

Acquisition and Representation of Knowledge for the FOREX Expert System

Jürgen Dorn¹ and Franz Mitterböck²

Abstract

In the FIW II project, an expert system for the management of forestries and especially for the rehabilitation of forest ecosystems dominated by Norway spruce shall be developed. From the AI point of view, the domain is characterised by its large number of partly contradictory theories supplied by locally distributed experts.

In this paper, the required knowledge is analysed and knowledge representation formalisms are proposed for this domain. It is tried to cover the whole domain and not only the rehabilitation problem to enhance extensions of the knowledge base to be conducted in the future. The knowledge base shall be accessible from an expert system as well as from an internet browser. The knowledge acquisition shall be supported by the internet browser.

A first prototype of a knowledge base for the bark-beetle problem was developed with PROLOG (Krzizala 1997). A first knowledge acquisition tool on base of HTML were developed in a second master thesis (Ledl 1996). This paper describes these systems, comprise first results and makes some proposal for a further improvement.

1. Introduction

The FIW³-project in Austria is looking back on a decade of research work directed to the forest decline problem. When monocausal hypotheses, contradicting each other, were violently discussed to explain the origin of the forest decline syndrome,

¹ Institut für Informationssysteme, Abteilung für Datenbanken und Expertensysteme, Technische Universität Wien, Paniglgasse 16, A-1040 Wien, email: dorn@dbai.tuwien.ac.at

² Institut für Forstentomologie, Forstpathologie und Forstschutz, Universität für Bodenkultur Hasenauerstraße 38, A-1190 Wien, email: mitter@h415.boku.ac.at

³ FIW is the abbreviation for "Forschungsinitiative gegen das Waldsterben" (research initiative against forest decline)

Austrian scientists tried to approach the problem from a more holistic view. Fundamental assumption was that the expression of acute weakness and disease of trees or forests usually has multicausal origin. Either synergical effects are produced by air pollution when coinciding with episodes of severe climatic stress, or prevailing ecological instability makes forests vulnerable to detrimental effects of stress episodes. All kind of stressors, abiotic and biotic, can be involved in the progress of decline. Thus forest ecosystems in central Europe, which have been exposed continuously to severe human impact during centuries, are expected to be highly susceptible to progressive loading by air pollution and climatic changes. These assumptions and the principle of predisposing and triggering nature of stresses lead to an “integral stress hypothesis” on the origin of the forest decline. Consequently, activities on ecological repair were urgently demanded on two levels:

- reduction of air pollution and
- ecological rehabilitation of forest ecosystems (Führer 1994).

The research programme FIW II is directed to the field of ecological rehabilitation of forest ecosystems, dominated by Norway spruce. It aims to a methodology for deducing concrete recommendations to foresters, how to restore and manage forest ecosystems, thus providing its sustainability. These scientific efforts are made, in order to develop well founded procedures, which allow the establishment of concepts for forest rehabilitation, adapted to the local situation. Investigations are performed by the way of three interdisciplinary case studies. More general guidelines for integral stress diagnosis, risk evaluation and conception of programmes for forest rehabilitation shall be defined in the course of a final synopsis, which also should offer an operational system on basis of artificial intelligence. The objective of such a system development is twofold:

- an expert system can act as an advisor for a forester, a forest consulting office or some governmental authority
- the formalization required for the system forces the experts to make the knowledge explicit and supports so the communication between experts of different subdomains.

From the point of view of artificial intelligence the system development is a great challenge because it is an interdisciplinary project with experts from many different domains. There are experts from economic management, forest regeneration, game ecology, meteorology, biochemical-indication and more having their own vocabulary and their own theories with unique assumptions. Theories in these domains are partly contradictory, partly redundant and often experts of one domain are not aware of certain conclusions and/or assumptions in the neighbour domain. For example, treatments typically recommended by game ecology experts may have negative effects in forest regeneration. Moreover, there exist not yet a unified problem solving procedure to solve problems. This method is still under

investigation by the experts. Practitioners in the forests usually apply a very limited amount of available knowledge to solve problems.

There are four principle tasks an expert system could support:

- the continuous supervision of a continuance,
- the determination of a goal continuance (e.g. what trees should exist in 30 years)
- the diagnosis of faults⁴ and
- the proposition of treatments (a therapy).

The operational system to be built is not the first expert system developed for forest management. Besides systems especially in North America, Germany and Scandinavia (for example by the European Forest Institute in Finland), there also some systems developed by different partners of the interdisciplinary FIW-project. There are expert systems for game ecology (Reimoser 1994), sensibility analyses (Eckmüller and Moser 1994) and other tasks not yet publicly described. However, it is not possible to unify these systems in order to support foresters adequately. As Saarenmaa et al. (1994) pointed out classical rule-based expert systems have a very limited applicability.

In 1996 we have started a new approach in the FIW II project. We have decided that not the operability of the system is the great challenge but the unified modelling of the interdisciplinary domain. In many interviews with experts from different domains and by literature studies we have tried to find out first the common knowledge structures. To support the knowledge elicitation we have established an interdisciplinary glossary of used terms in the internet. Then we decided to use the problem of bark-beetles as a first subdomain to investigate which knowledge representation formalisms are appropriate. A domain model was realised in Prolog (Krzizala 1997). A second subdomain was investigated later because in the first subdomain the number of different experts were relatively small. This second problem is that of forest revitalisation. A third subdomain will be game ecology addressed in the Ph.D. Thesis of E. Partl.

Since there exist such a huge body of knowledge about forest management, knowledge acquisition should be supported by a dedicated tool in order to enable the domain experts to enter the knowledge direct into the knowledge base. This is also necessary, because it is very likely that there will be continuously updates of the knowledge due to latest research results. Since the Austrian experts work at different institutes in different cities, the knowledge acquisition process is also

⁴ In Model-based diagnosis, a subfield of artificial intelligence, a fault is defined as a malfunction of a system explained by the difference between a correct system and a faulty system. In our domain it is more difficult to explain, what a fault is because it is not clear what a correct system is. There are ecological influenced objectives as well as economic goals that must be considered.

locally distributed. We have thus decided to support the knowledge acquisition process by an internet-based tool. With the first prototype implemented by Ch. Ledl (Ledl 1996) the expert can fill-out a knowledge representation formalism by means of a HTML-form.

2. Knowledge Analysis

As in most problem domains it makes sense to distinguish between taxonomic and assertional knowledge. The taxonomic knowledge can be seen as the theoretical knowledge of the experts that know or believe the definitional and causal relations between objects of the domain. This knowledge is generic and not restricted to concrete objects. The assertional knowledge is the knowledge about a concrete problem. For our domain the assertional knowledge contains data and knowledge about a certain geographical domain (e.g. the altitude) and characteristics of a certain continuance (e.g. stock of trees). Typically, the assertional data is partly stored in a Geographically Information System (GIS) or some other data base system. Another part must be inserted by a user manually. Of course, the adequate structuring of the data base is very important, but this will not be investigated here. Also the inferences necessary between a GIS and the expert system are of high relevance. So we must be able to abstract from large number of measurements and generate qualitative assertions about a domain. However, we do not want to elaborate on this in our paper. The taxonomic knowledge with the generally applicable inferences are elaborated on.

In comparison with other domains such as expert systems for industrial applications, forest management is characterised by the systematic scientific classification already available. Every student has to know classifications of plants, animals and minerals. These classifications lead very naturally to an object-oriented representation with inheritance structure, part-of relations and other associations (Saarenmaa et al. 1994). What is not yet encapsulated explicitly in the existing theories are such concepts as damage, symptom, diagnosis, therapy which are used when a problem is solved in the domain. In model-based knowledge acquisition these concepts are called roles of the problem solving task⁵. These objects are elicited when the process how experts solve the problems in their domain is

⁵ For example, in the KADS-methodology (Wielinga et al. 1992), (Bull 1993) meta classes are used to describe the role of objects in a general problem solving process. These roles act as a placeholder in the description of the inferences and point to the type of the domain object that can fill the placeholder.

modelled. Thus we have to model the general role “damage” and the relation that states that bork-beetles may be a damage.

Further objects, that are used in many inferences in forest management are spatial concepts such as measuring point, continuance, district and more. These objects, relations between them and inferences that may be concluded between them are a kind of commonsense knowledge. There are theories in artificial intelligence (Cohn 1995), (Kautz 1995) that support these inferences however they are computationally very expensive and it must be investigated further whether explicit inferences are sufficient for the domain.

In the scientific theory of the whole domain one typical inference structure are determination keys. These are a number of if-then-else clauses to determine for example which kind of parasite exists at a tree. Usually these rules are formalised as a decision tree in which the inferences are used in a very strict top-down sequence. They are of course influenced by the existing classification. First, attributes are asked that assign a given species to a certain class and then to a subclass. As known also from theories in machine learning, one should ask first those attributes that deliver the greatest information content. This is not always done in the existing theory due to the required sequence. Because the inference shall be generally valid and the limited representational capabilities in traditional theoretic frameworks the inferences are so inflexible. It seems possible to improve such inferences with AI-technologies because a rule-base system can handle such a determination process more flexible. For example, if such a determination inference is used, some knowledge may already exist that decides queries on a lower level of the determination tree. If this known knowledge is applied first, other queries on an upper level may be avoided.

Very symptomatic for the domain is uncertain and vague knowledge. The uncertainty has different sources. If investigations are made, there are usually large areas to be examined. It is impossible to make secure propositions over the whole area. Typically, measurements are made at points in regular distances and measured values are estimated for intermediate points. Another source of uncertainty is the inexactness of measurements. Many experiments were conducted in the three case studies to determine the repeatability of measurements. For example, experts were asked to determine aspects such as the density of tree tops or the severity of tree damages. It was recognised that different experts come to different results and moreover, if one expert is asked twice to evaluate a continuance after a longer period, the two evaluations will differ considerable. A third source of uncertainty are the long periods between investigations and changes in the environment. To see effects of treatments it takes often years and many environment variables change in the mean time. Thus a comparison of situations is very difficult and attributed by many uncertainties. These uncertainties result in uncertain inferences. With the rising of air pollution forest decline was detected. However, the causal relation

between both phenomena cannot be modelled by a simple certain relation. There exist several theories and inferences that are very likely but not certain. And there are different reasons for the forest decline. Each of these causal explanations have a certain impact.

There are not yet many systematic approaches to handle this uncertainty in the domain of forest management. The consideration of this uncertainty leads to much more complex inference structures that cannot be handled anymore without computers. For example, a determination tree can be attributed with uncertainty. This means, we cannot decide clearly which branch to follow. We must follow all branches (some branches may be cut because the likeliness is very small) to decide to which class a given species belongs. Furthermore, contradictory and redundant knowledge lead to a more complex reasoning model. Thus the assertion must be attributed by some likeliness that may be influenced by different inferences. This likeliness can be interpreted as a variable whose content can be increased or decreased by different rules.

Moreover, in an intelligent reasoning scenario for forest management the utility of knowledge must be evaluated. The acquisition of some assertional knowledge is very expensive. For example, it was recognised in the case studies that by means of bio-chemical indication, we can conclude with great certainty on certain stress factors of a forest. However, this procedure is very expensive and should only be applied if this additional knowledge will improve further conclusions that are relevant for a forest therapy.

An important type of knowledge are the goals and restrictions for forest management and rehabilitation. If a system shall support a user, these goals must be made explicit. This explicitness is not yet existent in theoretical investigations on forest management. Of course there is the forester's goal of maximal profit or return of investment. However, there are other goals that are difficult to be operationalised such as human recreation support, game population, or ecological goals as for example a potential natural society. Additionally, it is difficult to conclude from the financial goals any immediate procedures. The elicitation of goals was one of most difficult problems in our analysis because there exist very few objective knowledge. Many rules given by experts are based on unmentioned assumptions concerning the goals. In practise there should be the possibility of a compromise between different goals and restrictions. Thus the management can be interpreted as a multi-criteria optimisation problem. In this problem some restrictions such as governmental laws must be obeyed and other preferences must be set by the user (e.g. a forester).

If goals are existent and a given situation is evaluated as not confirming with the goals, treatments such as a rejuvenation of the forest or a fertilisation must be planned. There exist a great number of different possible procedures to fight against damages and influence the health of forests. Many of these treatments were also investigated in the three case studies. These procedures can be interpreted as a

distinguished knowledge type. Similar as actions in the domain of robotics which is examined very often in artificial intelligence, these procedures have certain results and conditions that must be true in order to apply them. This repository of possible treatments is the main tool to control forest management. Especially this repository will change continuously because scientific research will probably find new treatments or improves already described treatments. For a reasoning on the applicability of such treatments also a representation of similarity and adjustable variables seem to be important. A further important concept is temporal knowledge because the treatment will be dependent on many temporal aspects.

3. Knowledge Representation

The knowledge representation proposed by G. Krzizala in his master thesis (Krzizala 1997) is based on the representation and the meta-processor described in (Dorn 1993). PROLOG is used as the basic representational language and inference mechanism. The meta-processor supports different knowledge representation formalisms and reasoning techniques. However these techniques were not sufficient for the domain of forest management. Thus, new formalisms and techniques were defined but not implemented yet in the meta-processor. The next figure gives an overview on the knowledge base proposed for the domain as well as the required reasoning techniques.

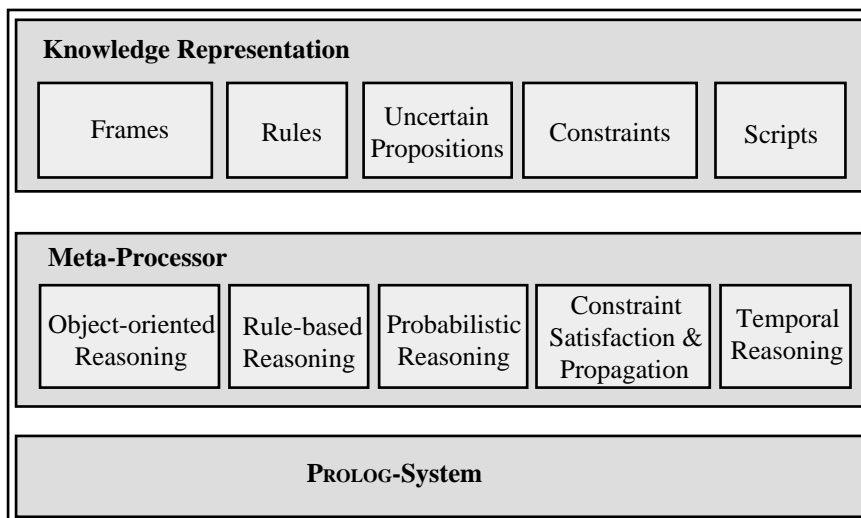


Figure 1: Knowledge representation formalisms and required reasoning techniques

There is not enough space to explain the full syntax of our knowledge representation system. However, we will give a short introduction into our representation and explain in the next subsection how frames are represented in a bit more detail.

PROLOG-predicates are usually written in a prefix notation, i.e. we first give the predicate and then in braces the arguments (e.g. `height(tree, 12)`.) PROLOG however, allows the definition of predicates as operators and operators may be defined as infix if exactly two arguments are present. We may write then "tree height 12." We use this technique to define an extended representation language that should be better readable than standard PROLOG.

3.1 Frames

Frames are used in knowledge representation to support object-oriented reasoning and to model objects of a domain and roles of a problem solving process. A frame describes an object prototypically by naming the attributes of the object. Some of the attributes may have values that are valid for all objects that are described by the frame. Frames are used to generate individual instances.

Attributes are described by so called slots. The symbol "::<" is defined as infix-operator that delimits the slots of a frame. A slot can be described by a number of facets that support a meta description of the slot. Facets are used to describe the type of the filler and procedures how to proceed if a value is asked but not stored in an instance. They are delimited by the ":"-operator.

The following frames shall be used as an intuitive example. The first frame "position" describes how coordinates are given for an object of the domain. The frame "object" describes general objects that may have a geographical position. The third frame describes a tree as a subclass of an object. The first two slots of this frame contain the name and the superclass. A "tree" is derived from "object" thus inheriting all attributes that are defined for "object". The third slot contains a "part-of" relation that describes from which parts a "tree" is built up. The remaining slots describe the height and the material of a tree.

```
frame(position :: X : type(int) :: Y : type(int) :: Z :
       type(int)).
frame(object :: at : type(position)).
frame(tree :: super : object :: parts : [top, trunk, root] ::
       height : type(int) :: material : wood).
```

If we generate instances of such frames we support two notations. The flexible one is where we name each slot name before we write the value. The short form does not name the slot names. Therefore, the attributes must given in the same sequence as in the frame description. If the value is not known we use the "_" symbol.

```
tree(t1 :: height : 16).
```

```
tree(t1 :: (143, 26) :: 16).
tree(_ :: _ :: 16).
```

An access of attributes is defined as follows, whereby the first clause queries the attribute and the second clause queries the truth:

```
?-tree(t1 : height : X)
?-tree(t1 : height : 16)
```

If attributes are queried in the left-hand side of a rule or in a constraint, a simplified call is possible:

```
if(tree : t1 : height < 16) then ...
```

Besides the sketched representation, the frame meta-processor supports inheritance, defaults, type checking, handling of sets and simple spatial reasoning.

3.2 Uncertain knowledge

For the forest management domain the representation of uncertain knowledge is required which is not yet implemented in the existing meta-processor. Belief networks as proposed by Pearl (1986) seem to capture the essential requirements. In belief networks we can assign probabilities to single propositions. Thus we may write in our language:

```
tree : t1 : high ! 0.8. or tree(t1 :: height : high ! 0.8).
```

stating that the probability that the tree t1 is high is .8. Inferences in the belief network can then change this probability if necessary.

3.3 Rule-based Inferences

An important reasoning mechanism in forest management is rule-based reasoning where from a given situation (constrained in the left-hand side of a rule) new facts are deduced. These inferences may be associated with a certainty of the inference as in belief networks.

3.4 Constraints and goals

As already mentioned, there are usually concurrent goals for forest management. Scientists have certain ecological goals such as a “potential natural forest community”, governmental authorities want recreational opportunities for human citizens and the forest owner wants a high return of investments. Since these goals

are contradictory, not all of them can be fulfilled totally. Therefore, soft constraints can be used to measure a degree of satisfaction. In (Dorn 1997) a model is elaborated that allows the definition of alternative constraint sets. Thus a user should first enter its preferences in an expert system which then generates possible future scenarios described by alternative constraint sets.

3.5 Event-oriented knowledge

If a diagnosis is made and a new goal continuance is defined, we must search for actions that transform the given situation into the goal situation. In artificial intelligence typically so called operators are used that contain a list of pre-conditions that must hold before the action is performed and a list of effects that will happen if the action is performed. Means-end analysis may be used to select then appropriate actions (Fikes 1971).

Scripts allow an improved representation with capabilities for describing temporal and causal interdependencies explicitly and abstraction over actions. For a more detailed description we refer to (Dorn 1995).

4. Knowledge Acquisition

The domain of forest management is so complex and extensive that it would be too expansive to use the typical knowledge acquisition process comprising a knowledge engineer. Furthermore, the assumption that the body of knowledge continuously grows forbids this approach. Therefore, we have decided to support individual experts so that they can enter their knowledge individually into the knowledge base. This attempt is not new but what is different to similar approaches is that there are so many experts from different domains working in different locations in Austria. This has lead to the idea to realise the knowledge acquisition tool in the internet.

A first prototype of such a knowledge acquisition tool was developed by Ch. Ledl in his master thesis (Ledl 1996). This system can be used as a browser through the knowledge base as well as a system to manipulate the knowledge base. Since the manual input of many experts will be very error-prone and furthermore the expert shall not be forced to learn any syntax, the system provides the user with forms to be filled out. These forms described by HTML-forms are then translated into the PROLOG representation of our data base. Certain control mechanisms must be defined to allow a pseudo parallel access of users to the knowledge base. These are realised by CGI-BIN scripts on the server side of the system. If inputs are made, these inputs should be checked with the existing knowledge base. If for example an object is defined and a new attribute is attached, it makes sense to check whether

this attribute is already defined for a superclass of this new class. In order to check this out, a search must be performed that is supported by our meta-processor realised in PROLOG. Thus there must be a PROLOG-process running to which the CGI-BIN process sends a query. The following graphic displays this architecture.

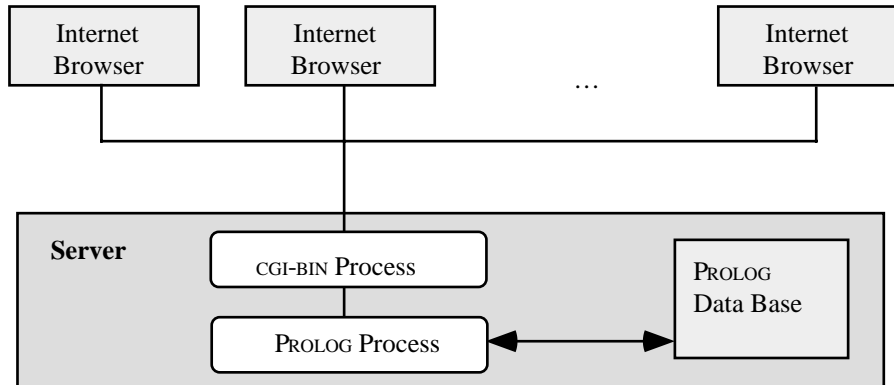


Figure 2: Architecture of the system

HTML forms offer user interface elements such as push buttons, check boxes, selection lists and text areas for free ASCII text.

The most important design criteria of the system was the ease of use. However, there are two aspects that conflict with each other. If we want to check inputs, we must submit the inputs to the server. In HTML the only possibility is to force the user to click the submit button. This results either in very small forms which will be submitted after few inputs or to define relative large forms in which many errors can be made by the user. It is also difficult in the large form to give good hints to the user where he has made the error.

5. Outlook

Forest decline and reasons for forest decline such as air pollution are not constrained to single nations. The research on such topics is of international relevance and moreover, treatments to overcome problems must be applied over frontiers. The distributed knowledge acquisition approach presented here could be used also for a more international research as well as an international advisor for forest

management problems. At the moment the whole knowledge base as well as the glossary are German language based. Of course Austrian foresters who are intended to be the users of this expertise should not be forced to use an English expert system⁶. However, due to our decision to implement the knowledge base in the internet, language translation tools as for example that supplied by the altavista search engine can be used to support the translation.

A second problem is that of authorization. If an international community wants to build up the knowledge base we have to define procedures and mechanisms for protection of knowledge entered by users as well as voting mechanisms for redundant or contradictory knowledge. In our project we have realized a very simple approach for the glossary. We have defined research groups that can authorize themselves by a password. For each term in the glossary, each group can enter now a definition. This definition may only be deleted by the system administrator or a member of the group. Although we have ten different groups, usually there are only two or three different definitions. For the knowledge base the process will be more complicated, because it will be necessary to eliminate certain redundant information to make an expert system correctly work. Furthermore, it is very important to formalize national specific situations such as legal and geographical issues⁷.

References

- Bull (1993) *Open KADS User's Guide*, Bull S.A. CEDOC 20, rue Dieumegard, 93406 Saint Quen CEDEX, France.
- Cohn, A.G. (1995) The Challenge of Qualitative Spatial Reasoning, *Computing Surveys* **27** (3) pp. 323-327.
- Dorn, J. (1993) *Wissensbasiertes Planen oder Prolog in Aventurien*, Skriptum zur Vorlesung, Technische Universität Wien.
- Dorn, J. (1995) Dependable reactive event-oriented planning in *Data and Knowledge Engineering* **16**, pp. 7-49.
- Dorn, J. (1997) Modellierung von Zielen beim Forstmanagement als Constraint-Satisfaction-Problem, *Forstliche Schriftenreihe*, Universität für Bodenkultur.

⁶ It is already anticipated that there will be a great problem of acceptance by the practitioners in the forest. There is the problem that they have to use a new technique but the greatest obstacle will be the need to make objectives explicit. To use a foreign language program would complicate the application.

⁷ Often experts give rules that are only valid in Austria due to certain Austrian laws or the Austrian climate.

- Eckmüller, O. and Moser, M. (1994) Sensibilitätsanalysen eines Expertensystems zur Bestandeshaltung in *Forstliche Schriftenreihe* Universität für Bodenkultur Bd. 7, Wien, pp. 257 - 269.
- Fikes, R. and Nilsson, N. (1971) STRIPS: A new approach to the application of theorem proving to problem solving, *Artificial Intelligence* 2, pp. 189-208.
- Führer, E. (1994) Forschungsinitiative gegen das Waldsterben Programm FIW II: von der Waldschadensforschung zur Waldökosystem-Sanierungsforschung in *Forstliche Schriftenreihe* Universität für Bodenkultur Bd. 7, Wien, pp. 1 - 10.
- Kautz, H.A. (1985) Formalizing Spatial Concepts and Spatial Language in J.R. Hobbs (ed) *Commonsense Summer: Final Report*, No. CSLI-85-35, Stanford University.
- Krzizala, G. (1997) *Wissensakquisition und Wissensrepräsentation im Rahmen der FIW II*, Diplomarbeit Technische Universität Wien.
- Ledl, Ch. (1996) *Entwurf eines Wissensakquisitionssystems für Expertensysteme im Internet*, Diplomarbeit Technische Universität Wien.
- Reimoser, F. (1994) Expertensystem „Wildökologie - Waldverjüngung“ in *Forstliche Schriftenreihe* Universität für Bodenkultur Bd. 7, Wien, pp. 237 - 256.
- Pearl, J. (1986) Fusion, propagation, and structuring in belief networks, *Artificial Intelligence* 19 pp. 241-288.
- Saarenmaa, H. et al. (1994) Object Oriented Modeling of Tasks and Agents in Integrated Forest Health Management, *AI Applications* 8 No. 1 pp. 43-59.
- Wielinga, B.J., Schreiber, A.T. and Breuker, J.A. (1992) KADS: A modelling approach to knowledge engineering, *Knowledge Acquisition* 4(1) pp. 5-53.