

e-mail: andre@lri.jur.uva.nl, tel: +31-20-5253499, fax: +31-20-5253495

Abstract

This paper describes ON-LINE (ONtology-based Legal Information Environment), an architecture for a legal workbench which combines two major functions: legal information serving and legal analysis. Some of the main features of ON-LINE are: the integrated storage and representation of legal text and knowledge by using interconnected knowledge and text repositories; a representation of legal knowledge based on a functional ontology of law; the emphasis on legal modelling as a central task in legal practice. ON-LINE comprises three main modules. The *Legal Information Server* is able to retrieve legal information based on either textual or conceptual search. The *Legal Information Modelling Toolkit* is a collection of integrated tools to transform legal text into legal knowledge. The *Legal Analysis Environment* contains reasoning tools to perform two of the central legal tasks: assessment and planning. The architecture is intended to be a basis for experimentation, and it is therefore highly extensible. ON-LINE is partially implemented in Common Lisp and it is supported by the LOOM system.

1 Introduction

This paper presents ON-LINE (ONtology-based Legal Information Environment), an architecture for storing and retrieving legal information and reasoning with legal knowledge. ON-LINE is partially implemented in a research prototype system, and is used to integrate and test proposed and formalized conceptualizations of legal knowledge representation and reasoning methods [15]. The representations are based upon an abstract ontology (summarized in section 2.2) that covers the legal domain [15, Chapter 4]. Part of the legal knowledge consists of the description of norms which are expressed as deontic functions. A formal account of these deontic functions, which provides an alternative for (standard) deontic logic approaches, has been presented in [16]. Moreover, an algorithm which draws conclusions about the normative status of a case (allowed; not-allowed) has been developed as a core method for analyzing legal consequences [16]. Also representational solutions for other types

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission.

© 1995 ACM 0-89791-758-8/95/0005/0307 \$1.50

of legal knowledge (world knowledge, causal/responsibility knowledge) have been developed [15]. In this paper, we focus on the multiple functions which these elements support when integrated into the architecture proposed. ON-LINE contains reasoning modules based on the models for legal assessment and legal planning developed in [17, 15], as well as on supporting tools for legal modelling and design (drafting regulations). This integration of functions and support makes ON-LINE a legal workbench, rather than a legal expert system annex data-base. The key for the integration is the theoretical framework provided by the ontology.

This article is structured as follows. In section 2, some design issues of ON-LINE are discussed: (i) a global functional description, (ii) a short overview of the ontology and its role in providing the modelling framework for the tools and reasoning modules, (iii) its information bases, which comprise both (legal) text, knowledge representation and (labeled) links between these and (iv) some details of the knowledge representation employed. The architecture (structure, modules, etc.) is explained in section 3. Section 5 discusses some of its problems and limitations, and section 6 compares it with related work. Finally, section 7 presents the conclusions.

2 Some Design Issues of ON-LINE

In conceiving ON-LINE, we started from the idea that the activities of legal practitioners invariably involve *legal modelling* (or *interpretation*), whether it refers to (hypothetical) cases, to regulations or to the intended effect that some regulation should have. These interpretations may have varying degrees of freedom, biases and idiosyncrasies, which prohibit full automation except for specific niches in the world of law (e.g. routine cases). However, this does not mean that everything in legal modelling is at variance. The major conceptions of what law is about remains relatively stable over decades of legal practice and jurisprudence, and can be expressed in the form of an ontology (see 2.2). Because legal sources are all in textual form, their access during all activities should be supported as well: getting this access may in fact be a major legal activity as e.g. in finding relevant statute or precedence for a case. ON-LINE is based on the idea that legal problem solving is to some extent a global modelling activity in which the practitioner ‘experiments’ with alternative models (interpretations) of the legislation and/or of a case in order to reason about their consequences. This modelling can be goal-oriented (i.e. aiming to reach a certain legal conclusion) or not. Its output is a model of the case and, sometimes, a specific model of the legal sources.

2.1 Tasks and functions in legal practice

In legal practice, three major kinds of tasks can be distinguished:

- Drawing the legal consequences of a case. This involves both the modelling of a case, and an assessment of relevant norms and responsibilities. These problems are typical part of a legal *assessment* task [5, 17].
- For most legal practitioners, their activities involve more than drawing or foreseeing the legal consequences of cases. They may also have to take or advice legal actions, which are performed by solving *planning* problems.
- A specific legal task is the drafting of regulations (statutes, contracts, legal documents). This means the translation of a model of intended states in the world into normative and other types of rules. The assembly of these rules into a coherent, well structured regulation involves the solution of typical *design* problems.

The latter two tasks are typical synthesis tasks, i.e. ultimately, they involve solving a synthesis problem (assembling a plan or a regulation). However, as a common denominator, they also involve legal assessment as a subtask (e.g. testing the draft of a regulation by assessing hypothetical cases). This common denominator is the core of the *legal analysis* service of ON-LINE. The legal analysis service of ON-LINE embeds a number of problem solving modules which represent together a general problem solving competence. Besides this problem solving competence, a workbench should also have information management facilities. As the use of legal sources is of decisive importance for the legal practitioner, ON-LINE is constructed around these information management facilities.

These two main types of architectural facilities actually correspond to two major types of AI & Law applications. *Legal information retrieval systems* store legal sources in textual form, and offer a number of mechanisms for their retrieval [2]. AI techniques can be used in the retrieval process, e.g. by using a representation of legal concepts as a basis for search (see e.g. [11]). *Legal analysis systems* are abundant in the literature, and they comprehend most known legal expert systems. Legal analysis is a broad category of applications which comprehends more or less all types of tasks which aim at finding the legal conclusions for a certain case, whether in a goal-oriented manner or not. Most AI & Law systems belong to either one of these categories. One of the major (and innovative) functions of ON-LINE is to *integrate* legal information systems and legal analysis through modelling (see also section 3).

2.2 Ontological basis for legal knowledge representation

Many legal information systems (e.g. [11, 7]) have employed a representation of the concepts used in a law as a twin representation of the text which was used for both searching and reasoning purposes — a technique named *conceptual information retrieval*. These systems usually represent only *domain concepts* i.e. the concepts which describe the domain the law is about, e.g. value added tax, traffic, etc. These concepts are represented using concept-based formalisms such as semantic networks or conceptual graphs. Somewhat in opposition, legal analysis systems have traditionally employed rule-based or case-based representations (or, far less frequently, some kind of logic). ON-LINE tries

to bridge this gap by shifting the problem. Instead of using one of these formalisms as a homogeneous knowledge representation formalism, we use a knowledge representation based on a specific theory of ‘what law is made of’ — an *ontology*. Specifically, we have elaborated a *functional ontology of law* which distinguishes a number of (primitive) functional types of legal knowledge based on a teleological view on the legal system. Each of these types of knowledge has its own representation in an abstract formalism which is (roughly) independent of the specific implementation formalism to be used for reasoning with it.

We have argued in [14] in favor of an ontological basis for AI & Law research in general. Briefly, we pointed out the following advantages. First, an ontology is a powerful basis for a divide-and-conquer strategy. When a number of primitive categories of legal knowledge are identified, and the nature of their interrelations is explained, the problem of representing legal knowledge can be split accordingly in parts which can be analyzed and solved separately and (at least to some extent) independently. Second, an ontology works as a set of principles behind the construction of knowledge representations. Third, ontologies can provide a bridge to legal theory through a common question — namely, “what is law?”. Fourth, legal ontological categories have a direct correlation with the structure of legal arguments, and as such provide a basic view on legal reasoning. In the particular case of ON-LINE, this translates into a number of advances in comparison with most present information retrieval systems. For example, ON-LINE represents not only the concepts of the domain (traffic, copyright, etc.) the law is about, but of law (legislation) itself. We will discuss some other advantages in the course of the article.

2.2.1 The Concept of Ontology

A key concept in understanding ontologies is *conceptualization*. A conceptualization is a description of the objects, concepts and other entities which exist in a certain part of the world, as well as the relations between them. By identifying these terms and relations, one commits oneself to some domain of discourse [10]. AI has borrowed from philosophy the term *ontology* to mean a specification of of a conceptualization.¹ An ontology can be a conceptualization of an abstract concept, such as time, action, space, or of a domain. Ontologies of domains can be more or less specialized, ranging from general descriptions of engineering mathematics to elevators.

Ontologies comprise a number of *concepts* which describe classes of entities or relations in the domain. These concepts may be arranged in *typologies* which connect them in class/sub-class relations. The roots of these typologies are *primitive concepts*. The set of concepts in an ontology defines a *vocabulary*. In addition to this vocabulary, an ontology should provide the basic relations between its concepts. These two parts specify how the world is ‘divided’, i.e. which basic types of things are distinguished.

2.2.2 The Functional Ontology of Law

The functional ontology of law, which is used as a representational foundation for ON-LINE, has been described in [15, Chapter 4]. Its description in the formal language

¹In philosophy, an ontology is a systematic account of existence. What is called ontology in AI is perhaps closer to what has been called in philosophy an *ontological theory*: “a theory that contains and interrelates ontological categories, or generic concepts representing components or features in the world” [6, pag. 11].

Ontologia [10] is available in the World-Wide Web at the address <http://iri.jur.uva.nl/people/andre/follow.lisp>. The functional ontology of law has a somewhat different character than many AI ontologies. Some people stress the vocabulary character of ontologies and envisage them as a 'dictionary' of the terms used in a specific domain or application, frequently seeing ontologies as only a taxonomy of the domain terms. So, for instance, an ontology of medicine would contain a taxonomy of diseases. In contrast, the functional ontology distinguishes *primitive categories* of legal knowledge in general. It does not contain application-specific legal concepts. It is more like an ontology of medicine which defines major abstract medical concepts such as disease, human bodily functions, organs, etc, which define what medicine is about, but not what specific types of diseases, etc, are known. Using another analogy, the functional ontology of law tries to be a grammar, not a dictionary or thesaurus. That is why it can be used as a foundation for knowledge representation.

The functional ontology of law has foundations in legal theoretical concepts from theorists and philosophers like Hart, Kelsen, Bentham and Lindahl. It is outside the scope of this paper to describe the ontology in full detail; the interested reader is referred to [15, Chapter 4]. Below, we will briefly characterize the main primitive categories of legal knowledge distinguished in the ontology:

- *normative knowledge* consists of all knowledge, which role it is to specify which behaviours in the world are to be allowed, and which disallowed. It comprises basically norms that address behaviour directly (Hart's 'primary rules'). A formalization of this knowledge category, including its classical deontic concepts, is proposed in [16].
- *world knowledge* comprises the knowledge about the world which is regulated by the law. The assumption is that the law contains a model of the sub-part of the society which it intends to regulate [3]. This "world model" specifies which types of agents and objects are considered, how these are related in terms of types/sub-types (i.e. taxonomically), which behaviours may occur, etc. This type of knowledge is usually recognized as a terminological one, because it is specified through the definition of terms. Further, world knowledge comprises *causal knowledge*, which specifies which behaviors is to be considered a cause of certain events or other behavior. World knowledge is the interface between purely commonsense knowledge and legal knowledge, since some terms or causal connections specified in the law are assumed to be interpreted in terms of other commonsense terms or connections.
- *responsibility knowledge* is the knowledge used to assign responsibility to agents when a violation of a norm occurs. It includes both limitations of responsibility (e.g. limited liability of manufacturers) and the assignment of responsibility without a clear causal connection (e.g. parental responsibility for civil damages of their children). Responsibility knowledge is built upon (but is distinct of) causal knowledge. We have already sketched the formalization of this connection between causality and responsibility in the law in [13].
- *reactive knowledge* specifies the reactions the legal system can or should take towards an agent which is responsible for a certain situation. This is usually a punishment (as in the penal code of many countries), but it also comprises rewards such as tax rebates.

- *meta-legal knowledge* is legal knowledge about legal knowledge (Hart's 'secondary rules'). It includes for instance data about elements of legal knowledge (e.g. the enacting authority or issuing date), knowledge about the solution of conflicts between norms, and knowledge about the validity of legal knowledge.
- *creative knowledge* is a small knowledge category which contains the laws which create new legal entities, e.g. an authority, department or a school.

The categories discussed above are but the primitive ones. Other important categories of legal knowledge, such as rights and legal procedures, are defined in terms of the primitive ones, either as compositions or variations in form (see [15] for details).

2.3 Text and concepts: dual representation of legal information

One of the ways to integrate information retrieval and legal analysis is a dual representation of legal information. ON-LINE stores this both as text (in a *text base*) and as knowledge (in a *knowledge-base*). The knowledge base is a model of the text base. That is, the knowledge base and the text base represent the same thing (a law), but expressed in different languages. The text base provides the law in one or more natural languages,² while in a knowledge base it is expressed in a formal knowledge representation language based on the ontology. Of course the status of these two ways of expressing is different. The textual sources are the standard, as their legal status (in general) will imply. However, legal sources may have many interpretations. Moreover, representing knowledge involves always some way of interpretation. Multiple (relevant) interpretations of a legal text can and should be represented. The advantage of a knowledge representation is that it can describe ambiguities unambiguously, provided one has traced the ambiguity. As a consequence, there is no distinction in ON-LINE in the modelling support of the tools to model cases (in general performed by a legal practitioner), or to model a regulation, which may be part of a knowledge acquisition activity to load ON-LINE with a specific domain.

Another important aspect of ON-LINE is that the coupling between the text and knowledge bases is not restricted to formal legal sources, but applies to *legal documents* in general: precedents, jurisprudence, commentaries, cases (actual and hypothetical), etc. The idea is that the storage of all these documents as first class elements provides that the information server works not only as a static database of legal sources but also as a dynamic, group-specific library of both text and knowledge elements which can help improving the usability and reusability of legal information. In addition, there is a homogeneous treatment of legal information. For example, the user can employ a textual description of a case to build a knowledge version of it, which can be used by the legal analysis modules (see section 3).

Text are stored as objects. Objects may be atomic (e.g. an article, or a simple proposition in an article), or composite (e.g. a regulation). Composite objects may embed other (composite) objects, thus forming part-of hierarchies. The text-objects ("sources") are also (system) typed. Some of the built-in types are *official sources* (which can be: statutes, court decisions and contracts), *cases* and *jurisprudence*. In

²Legal sources may occur in a multi-lingual form, e.g. EG legal directives.

addition, the user may define other types, e.g. by specializing one of the built-in types.

2.4 Some Details of the Knowledge Representation

As with the text repository, ON-LINE represents (internally) the elements of the knowledge base as structured objects. Such representation has many advantages from an architectural point of view, since it provides a homogeneous method of storing and working with these elements. Further, it is meant to be accessed through the browsing and editing tools in the Legal Information Modelling Toolkit, which should enable semi-automated manipulation of these objects without being concerned about their internal structure: the tools should hide this internal structure from the end user. On the other hand, for the advanced user, a more explicit and readable representation should be available, if only to reuse domain knowledge specified by other groups and with other tools. To obtain a middle ground which allows both efficiency and reusability/readability, ON-LINE employs an 'executable language' in the style of Ontolingua [9] or LOOM [12]. The idea is that, instead of providing only a declarative language, the architecture supports a number of Common Lisp forms which, when executed, create or manipulate the internal representation of the knowledge elements as objects. Therefore, each of these forms (implemented as lisp functions) has a double role: it is a declarative representation of the knowledge, and it can be immediately used to create the corresponding knowledge objects. There are functions to create all three types of objects: knowledge, text and architectural objects, i.e. contexts and modelling links (see section 2.5). This Ontolingua-like language facilitates reusability because it has a stable external interface format to exchange knowledge. The language is not restricted to legislation, but to any knowledge element, e.g. the representation of a case. In addition, similar forms can be used to represent and to refer to the architectural objects.

The following is a (toy) example of forms in this language.

```
(defcontext new-context :theory (fol))

(in-context 'new-context)

(deflegconcept place
  :is-primitive object)

(deflegconcept classroom
  :is-primitive place)

(deflegconcept smokes)

(deflegrelation in.place
  :domain person
  :range place)

(defprimnorm n2
  :deontic-concept F
  :generic-case ((?x ?y)
                 (smokes ?x)
                 (in.place ?x ?y)
                 (classroom ?y)))

(defcase case1 (John BuildingB RoomB2)
  ((person John)
```

```
(building BuildingB)
(classroom RoomB2)
(smokes John)
(in.place John RoomB2)
(within.place RoomB2 BuildingB))
(new-context fol))
```

This example shows several knowledge and architectural elements. The first two forms specify a new context new-context. Then there are three forms defining some simple concepts and relations. These elements are part of the world knowledge category. The primary norm n2 is part of the normative knowledge category. Case1 is described as a set of assertions about instances in the world (John, BuildingB) using the concepts and relations defined as part of the world knowledge.³ ON-LINE comprises facilities for conceptual search which may be employed directly by the user to browse or search elements in the knowledge base. For example, one can retrieve all superconcepts and subconcepts of place using the following queries (note the internal 'root concept' legal-thing):

```
(retrieve (?x) (superconcepts place ?x))

-> ((|C|LEGAL-THING) (|C|NON-AGENT) (|C|OBJECT))

(retrieve (?x) (subconcepts place ?x))

-> ((|C|BUILDING) (|C|CLASSROOM))
```

2.5 Linking: information management support

To have a legal knowledge base which is an abstract model of the legal text is not very helpful if the correspondence itself is not explicitly stored. It is necessary to keep track of the modelling process in order to be able to reason about it, if only to retract falsified beliefs or interpretations ("truth maintenance"). Furthermore, there are also reasons of a more practical nature, such as allowing a multidimensional way of searching and updating, i.e. for hypertext documentation. Therefore, ON-LINE supports the storage of modelling steps as first class components called *modelling links*. There are two types of modelling links in ON-LINE (see figure 1):

definitional links keep track of where the information came from that has been used for modelling a certain element of the knowledge base. For instance, a definitional link keeps track of the fact that the definition of the concept *software* was based on the use of the term 'software' in Article 2 of the Brazilian Software Regulations. Because this implies an interpretation process, this link has additional slots, such as *author*, or *source*, and optionally a *rationale*. This rationale could be turned into a full knowledge item, but the default is a simple textual note. In addition, there are user defined labels. Labels may be used for reasoning or otherwise to restrict search.

referential links keep track of multiple references to the same defined concept. These links have no labels. For instance, the concept *intellectualProperty* is referenced

³Since ON-LINE relies on LOOM for managing and creating concepts and relations (see below), the syntax of some forms is very similar to the syntax of some LOOM commands.

in Article 1 of the Brazilian Software Law, but it is defined elsewhere (actually, in another law). Referential links can be used to cross-reference larger pieces of text or knowledge. For instance, there is in figure 1 a reference between Article 3 and the whole definition of the concept *computerProgram*.

Modelling links are a way to obtain the same advantages of isomorphism [1] without some of its restrictions. While there is not necessarily (and not likely) a one-to-one correspondence between elements in the text and in the knowledge base, the correspondences between the elements can be traced if necessary.

3 The Structure of ON-LINE

ON-LINE comprehends three basic modules (see figure 2): (i) a *Legal Information Server* which retrieves information in both textual and conceptual (knowledge) form, and stores information in the text form; a *Legal Information Modelling Toolkit* which comprehends a number of tools to aid in the translation of legal texts into the knowledge representation language (based on the functional ontology); and a *Legal Analysis Environment* which contains a variety of reasoning sub-modules, which use the knowledge base. The modules of ON-LINE communicate by their common access to the two information repositories. These modules will be further described below.

All modules share a *scope restriction mechanism*. This mechanism restricts searching or browsing to a specific subset of the information: a *context*. Scope restrictions can be defined either in the text base or in the knowledge base. They may be used for pragmatic reasons (e.g. making search more efficient), but also for more principled reasons, e.g. to specify the semantical contexts for the interpretation of concepts. There are three basic methods for restricting the scope in ON-LINE, based on (i) the type of the information (e.g. its ontological category), (ii) the contents (e.g. information about a specific concept such as 'software'), or (iii) modelling links (i.e. elements referred by some links).

3.1 Legal Information Server

The services provided by this module are directly related to the Legal Information textual base. Its design does not differ very much from the usual structure of legal information databases, with modules for adding, deleting and modifying items in the text database, and version control. The main extra of this Legal Information Server is that it is also able to search for information using the knowledge base. One can search for the term 'software' (in the text base) or for the concept 'software' (in the knowledge base). That is, ON-LINE enables *conceptual retrieval* (cf. [7]). Additional features of this conceptual retrieval are (i) the search is made not only to world concepts, but using elements in the whole ontology; and (ii) retrieval is not only by queries, but also based on the description (in knowledge terms) of a case, which allows various modes of intelligent consultation.

3.2 Legal Information Modelling Toolkit

The Legal Information Modelling Toolkit contains a number of tools for modelling legal information. These tools are *browsers* of various types, which present different views of the elements in the text and knowledge base, and *editors* for

adding and deleting material from the knowledge base.⁴ The Legal Information Modelling Toolkit has also the tools for creating and managing the modelling links.

3.3 Legal Analysis Environment

This environment contains an extensible number of modules which execute legal reasoning tasks. Presently, two modules are supported:

- *Legal assessment* is a task which assesses a *case description* (a description of relevant facts in the world) based on a body of legal knowledge. The task can be executed in two modes. In the *case analysis* mode, a specific case already modelled and stored in the system is matched against a knowledge base (or a subset of it defined with the context definition tools). In the *goal-oriented* mode the task is 'inverted' in the sense that instead of looking for all possible consequences of a certain case, the user wants to look for conditions which are sufficient to warrant a certain (desired) conclusion. There are several specific goals which may be used (e.g. what is needed for concept X to apply to individual Y?).
- *Legal planning* is a task in which a *plan* (list of actions to be taken or procedures to be followed) is generated to achieve a certain (legal) *goal* (described in terms of legal concepts and norms which apply in the final state), departing from an *initial state*.

An important question is how these modules are integrated with the rest of the architecture. A complete discussion of the issue is outside the scope of this paper; the interested reader can find the complete discussion in [15]. Briefly, the solution to this problem is based on the use of *knowledge-level models* adapted from the CommonKADS Library [4] to specify the tasks implemented by the reasoning modules. These models describe the tasks using (functional) inference steps, and roles that domain knowledge plays in problem solving. Figure 3 pictures a graphical representation of the model of normative assessment employed in ON-LINE. On the left side, we have the model extracted from the CommonKADS Library; on the right side, the 'adapted' model for normative assessment. The adaptation corresponds to a mapping between the problem-specific roles to law-specific roles, i.e. we must verify which types of legal knowledge play these roles (for more details on the methods and the adaptation see [17], or [15, Chapter 7]). For instance, one of the steps in assessment in general is to *abstract* the description of a case by translating some terms into more abstract terms (commonly those used in a set of norms at hand). To make this step, knowledge is needed to play the role of *system description*, e.g. a description of the terms used to describe the case (system) organized in an abstraction hierarchy. The integration relies on the fact that the knowledge which fills these (static) problem-solving roles can be defined as one or more of the knowledge categories of the ontology. In this case, for example, the knowledge which plays the role of *system description* is precisely the knowledge belonging to one of the categories of the functional ontology of law, namely *world knowledge*. We find a similar mapping for the model selected for legal planning. This correspondence implies that the integration between legal information modelling and analysis is basically straightforward, since knowledge in ON-LINE's knowledge base is already structured in

⁴The editing of the text base is performed by tools in the Legal Information Server

the way it is needed to perform the tasks in the Legal Analysis Environment. Indeed, this close correspondence between the problem solving roles in the models for assessment (and planning) as described in the CommonKADS Library, and the categories of knowledge we use in the knowledge representation of ON-LINE, gives some evidence for the validity of our functional ontology for legal reasoning. We are not claiming, however, that all knowledge necessary to solve legal problems is in the knowledge base. In planning, for example, additional knowledge is needed to specify the structure of the plan, for example whether it is a total or partial order of actions, whether or not there are resources associated to the plan steps, and so forth. However, this additional knowledge is not legal knowledge strictly speaking, but is instead part of the specification of the method used to solve the problem. For this reason, it is normally hidden in the specification or implementation of the method, instead of stored in the knowledge base.

4 Some Implementation Details

ON-LINE is a research prototype. It is written in Common Lisp, using also the LOOM representation system [12]. The use of LOOM is a key factor in the implementation, since it provides sophisticated facilities for concept definition and search for free. Most ON-LINE facilities rely heavily on the corresponding LOOM facilities. For example, all the elements in the knowledge and text repositories are LOOM objects. By fully exploiting these facilities, the development of the prototype became feasible as part of a PhD thesis project. However, for some of the services (particularly the ones that deal with text objects) the use of LOOM is not particularly efficient, and they will have to be reimplemented directly in Common Lisp to arrive at an operational system. It is important to note that only for the world knowledge LOOM is used as a knowledge representation system as well. All other categories of legal knowledge have their own languages which are not (directly) supported by LOOM. That the knowledge representation language of LOOM is used for world knowledge is not a coincidence. In [15, Chapter 6] we show that terminological knowledge representation, a type of concept-based representation of the family of semantic nets and conceptual graphs, is the most adequate type of formalism to represent legal world knowledge.

The current version of the prototype system does not contain all facilities described. We have concentrated on the knowledge representation and conceptual search services, and the assessment module. The planning service of the Legal Analysis Environment and many graphical browsing and editing facilities will be part of the second version of the implementation, planned by late 1995. We are currently testing the system for some types of law, but full results are still to be obtained. For the testing phase, we use legal domains such as the agriculture section of the GATT treaty, and parts of other legislation including the Dutch and the Brazilian Copyright Laws and the Brazilian Traffic Law.

5 Problems and Limitations

The proposed architecture has a number of problems and limitations which can be already recognized. The main problem, and one which implies a restriction on the scope of the architecture, is that the architecture is modelling-intensive. The modelling effort necessary for representing a single regulation (e.g. a traffic law, a copyright law, a treaty) is rather large. Furthermore, even if some help is

provided by browsers and editing tools, most of this work needs to be done or at least checked by a specialist. It may be necessary to read books or commentaries about the specific regulation being modelled in order to come up with a well-structured model. Finally, some compromise must always be worked out with respect to how much of common-sense knowledge will be included in the knowledge base. All these factors, which are mostly the usual ones in knowledge acquisition anyway, preclude or make it very difficult any attempt to use the architecture for modelling very large amounts of legislation, e.g. as in CD-ROMS containing legislation and/or precedence law of full divisions of legal systems (e.g. penal law). The scope of the architecture is restricted to model in detail limited amounts of legislation. Nevertheless, it must be realized that these problems are a direct counterpart of the strengths of the architecture, particularly its capacity to retrieve information and reason with the integrated knowledge and text repositories. A compromise is always necessary between quantity of information available and the quality of its model. Moreover, available technology for knowledge acquisition from texts is far from providing an alternative. In summary, there is no shortcut available: if a representation of a regulation is wanted, painstaking and extensive modelling work must be done.

Another issue which is not yet covered, is the broadness of the application scope of the architecture — and, in the end, of the ontology itself. While it is expected that the ontology is able to represent adequately legal knowledge in several types of legislation and legal systems, this issue was not yet tested in practice. Further, in order to model legal knowledge with ON-LINE the user or knowledge engineer must understand the ontology and analyse the legislation in the terms it defines. This may or may not cause problems, depending on the specific knowledge engineer or user in question, but some of the comments received so far point to a mismatch between the way these experts reason and the framework provided by the ontology: the ontology, they say, is not 'intuitive'. This problem can be corrected in part by training the knowledge engineers, but a more complete solution would require an in-depth study of what these 'intuitions' exactly are, and the construction of interfaces which present the knowledge in a form in accordance with these intuitions.

6 Comparison with Related Work

The architecture proposed here seems to be innovative in proposing a representation of a broad range of types of legal knowledge integrated to the representation of the text, and in employing an ontology to drive legal information modelling. However, there are many legal information systems and architectures which incorporate conceptual models of the underlying domain, i.e. of the world (and mainly definitional) knowledge. This has been called *conceptual information retrieval*. Some examples are the systems by [11] and [7]. These systems usually employ some sort of conceptual representation such as semantic networks and conceptual graphs which are very similar to the terminological knowledge representation used here. This can be viewed as a convergent view that this type of knowledge should be represented concept based. The use of arguments in the system by [7] is an interesting idea which may be incorporated in the future to the architecture.

Another architecture of a comparable (broad) scope and a similar 'toolbox' strategy is DataLex [8]. Similar features are: the use of integrated text and knowledge repositories;

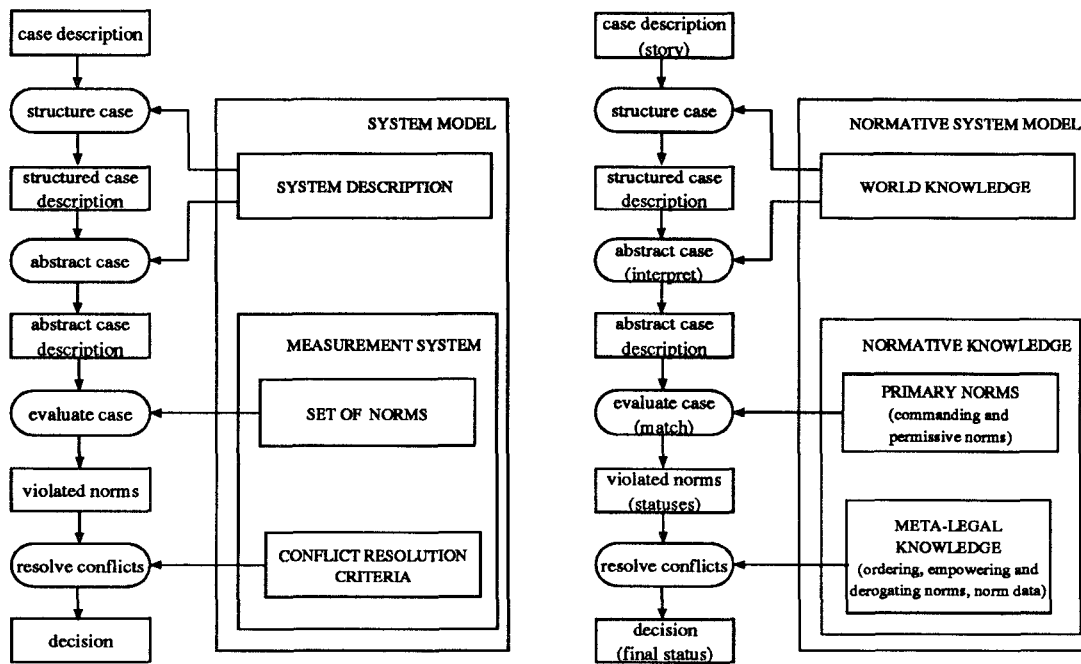


Figure 3: Graphical representation of the problem solving method used for normative assessment. This is a notation used for representing what is called in CommonKADS a *function structure*. Only the main functional steps and their dependencies are represented; control is represented elsewhere.

the use of cross referencing (in ON-LINE with the modelling links and in DataLex with a hypertext tool); and the storage of all types of legal information (legislation, commentaries, cases, etc.). The most important difference, however, is ON-LINE's use of an ontology as a basis for knowledge representation. In contrast, DataLex uses a simple rule-based formalism. The implications of this are far-reaching, and comprehend for instance the integration of the reasoning modules as described above. ON-LINE also improves some of DataLex' services, for example by providing a broader cross referencing mechanism (not limited to the legislation text) and a scope restriction mechanism.

7 Conclusions

The ON-LINE architecture proposes an ambitious view on legal information serving, where legal information is stored and manipulated both in its original form (text) and modelled as legal knowledge. Both forms are linked. The basis for the knowledge representations is a functional ontology of law [15, Chapter 4] as well as the specific formalisms proposed for normative [16] and world knowledge [15]. It also contains reasoning modules based on the models for legal assessment and legal planning [14] as well as supporting tools for legal modelling and design. The main features of ON-LINE are the following:

- *Integrated storage and representation of text and knowledge.* ON-LINE stores legal information in both text format and modelled as knowledge, forming two repositories, resp. a *legal information database* (or text base) and a *legal knowledge base*. The mapping is supported by *modelling links* which store references and modelling information about the connections between items

in both repositories. Modelling links are first class objects which can be used for intelligent retrieval and reasoning purposes. The architecture provides graphical editors and browsers for accessing both repositories.

- *Emphasis on legal modelling.* The architecture provides several facilities to help or partially automate the process of modelling legal information (text) as knowledge, including the storage of multiple interpretations.
- *Documents and sources are structured objects.* The text base is an object-oriented database of documents and legal sources. The classes of documents are arranged in a hierarchy which can be extended by the user. The text base supports the storage of any textual source used in legal practice.
- *Ontology-based knowledge representation.* The knowledge base is structured based on the proposed ontology [15, Chapter 4] — that is, the main classes of constructs are the ones proposed in the ontology. The knowledge representation of the legal information is used for both information retrieval and reasoning.
- *Information retrieval services.* There are a number of different information and knowledge retrieval services, including conceptual search, goal-oriented search and a simple full text boolean search as a fallback. The retrieval process is aided by the specification of scope restrictions, i.e. subsets of the text or knowledge bases specified by the user.
- *Reasoning services* The architecture provides a number of reasoning services centered on helping in the legal analysis of cases. These services are built based on models of typical legal tasks.

Acknowledgements André Valente was supported by the grant number 203182/90.1 of Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil.

References

- [1] T. Bench-Capon and F. Coenen. Exploiting isomorphism: Development of a KBS to support british coal insurance claims. In *Proc. of the Third International Conference on AI and Law*, pages 62–68, Oxford, 1991.
- [2] J. Bing. Designing text retrieval systems for ‘conceptual searching’. In *Proc. of the First International Conference on AI and Law*, pages 43–51, Boston, May 1987.
- [3] J. Breuker and N. den Haan. Separating world and regulation knowledge: Where is the logic? In M. Sergot, editor, *Proc. of the Third International Conference on AI and Law*, pages 41–51, 1991.
- [4] J. Breuker and W. van de Velde, editors. *CommonKADS Library for Expertise Modelling*. IOS Press, Amsterdam, 1994.
- [5] Joost Breuker. Modelling Artificial Legal Reasoning. In G. Boy, N. Aussenac, B. Gaines, and J. Boose, editors, *Proceedings of the European workshop on Knowledge Acquisition*, pages 34 – 45, Berlin, 1993. Springer Verlag.
- [6] M. Bunge. *Ontology I: the Furniture of the World*, volume 3 of *Treatise on Basic Philosophy*. D. Reidel, Dordrecht, Holland, 1977.
- [7] J. Dick. Reasoning with portions of precedents. In *Proc. of the Third International Conference on AI and Law*, pages 244–252, Oxford, 1991.
- [8] G. Greenleaf, A. Mowbray, and A. Tyree. Expert systems in law: the DATALEX project. In *Proc. of the First International Conference on AI and Law*, pages 9–17, Boston, 1987.
- [9] T. Gruber. Ontolingua: A mechanism to support portable ontologies. version 3.0. Technical report, Knowledge Systems Laboratory, Stanford University, California, 1992.
- [10] T. Gruber. Toward principles for the design of ontologies used for knowledge sharing. In N. Guarino and R. Poli, editors, *Formal Ontology in Conceptual Analysis and Knowledge Representation*. Kluwer, 1994.
- [11] C. Hafner. Representation of knowledge in a legal information retrieval system. In R. Oddy, S. Robertson, C. van Rijsbergen, and P. Williams, editors, *Information Retrieval Research*. Butterworth, London, 1981.
- [12] R. MacGregor. Inside the LOOM classifier. *SIGART Bulletin*, 2(3):70–76, June 1991.
- [13] A. Valente. Preliminary steps in causal legal reasoning. In A. Cohn, M. Lee, C. Preist, and C. Price, editors, *Proc. of the AISB Workshop on Qualitative and Causal Reasoning*, Birmingham, March 1993.
- [14] A. Valente. Ontologies: the missing link between legal theory and AI and law. In *Legal Knowledge Based Systems and Legal Theory*. Koninklijke Vermande, 1994.
- [15] A. Valente. *Legal Knowledge Engineering: A Modelling Approach*. IOS Press, Amsterdam, 1995.
- [16] A. Valente and J. Breuker. A commonsense formalization of normative systems. In Joost Breuker, editor, *Proc. of the ECAI'94 Workshop on Artificial Normative Reasoning*, pages 56–67, Amsterdam, August 1994.
- [17] A. Valente and J. Breuker. Making ends meet: Conceptual models and ontologies in legal problem solving. In *Proc. of the XI Brazilian AI Symposium (SBIA '94)*, pages 1–15, Fortaleza, Brazil, 1994.