AN APPROACH FOR BUILDING GENERIC DIAGNOSIS MODEL IN AGRICULTURAL DOMAIN

Abeer El-Korany **El-Sayed El-Azhary** **Mohamed Yehia**

Central Laboratory of Agricultural Expert System
e-mail (abeer/sayed/mdyehia)@mail.claes.sci.eg

Abstract: Expert systems development is a complex and expensive process that needs to be applied in an organized manner. Building expert systems has been seen as a modeling activity. The aim of this research is to present an approach for building generic diagnosis model in the agricultural domain. This generic model has the ability to derive diagnoses for differently structured individual systems from the agriculture domain. The generic model would serve both the developers and the implementers of a diagnosis expert system. The primary goal for developing this model, and consequently a tool based on it, is to facilitate the rapid development of a diagnosis component by offering the system builder a template that can be easily filled.

Keywords: Diagnosis- Expert system- Modeling – Problem solver

1 INTRODUCTION

Diagnosis is the problem of trying to find the causes of abnormal observations. There are mainly two distinct approaches to diagnosis: heuristic based approach, such as [Shortliffe, 1976], [Reggia, 1983]. This approach uses symptom-causes association rules gathered from experts in a domain. The other approach is model based approach such as [Genesereth, 1984], [Davis, 1984], [de Kleer and Williams, 1987], [Struss and Dressler, 1989], [Nejdl, 1995], [Purna, 1995]. This approach is based on the availability of structural and behavioral device model. These models enable the diagnostic problem solver to attain its goal. There is also another approach which combines these two approaches, such as [Koseki, 1989], [Lee, 1990]. The differentiation between these approaches depends on the applicable domain. The most suitable approach for the agricultural domain is the heuristic based approach because it is easy to get symptom-causes association rules from the experts rather than getting the structural and behavioral plant model.

The aim of this work is to present an approach for building generic diagnosis model in agricultural domain. This approach has been achieved by identifying and capturing all knowledge related to all diagnosis expert systems that have been developed at the Central Laboratory of Agricultural Expert System (CLAES). By analyzing this knowledge we identified the common knowledge for each class of crops as well as specific knowledge that depends on the crop.

According to the crop classification which are vegetables, field crops and trees, the common knowledge for each class of crops has been obtained by integrating the related knowledge of each crop in the class. This knowledge contains domain, inference and task knowledge.

A tool has been developed to facilitate the rapid development of diagnosis expert system. This tool supports the generic model and contains the common knowledge for each class of crops and the specific knowledge for each crop. So, the developer of a diagnosis expert system can use the common knowledge in building his own expert system. He/She can use also a specific knowledge of a crop in the same class or add new knowledge. This tool was used in the actual ICARDA training held by CLAES last year.

A key advantage of the generic diagnosis model is providing a common ontology for different crops, saving the time and reducing the effort consumed in building the expert system.

Section two presents earlier research efforts for developing diagnostic expert systems in agricultural domain, approaches of building generic model and tool. Section three demonstrates the diagnostic generic model design based on CommonKADS methodology (Breuker and Wielinga 1994; Schreiber 1999). The implementation of the diagnostic generic tool has been demonstrated in section four. Section five presents an example run on the tool. Section six presents the conclusion of this work.

2 RELATED RESEARCH EFFORTS

This section gives a brief history about different expert systems that have been developed in the agriculture domain. It also presents approaches used in the generic modeling of knowledge-based system. Finally, it illustrates different tools that have been developed at CLAES to build agriculture expert systems.

2.1 Agriculture Expert Systems

Knowledge-based expert system technology has been applied to a variety of agricultural problems since the early 1980s. The following paragraphs present the more recent research efforts of how expert systems were considered in agriculture.

CUPTEX [Rafea, 1995] is an expert system, which has been developed for handling management of cucumber disorders. The main objective of this expert systems is to identify the cause of an observed disorder, its severity, and then proposes the appropriate remediation. The user can consult directly the remediation part if he knows the cause of the disorder. However, the remediation part, in this case, verifies the cause given by the user before giving the remediation advice.

NAPER Wheat expert system [El-Beltagy, 1995], is developed for handling the production management aspects for wheat crop. This system includes varietal selection, planing the agricultural operation's dates, irrigation and fertilization requirements, insect and disease identification, insects and disease remediation, and harvest management. The methodology adopted for this system is the generic task methodology [Chandrasekaran, 1988].

Yialouris and others [Yialouris, 1996] developed an expert system for tomato to handle tomato diseases identification problem. Frame knowledge representation scheme has been used for representing the knowledge base. It is worth noticing that fuzzy logic is used for handling the uncertainty in diagnosis. However, references for other expert system in this domain may be found in [Liebowitz, 1998].

PCEST [El-Azhary, 2000] is a pest control expert system for tomato. It contains two phases, diagnosis and treatment. The designed expertise model of this system has been done following the CommonKADS methodology [Wielinga 94].

2.2 *Generic Modelling of Expert Systems*

The idea behind building generic domain-specific model is to capture specifications and automatically generate or configure the target application in particular field. The generic knowledge models which can be reused in different applications are an essential issue for a more efficient application of knowledge based techniques in design. The recently developed general knowledge modeling schemes can be roughly grouped into two families:

• Modeling of *domain knowledge* in the sense of KIF (Genesereth and Fikes 1992) and Ontolingua (Gruber 1992), CYC (Lenat and Guha 1990), TOVE (Fox et al. 1993), etc. Such techniques allow us to represent the objects, attributes, relations, and constraints which describe a domain. These approaches are very expressive on the domain layer but they do *not* focus primarily on problem solving methods and their interactions with domain knowledge.
• Knowledge modeling which *integrates domain and problem solving knowledge* like, for instance, CommonKADS (Breuker and Wielinga 1994; Schreiber 1999), or more recently, UPML (Fensel et al. 1999). CommonKADS is based on a sophisticated general scheme to describe the various kinds of knowledge: domain, task, and inference knowledge2. But the currently existing realizations are only dealing with relatively simple domains and problem solving techniques. Strategic knowledge isencoded as task control of relatively simple forms. The translation of the proposed general principles into working techniques needs further work.

2.3 *Expert Systems Development Tools*

Different expert systems development tools have been developed by CLAES. These tools are heavily used to build different expert systems tasks in the agriculture domain.

2.3.1 *KROL (Knowledge Representation Object Language)*

KROL (Shaalan,K.,Rafea,m.,1998)is built on top of SICStus Prolog language. Thus, the full power of Prolog, as a logic programming, and the object-oriented programming style are joined. This language provides a good environment for the development of second-generation expert systems

The main facilities that KROL provides are: the expressive power to represent complex knowledge, the multi-paradigm knowledge representation, modularising of the knowledge base, controlling inheritance of properties thorough a concept hierarchy, writing inference mechanisms at different level of granularity, the primitives that allow for higher level knowledge base modeling approaches to scale to large problems, and the synergy of different inference mechanisms in one system.

2.3.2 *KSR (Knowledge Share and Reuse)*

KSR is a tool that extends the facilities of KROL and introduces the concept of knowledge sharing between different applications. This tool holds reusable knowledge to facilitate rapid expert systems building. KSR was implemented in C++ taking the advantage of object-oriented techniques, to facilitate building expert system on multi-user environment, and to facilitate running expert system on WEB. KSR is used by CLAES for building many expert systems that run on the WEB, for example: cucumber, wheat, and rice, and many of expert system that run on desktop. For example: cucumber, bean, strawberry and melon.

3 AN APPROACH FOR BUILDING GENERIC DIAGNOSIS MODEL

"Generic Modeling" is the process of constructing a model that represents features common to a class of phenomena. Generic modeling involves trying to make models flexible and sufficiently robust so that when new requirements inevitably arise, neither the model nor the applications have to change substantially. Also, generic structures result in smaller models, which are easier to understand than ones with large numbers of entities. Many approaches have been introduced for this purpose, but theoretical approaches must be supported by practical guidelines in order to apply these methodologies in real life applications. The proposed approach for building generic diagnosis model is the result of accumulated experience gained through many years of developing several expert systems in the agricultural domain by (CLAES). This generic model has the ability to derive diagnoses for differently structured individual systems from the agriculture domain. To do so, we have aimed to identify and capture all knowledge that is related to the diagnosis task, regardless of the crop, and determine the common knowledge that construct the generic model. Then we identify concepts that vary from one crop to another. This knowledge is used as a guide to complete the construction of a new diagnosis expert system. The following subsections illustrate the approach used in building the generic diagnosis model.

3.1 General and specific knowledge

In the agriculture domain, plants are classified into three main categories: field crops, vegetables, and trees. At CLAES, many expert systems for different plant under each of these categories have been developed and used in the field. These systems are for:

- Vegetables: Bean, Cucumber, tomato, Melon and Strawberry
- Trees: grapes, mango, and citrus
- Field Crops: faba-bean, rice, and wheat

Each of these expert systems, has its own domain specific knowledge. By integrating the domain specific knowledge for all the plant under each of the above category, we can extract knowledge that are common between plants. Thus, constructing the general knowledge that can be shared and reused by any plant under each of the three categories. Consequently, we determine specific knowledge that belongs to specific plant. A knowledge base library that contains general and specific knowledge for each plant category is constructed. The architecture used in building this library is similar to that of building the KB. i.e. we have general and specific knowledge for either the domain and the problem solving knowledge. The aim of this technique is to integrate different types of knowledge within a unique model, making the development of any diagnosis system in the agriculture domain an efficient and.

3.2 The Generic Diagnosis Model Architecture

Building knowledge based systems (KBSs) has been seen as a modeling activity .i.e. as the construction of several models capturing different types of knowledge (Calancey, 1989). A knowledge based system model should offer similar results in problem solving for problems in the area of concern. The model resulting from this process must contain different kinds of knowledge necessary to solve a task in hand. According to the KADS model of expertise (Schreiber, 1994), the proposed generic diagnosis model integrates the domain knowledge and the problem solving knowledge. Domain knowledge contains all knowledge about diagnosis in the agriculture domain. The problem solving process has two main categories: knowledge about the problem solving steps, their interactions, and their relations to the domain knowledge; and task knowledge - how the problem solving process is controlled.

3.2.1 Domain knowledge

On the domain level, the main knowledge categories can be summarized as follows:

- *General domain knowledge (Ontology)* containing generic as well as specific knowledge about concept, attributes, and their values of different diagnostic systems. As mentioned in the previous section, each plant category has a common and a specific domain ontology. Generic domain ontology contains the common concept with their attribute. Values of each attribute is the union of all the attribute value used in different diagnosis system of all the plant under the same category. The domain ontology consists of three main concepts ontology:
    - *Plantation ontology:* Which contains concepts related to the plant environment such as: soil, water, climate, plant and plantation. These generic concepts can be reused across different task in the agriculture domain.
    - *Disorder ontology:* For each crop class we integrate the disorder taxonomy that affect all the plant under this category.
    - *Observation ontology:* This ontology contains concepts that represent different plant parts of each crop class. Attributes and their values of each concept represent the features that may appear on different plant parts as a result of disorder attacking.

Its significant to mention that, during construction of these common ontology, it has been found that some systems use different vocabrary (names) for concepts, or attribute or values that has the same meaning. For example, concepts (climate, weather) attribute (status, appearance), etc. To prevent ambiguity, this terminology has been unified.

- *Domain model knowledge*

Domain model knowledge contains the required relationship between concepts attribute. The generic domain models consist of the complete relationship between concept attributes. While the for specified domain models, we determine the input/output concept attribute pairs that can be used in building this model. The developer is free to fill this model with the appropriate knowledge.

3.3 Problem solving method

In general, diagnosis is considered one of the expert system classification problem solving method [Clancy, 1987]. The basic steps that construct the classification expert system are:

1- Data abstractions
2- Heuristic match
3- Solution refinement

For complex systems, these essential steps can be extended to contain more features that make the matching between data abstraction and solution is immediate and thus improves the total system efficiency. The data abstraction step is used to collect all the primary knowledge about the diagnostic problem. For example, knowledge about different plant environmental status like: date of plantation, soil type, etc. Also, knowledge about the user complaints of a particular case. The complaints describe symptoms that affect different plants parts as a result of disorder attack. In the agriculture domain, there is a wide range of symptoms that may affect the plant. To reduce this input space, the data abstraction step is divided into the following steps:

1- Determine plant age
2- Determine growth stage
3- Determine combatable plant part
4- Generate primary observation

Since the plant age affects the plant observations that should applied to the user. For example, small plants can not have fruits. Thus, we begin by determine the plant age which in turn used to determine the plant growth stage. Then, we are able to detect the plant parts that are suitable for this stage. Moreover, since each plant class (vegetables, trees, and field crop) has its own observations, only the class observation is applied to the user so that he/she focuses and selects the symptoms of his case. This leads to reduce system inputs and improve the system performance.

The heuristic match step (which we call *predict disorder*) is used to perform a match between the user primary complaints and a broad class(s) of disorders. Having identified a solution abstraction which narrow the solution space, we still need to specify a solution in that space. This is accomplished through the solution refinement step which in turn divided into the following steps:

1- Generate additional observation
2- Confirm disorder

To reduce the search space, generate additional observation inference step is used to collect the observation related to the suspected set of disorders deduced by the predict inference step. Thus, only the symptoms that related to the narrow space of suspected disorder is applied to the user. This leads to accelerate the reasoning process and improve the system efficiency. The modified inference structure is shown in figure1.

4 THE GENERIC DIAGNOSIS TOOL

The generic diagnosis tool is a configurable tool that supports the easy creation of diagnosis expert systems in the agriculture domain. This tool has been developed using C++ which supports object-oriented concepts and COM technology which provide a good environment for the development of second-generation expert systems. The general structure of the tools is:

1. Problem solver editor
2. Concept editor
3. Domain model editor

The following subsections illustrate the basic functions of each of these editors.

The main script of the generic diagnosis tool is the problem solver editor that represents the problem solving method. As described earlier, the generic diagnosis model provides a sequential problem solving method that consist of seven inference steps.

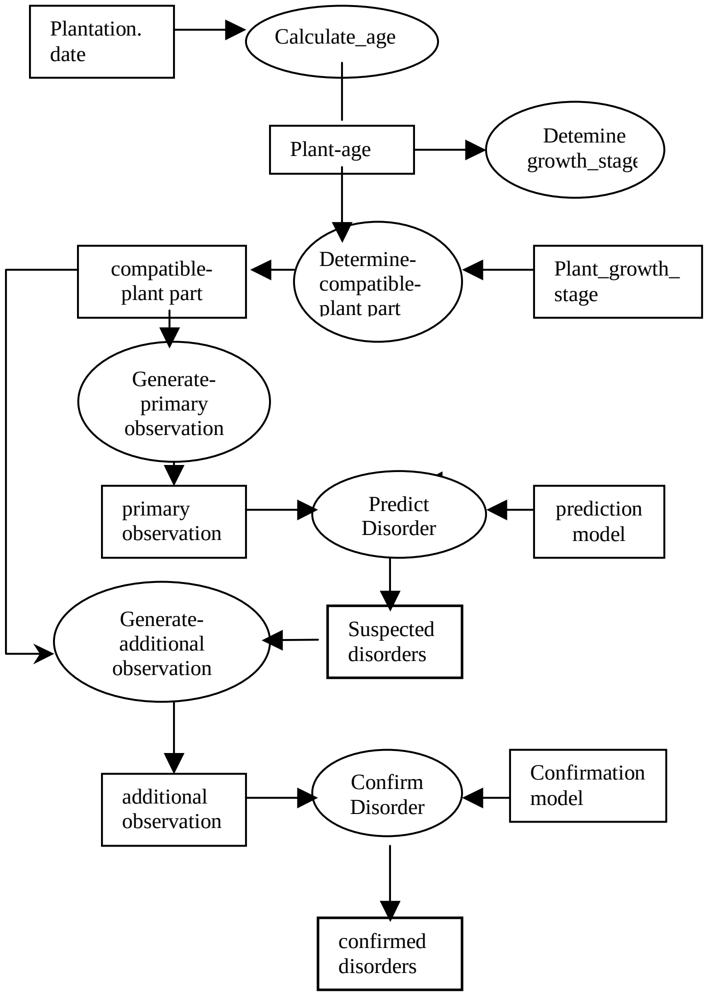

Fig.1: The inference structure of the generic diagnosis model

4.1 *Problem solver editor*

Some of these steps are generic for all plants under all categories. These steps are built in the tool and are hidden for the expert system developer. These steps are:

1. Determine plant age
2. Determine combatable plant part
3. Generate primary observation
4. Generate additional observation

Specific inference steps are applied to the developer in order to fill them with the appropriate knowledge. Figure2 demonstrates the diagnoses problem solver which contains the three plant dependent inference steps that can be modified by the developer.

4.2 *The Concept editor*

The concept editor contains the common and specific ontology of the generic diagnosis model. The default inheritance mechanism is provided, which implemented by the delegation mechanism.

4.3 *The domain model editor*

The domain model editor allows the developer to provide the domain models of the generic diagnosis model. Since all the domain models are in form of rule clusters, the domain model editor is simply a rule editor.

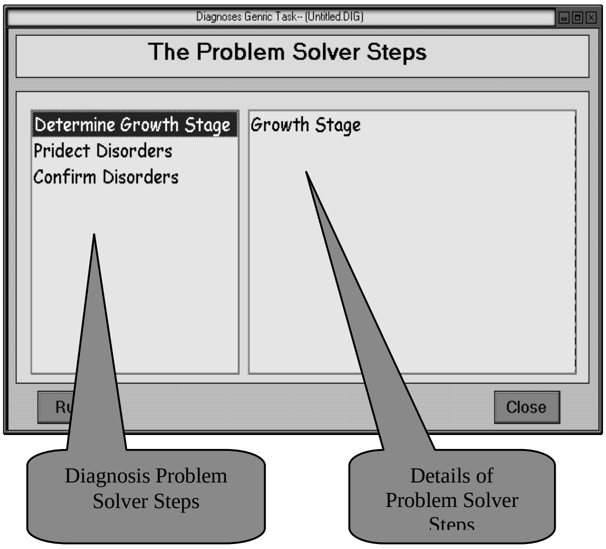

Fig2. The Problem Solver Editor

4.4 *The domain model editor*

The domain model editor allows the developer to provide the domain models of the generic diagnosis model. Since all the domain models are in form of rule clusters, the domain model editor is simply a rule editor.

5 EXAMPLE OF UTILIZATION

This section presents the result of using the generic diagnosis tool in building real world agriculture expert systems. The example illustrated here is for diagnose Soybean Fungal Diseases. These examples were developed by the trainer of the ICARDA that has been held by CLAES at Oct. 2003. As described in the previous section, to build a new expert system the developer should construct the following inference steps:

1. Determine Growth stage
2. Predict Disorders
3. Confirm Disorders

The following subsections demonstrate the development of each of these steps.

5.1 *Determine Growth Stage*

According to the design of the generic diagnosis model[ref], the determine growth stage inference step use a rule based model to construct this relation. As

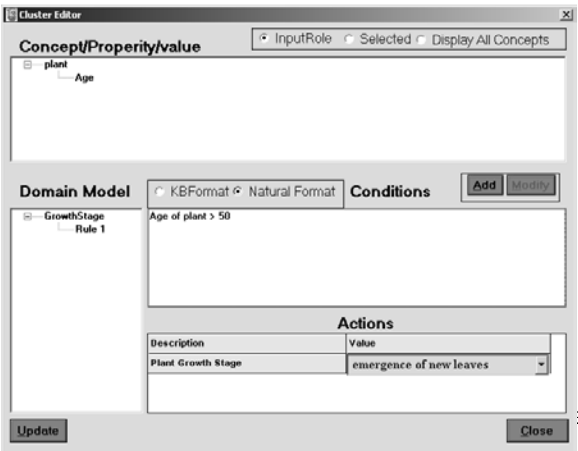

Fig.3 The Rule Editor of Growth Stage

writing rule condition in natural text format, selecting the matched growth stage in the action part

5.2 *Predict Disorders*

The predict disorder inference step use the prediction model to determine the primary set of disorders that may affect the plant due to the user primary complains. Thus, this model has a direct contact with the concept editor in order to determine the disorder taxonomy of the plant. The disorder taxonomy of the plant is a subset of the global disorder taxonomy of the plant's category. Accordingly, the concept editor provides the developer the facility to add/delete a concept from the concept hierarchy. Furthermore, the direct link between concept editor and rule editor reflects the removal of disorder from concept hierarchy on the rules that contain this disorder. Figure 4 shows the concept hierarchy editor of the generic diagnosis tool.

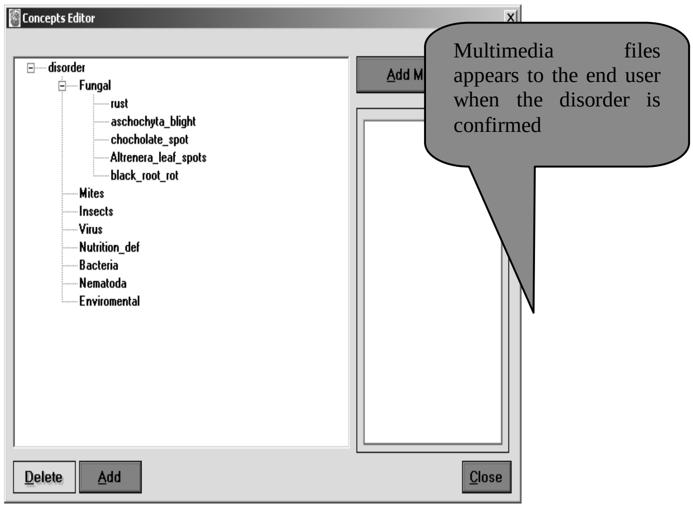

Fig.4 The disorder Concept hierarchy

After determine the required disorder set, the developer can add/edit/delete rules to the prediction model as shown in figure 5

5.3 *Confirm Disorders*

The last step, the developer can add/edit/delete the rules of the confirmation model.

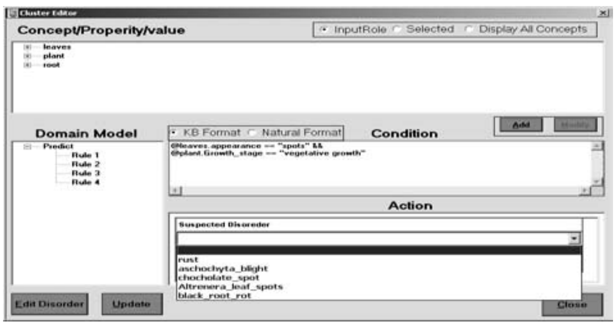

Figure 5 Rule Editor of Prediction Model

6 CONCLUSION AND FUTURE WORK

Generate generic model for diagnosis expert system will increase the use usage of this model in building other expert systems for different, which minimize the cost and effort of knowledge based systems development. Future work that can improve the use of expert system application in the agriculture domain are:

1. Building other generic models for other agricultural operations such as irrigation, fertilization, treatment, plant care.
2. The tool can be extended to support the knowledge acquisition stage.

REFERENCES

Chandrasekaran, B., Generic Tasks As Building Blocks for Knowledge-Based Systems: The Diagnosis and Routine Design Examples. To appear in Knowledge Engineering Review, 1988.

Davis,R., Diagnostic Reasoning Based on Structure and Behavior. Artificial Intelligence, 24:347 - 410, 1984.

De Kleer, J. & Williams, B. C. Diagnosis Multiple faults. Artificial Intelligence, 32:97 - 130, 1987.

[El-Azhary, 2000] El-Sayed El-Azhary, Hesham A. Hassan, A. Rafea, "Pest Control Expert System for Tomato (PCEST)", Journal of Knowledge and Information Systems (2000) 2: 242-257, Springer-Verlag London Ltd.

El-Beltagy, S., Rafea, A., Kamel, A., Sticklen, J., Schulthess, U. and Ward, R. An expert system for wheat disorders diagnosis and treatment using a hierarchical classification problem solver. In :A.J. Udink ten Cate, R., Martin-Clouaire, A.A. Dijkhuizen and C. Lokhost (Editors), Preprints of the 2nd IFAC/IFIP/EurAgEng Workshop on Artificial Intelligence in Agriculture, 29-31 May 1995, Wageningen, The Netherlands. Pp. 277-281, 1995.

Fensel, D., Benjamins, V.R., Motta, E., andWielinga, B.: 1999, UPML - a framework for knowledge reuse, in: Th. Dean (ed.): Proc. IJCAI–99, Stockholm, Morgan Kaufman Publ., USA, pp. 16-23.

Fox, M., Chionglo, J.F., and Fadel, F.G.: 1993, A Common Sense Model of the Enterprise, Proceedings of the 2nd Industrial Engineering Research Conference, pp. 425-429, Norcross GA: Institute for Industrial Engineers.

Genesereth, M., and Fikes, R.: 1992, Knowledge Interchange Format Version 3.0 Reference Manual,

Genesereth, M., The use of Design Description in Automated Diagnosis. Artificial Intelligence, 24:411 - 436, 1984.

Gruber, Th.: 1992, Ontolingua: a mechanism to support portable ontologies, Technical Report KSL 91-66, Stanford University Knowledge Systems Lab, 1992.

Koseki, Y. Experience Learning in Model Based Diagnostic System. Proceeding of IJCAI: 1356 - 1362, 1989.

Lee, S. N. DM^2: an algorithm for diagnostic reasoning that combines analytical models and experiential knowledge. Machine Learning and Uncertain Reasoning, Knowledge Based System(vol.3). Edited by B. R, Gains & J. H. Boose, Academic press, 1990.

Lenat, D.B., and Guha, R.V.: 1990, Building large knowledge based systems. Representation and inference in the Cyc project, Addison Wesley, 1990.

Liebowitz, J., The handbook of Applied Expert Systems, CRC Press LLC, USA, 1998.

Nejdl et al., "A Formal Framework for Representing Diagnosis Strategies in Model-Based Diagnosis systems", Proc. International Joint Conference on Artificial Intelligence, August 1995, pp. 1721-1727.

Purna et al., "A Framework for Hypothesis Generation in model-Based Diagnosis", Expert System With Applications, volume 9, No. 1, 1995, pp.41-54.

Rafea, A., El-Azhari, S., Ibrahim, I., Edrees, S., and, Mahmoud, M., Experience, with the development and deployment of expert systems in agriculture. Proceeding of IAAI-95. 1995.

J.A. Reggia, D.S.Nau, and P.Y.Wang, Diagnostic expert systems based on a set covering model, Int.J.Man-Machine studies 19, Pages 437-460, 1983.

Shaalan,K.,Rafea,M.,Rafea,A. 'KROL: A Knowledge Representation Object Language on Top of Prolog'. Expert Systems with Applications, An International Journal Vol 15 pp 33-46 Elsevier Science Ltd,1998

Schreiber,A.TH., Wielinga,B., Akkermans,J.M, Van DE Velde,W., de Hoog, R.(1994). CommonKADS: A comprehensive methodology for KBS development, IEEE Expert, Vol.9(6), pp. 28-37.

E. J. Shortliffe, Computer Based Medical Consultations: MYCIN, Elsevier, New York, 1976.

P. Struss and O. Dressler, "Physical Negation", Integrating fault Models into the General Diagnostic Engine, Proceeding of IJCAI: 1318-1323, 1989.

Technical Report KSL 91-1, Stanford University Logic Group, 1992.

Technical Report TR\CLAES\272\2003.9, 2003

Yialouris, C.P. and Sideridis, A.B. An Expert System for Tomato Diseases. Computer and Electronics in Agriculture 14(1996), pp. 61-76, 1996.

Wielinga,J. Expertise Model Definition Document, ESPRIT Project P5248 KADS-II, Document Id. : KADS-II/M2/UvA/026/5.0, University of Amsterdam, 1994.