Legal Data Modeling:
The Prohibited Transaction Exemption Analyst

Keith Bellairs
Computer Law Systems, Inc.
11000 West 78th Street
Eden Prairie, Minnesota 55344
U.S.A.

ABSTRACT

This paper addresses issues in the design of legal expert systems. The emphasis is on the nature of the underlying knowledge that is incorporated in the knowledge base of a legal expert system. Examples of different approaches are discussed. The "legal data modeling" approach, used by the author in the construction of an expert system, is described. Legal data modeling emphasizes the construction of a conceptual model of situations in which legal problem solving occurs.

INTRODUCTION

The selection of a legal task domain does not in itself constrain an expert system builder to one particular method for structuring that system's knowledge base. A structure that seems natural for certain legal tasks or domains may be less satisfactory in others. This paper identifies some of the factors that affect such choices.

There are alternative strategies for designing a legal expert system. The method proposed here is based upon the design of a conceptual model of the important objects in the problem domain. This model is called the "legal data model." The structure of the legal data model has implications for the overall capabilities of a legal expert system.

To support effective legal problem solving the knowledge base of a legal expert system should be based on a conceptual model of the objects and events that comprise legal problem situations. For example, to deal with the problems related to a pension fund's transactions, we need knowledge that reflects information about the pension fund and its transactions. In addition we need to identify the legal consequences of this information.

Legal data modeling is analogous to the practice of semantic or conceptual modeling in database design. [Brodie, 1984.] In knowledge engineering the conceptual modeling approach can help to focus the knowledge engineer's attention on the kinds of knowledge needed to solve a legal task.

The organization of this paper is as follows. First, some of the principles of conceptual modeling are discussed. A graphic representation, called Logical Data Structures (LDS for short), will be introduced. The importance of "representational independence" in conceptual modeling will be discussed. The next section of the paper will analyze the conceptual models that have been used in some legal expert systems. The legal data modeling approach will then be introduced. In this context the Prohibited Transaction Exemption (PTE) Analyst will be discussed. The PTE Analyst is an expert system which was designed using the legal data modeling concepts.

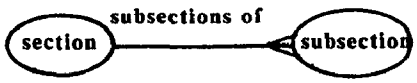
CONCEPTUAL MODELING

The goal of conceptual modeling is

> to develop descriptions of a world/enterprise/slice of reality which correspond directly and naturally to our own conceptualizations of the object of these descriptions. [Mylopoulos, 1984. p. 11-12. emphasis in original]

Some of the most interesting techniques for conceptual modeling have relied heavily on the use of diagrams: for example, Chen's entity-relationship approach [Chen, 1976] and Sowa's conceptual diagrams [Sowa, 1984]. In this paper we use a tool called Logical Data Structures (LDS). [Carlis]

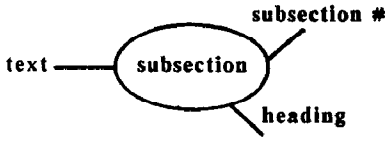
The following is an LDS diagram:



This model has two types of objects: sections and subsections. These objects are related by a relationship which has two "relationship descriptors." A section can have more than one subsection, as indicated by the "chicken-foot" (≺) on the relationship arc. The diagram can be read as follows:

> A section has zero or more sub-sections. Each subsection is a subsection of one and only one section.

An LDS diagram also represents information about the attributes of an object:



This diagram says that each subsection has certain attributes and describes the type of each attribute. An attribute is single-valued. Multiple-valued attributes are represented by relationships between two objects.

The LDS diagram is simple because there are few constructs to learn; it is also general -- any collection of objects, attributes, and relationships can be represented. But the LDS makes no attempt to specify a data structure suitable for the concepts described. One of the motivating features of database technology is data independence. [Chen, p. 9; Date, p.12] Systems like LDS were developed with a view to redirecting attention away from details of how to represent data in a file toward what data is needed by a class of applications. [Carlis] This approach helps to identify data that is relevant to the purposes of the system. It focuses on relationships between conceptual objects and their meaning to the users of the system. It also encourages a system designer to make use of generalization, aggregation, and abstraction to find the common structure of information in a system. [Borgida, 1984, p. 90]

In an AI problem we must be concerned with representational independence: a model of knowledge needed for a problem which is independent of the different representational mechanisms that may be used to implement the knowledge-base. Examples of such mechanisms include production rules, frames, semantic nets, logic programming, and so on. There are serious issues to be considered in the choice of a representational mechanism for implementing a knowledge-based system. For a discussion of some of these issues see [Israel, 1984]. This paper does not attempt to deal with issues of representation.

The issue of concern in this paper is, what knowledge is represented in the knowledge base. What kinds of knowledge can be brought to bear in a legal problem solving task? The following list gives a few examples:

1. Statutes. A legal expert system can attempt directly to encode a set of statutes. In an extreme case such a knowledge base may be structured on the basis of section and subsection of the statute in question.

2. Legal concepts. Rather than modeling a statute, a system builder may analyze the meaning of some legal rules and create a formalization of those rules. This differs from statutory knowledge in that issues of ambiguity and inconsistency must be resolved in order to achieve a meaningful hierarchy of concepts.

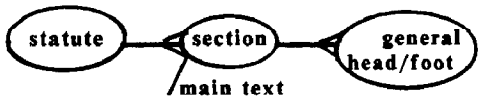
3. Situational knowledge. This is knowledge that an expert might use to classify objects or events in problem situations. This is the kind of knowledge emphasized in legal data modeling. In addition to formal legal knowledge, this kind of knowledge can include knowledge about non-legal domains and even common-sense knowledge.

4. Knowledge of precedents. The actual facts and results in legal cases can also be represented in a knowledge base.

Choices at this level have direct consequences concerning the capability of the expert system. The power of a knowledge based system lies in its ability to make inferences from the knowledge it possesses. A system with situational knowledge may be able to make classifications and draw conclusions from partial knowledge. In contrast, a "pure" conceptualization of a legal domain might be incapable of drawing any conclusions if there are missing pieces of data.

EXAMPLE CONCEPTUAL MODELS

Statutory models. In one reported project the "entity-relationship" data modeling concepts were explicitly used. The goal was to model the structure of United Kingdom statutes. [Heather, 1985] This knowledge was used to build an expert subsystem used for information retrieval which can "read and 'understand' the structure of a British Act of Parliament without human intervention." [p. 16] The structure of a statute is modeled more or less as follows:



The model includes information on matters such as type face and indentation, so that it can be used to format statutory text as well as support analysis of it.

Statutory knowledge can be important in some contexts. For example, the Internal Revenue Code is a Federal statute which is very self-conscious of its structure. It is not unusual to find provisions of the Code such as "other provisions of this subpart notwithstanding" or "with respect to organizations subject to the rules of subchapter C." Clearly, such provisions cannot be understood without detailed knowledge of the structure of the law in its various chapters and sections.

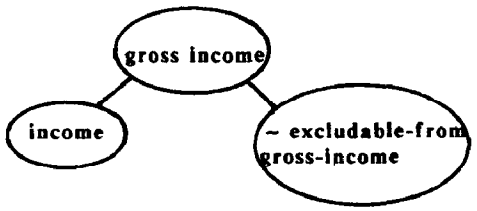
On the other hand, knowledge of statutory structure can be of fairly limited value in many cases. In the American legal system large, self-referential codifications of law are still the exception rather than the rule. Many statutes stand more or less alone. So a knowledge base which focuses primarily on the structure of the statute in question may be paying an inordinate amount of attention to a minor factor.

"Legal concept" models. The law uses "rules" as a fundamental unit of argument. It is therefore natural to attempt to model legal knowledge in terms of the concepts embodied in those rules. An early, hypothetical legal AI system demonstrated this approach with a model of tort law. [Meldman] A more recent example is the system designed by MacRae for problem solving in Federal income tax law. [MacRae] MacRae uses Prolog to model production rules "describing the income tax law." (p. 195) An example is

```
rule('Reg.Sec. 1.61-1(1)',
([Item,is,gross,income,of,Taxpayer]
if [Item,is,income,of,Taxpayer] or
~[Item,is,excludable,by,law,from,
gross,income,of,Taxpayer])).
```

Note that a concept is defined in terms of other concepts. The LDS of such a knowledge base might be drawn as follows:

In the actual rule shown, the concept and subconcept types are instantiated as follows:



The British Nationality Act program discussed by Sergot et al. is another outstanding example of a legal expert system based on knowledge that is an axiomatization of a set of concepts derived from a statutory rule base. [Sergot] The authors emphasize both "the resemblance of the knowledge base to the original form of the legislation" (p. 371) and the design of "an axiomatic theory that represents the legislation." (p. 376) Advantages noted by the authors include the minimal knowledge acquisition problems, since expert heuristic knowledge was not needed, and the value of logic as a comprehensible language for communication of information in a knowledge base.

A disadvantage of the "legal concept" approach is that no attempt is made to supply heuristic knowledge. Hence the analysis of a statute by the expert system will be exhaustive. This has two serious drawbacks: (1) the order of data requests in a dialog with the user will not necessarily seem relevant to the task and (2) the lack of heuristics may lead to search paths that are too long. This occurs because the order of inquiry is related only to the way in rules have been written and because such a system relies on depth-first exhaustive search. The larger and more complete such a system is, the greater these problems are because it will inquire exhaustively with respect to every set of facts which might constitute an exception to a rule. For reasons which an expert will understand, virtually none of these possible exceptions would be relevant in a given case. Note that MacRae has proposed a method of permitting the user to supply heuristic control knowledge for the use of such databases. (MacRae)

Another disadvantage is that a "naive" knowledge engineer may make unwitting interpretations in attempting to create a simple map of the logic of a statute. [Cf. Allen.] Any experienced attorney has reason to be skeptical of claims like the following:

> In fact, we found fewer such examples [of ambiguity] than we originally expected. In practice, where imprecision or ambiguity did exist, it was usually possible to identify the intended interpretation with little difficulty. [Sergot, 1986. p. 371.]

Such interpretation is the stuff of which litigation is made. An attorney would have reason to be very nervous about a knowledge base in which the knowledge engineer announced that he had worked out the "obvious" interpretation of what the framers intended by the law in question.

Based on these observations we would expect that the direct modeling of legal concepts can be effective in areas where there is a small set of rules, the concepts are reasonably unambiguous, and the objective of the system is a complete

analysis of the logic of a system of such concepts. Note also that "legal concepts" are a kind of knowledge that will be represented in all legal expert systems. The issue here is the extent to which those concepts are the focus of analysis in the structuring of a knowledge base.

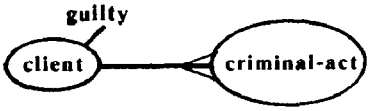
LEGAL DATA MODELING

"Legal data modeling" proceeds from the assumption that legal problems arise in the context of situations in the real world. The task of the attorney is to understand and classify the objects and events that occur in such situations. Statutory rules operate to classify the features of an event and to define the legal consequences of some kinds of situations. In pension law for example, some persons may be classified as "parties in interest" with respect to a transaction with a pension fund, and certain kinds of "transactions" between the fund and parties in interest are prohibited.

Legal data modeling does not replace statutory or conceptual knowledge. But additional knowledge about the significant objects and events in legal problem situations is needed.

An effective legal expert system must capture some aspects of an expert attorney's knowledge about the task domain. Johnson has called the knowledge that characterizes expertise the "operative knowledge." [Johnson; see also Clancey.] This is not necessarily the knowledge that an a priori conceptual analysis of an area of law might suggest. An expert attorney is highly adapted to problems in her area of expertise. This adaptation includes knowledge about ways to characterize facts and a sense of the order of importance of different features of a problem. Thus legal data modeling is oriented toward identifying significant parts of this operative knowledge. This is accomplished, of course, by working with an expert.

In legal data modeling a specific area of the law is selected and goals for an expert system in that area are set. Then an expert describes a typical problem. A knowledge engineer uses this kind of problem description to develop an LDS diagram representing the conceptual objects used in the problem solving process by the expert. The LDS is then used as a tool for communication between the expert and knowledge engineer to refine the initial conception of the problem space. Attributes of the conceptual objects are identified. In some cases the values of attributes can only be determined by asking the user. In other cases the value of an attribute can be inferred from values of other attributes; the basis for this inference may be legal or heuristic rules. As a gross example consider the following:



The value of the "guilty" attribute (true, false, or unknown, for example) can be inferred by using a rule which says

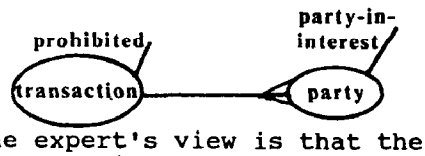
if there exists a criminal-act of client
then guilty = true.

Thus the legal data model assists the system designers in identifying the important legal rules in the problem domain.

The PTE Analyst was designed to be a commercial product for sale to attorneys as a law office productivity tool. Primary goals include ease of use and crispness of reasoning. The dialog with the system must be limited to relevant queries, and the system should not ask stupid questions. We assume that an attorney user will rapidly lose interest in a system that wastes the attorney's time by asking questions that are irrelevant.

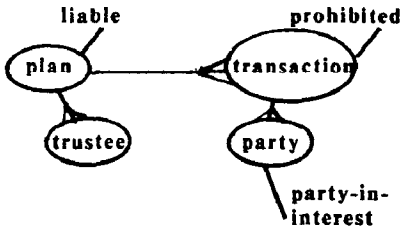
The domain of the PTE Analyst is analysis of employee benefit plan transactions under the rules of the Employee Retirement Income Security Act of 1974 (ERISA), the federal law which regulates the conduct of pension plans. (Employee benefit plans are entities like pension and profit-sharing plans.) Plans are prohibited from entering into transactions which are subject to abuse, but exemptions are provided for some kinds of transactions. In a law office context, an attorney will be called on to prepare a formal written opinion which carefully analyzes the facts and refers to the relevant portions of the law. A goal of the design of the PTE Analyst is to support the research leading to the preparation of a formal opinion concerning pension trust transactions. The analysis of such problems is not a hard problem for an experienced attorney, but it is tedious to pursue all the needed statutory references. The PTE Analyst assists an experienced attorney in locating the needed sections of the statute and provides a check to make sure that relevant issues have been considered. For an inexperienced attorney the research is more difficult because of the need to discover the structure of the statute and the relevant provisions of the law. Since this knowledge, too, is incorporated in PTE Analyst, it is a tool for the novice attorney as well.

The legal data model used in PTEA was developed in conferences between the author and an attorney specializing in pension law. Early discussions led to the focus on the problems presented in analysis of prohibited transactions. The first LDS of the problem was simply the following:



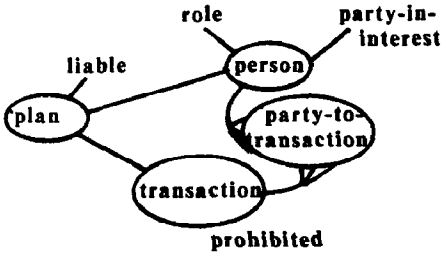
The expert's view is that the transaction is the main problem and that a transaction is frequently prohibited because a party to the transaction is a "party in interest."

During the refinement process it became apparent that the attorney needs to think about his client, in this case the employee benefit plan. Sometimes the determination of whether a transaction is prohibited depends on information about the plan and its trustees as well as the other parties to a transaction. This led to the following LDS:



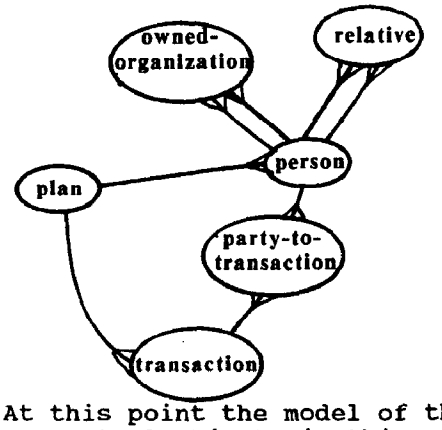
which reflects the high level goal of the user of the system: is the plan liable for having engaged in a prohibited transaction?

At this point the system was well enough defined so that a prototype could be implemented. Simple versions of critical rules could be implemented and the expert could observe the program. It became apparent that information about persons related to the plan as well as persons directly involved in transactions was needed. This problem arises because a party to a transaction may be a party in interest by reason of a relationship to a trustee, owner of the business, union, or other critical party. This led to the following revision of the LDS:



One person can be a party to more than one transaction and that a transaction may have more than one party to it. Thus the conceptual object representing a "party to a transaction" is an association between a person and a transaction.

The final LDS of the legal data model expands on the importance of relationships between persons. Personal kinship relationships, such as marriage, must be considered. Legal relationships, like ownership of a corporation, may also be significant. Thus the final LDS is shaped



At this point the model of the significant conceptual objects in this particular problem area is reasonably complete. The number of attributes of the objects is much larger. (They are not shown in this figure in an attempt to preserve some clarity.)

Rules have been designed by reference to ERISA to provide a means of inference for filling in unknown values of attributes. For example, rule 51 deals with determining if an exemption applies:

> **if loans are available to all participants,and loans do not favor highly paid participants, and loans are specifically provided for in the plan, and reasonable interest and adequate security are required, then a transaction is exempt pursuant to ERISA section 408(b)(1).**

So here we see rules that embody legal concepts but which are organized in terms of entities in the legal data model.

The transformation from the legal data model to the implementation language, Personal Consultant Plus [Texas Instruments, 1986], a microcomputer based version of EMYCIN proceeded more or less in parallel with the development of the legal data model. Using the Personal Consultant Plus terminology, conceptual objects are mapped into frames, attributes into parameters attached to those frames. As might be expected, the implementation language does not exactly implement the concepts of the modeling tool. As a result, it is necessary to define some additional frames in the Personal Consultant plus implementation. Rules are collected into rule groups which are attached to frames.

CONCLUSIONS

Legal data modeling has been used to implement the PTE Analyst. Preliminary examination of the performance of the system supports the following conclusions. First, the dialog between the user and the

system is arranged so that the system's focus of attention is compatible with the users'; from the users' point of view, questions are asked in a logical order. Second, knowledge associated with different conceptual objects can be used to cut off irrelevant lines of inquiry that might occur in a knowledge base which relies on depth-first search.

The data modeling approach is not novel. The design of the MYCIN knowledge base reflects the kind of analysis advocated here. One insight of those pioneers was that physicians solve diagnostic problems by collecting data about patients, bacterial cultures, and organisms, and by applying knowledge closely associated with entities of those types. [Buchanan] Thus the MYCIN project identified kinds of knowledge that are used in a problem domain. Subsequent work in database, and to a lesser extent artificial intelligence, has formalized conceptual modeling methods directed toward identifying and modeling this kind of knowledge. Practical legal expert systems, those directed toward effective problem solving, will benefit from consideration of these developments.

A final comment remains on the future directions of legal expert systems. The reader will have noticed that we have not discussed the use of knowledge of precedents, one of the kinds of legal knowledge mentioned at the beginning of this paper. Legal commentators have frequently mentioned the importance of reasoning by example from precedents in our legal system. [Levi] AI researchers have begun to build a framework for understanding and modeling the processes in analogical reasoning. [Kolodner; Winston.] A tool like the PTE Analyst can provide useful functions for the law office without attaining the true expertise in legal reasoning associated with reasoning from cases without incorporating case based reasoning. But the units of analysis in legal data modeling -- legal situations -- are the units that must be understood in order to make use of knowledge of precedents. So legal data modeling can be seen as a further step in a progression from naive rule-based systems toward systems that achieve true legal expertise.

References

1. Allen,L.E. "Towards a normalized language to clarify the structure of legal discourse," in Deontic Logic, Computational Linguistics and Legal Information Systems, v. II. Martino, A.A. (ed.) North-Holland Publishing, 1982, pp. 349-407.
2. Borgida, A. et al. "Generalization/ specialization as a basis for software specification," in [Brodie, 1984] pp. 87-117.
3. Brodie, M.L. et al. (eds.) On conceptual modeling. Springer-Verlag, New York, 1984.
4. Brodie, M.L. et al. (eds.) On knowledge base management systems. Springer-Verlag, New York, 1986.
5. Buchanan, B.G. and Shortliffe, E.H. (eds.) Rule-based expert systems. Addison-Wesley, Reading, Mass. 1984.
6. Carlis, J.V. Logical data structures. Univ. of Minn. Computer Science Dept. 1985.
7. Chen, P.P.S., "The entity-relationship model: Toward a unified view of data," ACM TODS, v. 1 n.1, March, 1976.
8. Clancey, W.J. "Heuristic Classification," Artificial Intelligence v. 27 n. 3 (Dec. 1985) pp. 289-350.
9. Heather, M.A. "Semiotic orders in law," presented at Second Annual Conference on Law and Technology, University of Houston Law Center, June, 1985.
10. Israel, D.J. and Brachman, R.J. "Some remarks on the semantics of representation languages" in [Brodie, 1984] pp 119-146.
11. Johnson, P.E. et al. "Expertise in trial advocacy," 7 Campbell Law Review 120 (Fall, 1984).
12. Kolodner, J.L. "Towards an under-standing of the role of experience in the evolution from novice to expert," Int.J. Man-Machine Studies v. 19 (1983) pp. 497-518.
13. Levi, E.H. An introduction to Legal reasoning. University of Chicago, 1949.
14. MacRae, C.D. "User control knowledge in a tax consulting system," in Artificial intelligence in economics and management, Pau, L.F. (ed) Elsevier Science Pub, B.V. (North-Holland) 1986, p. 193.
15. Meldman, J.A. "A preliminary study in computer-aided legal analysis," MAC TR-157, MIT (Nov. 1975).
16. Sergot, M.J. et al. "The British Nationality Act as a logic program," Communications of the ACM v. 29 n. 5 (May 1986) pp 370-386.
17. Winston, P.H. "Learning new principles from precedents and exercises," Artificial Intelligence v. 19 (1982) pp. 321-350.