

# A Web-based System for Automated Medical Knowledge Acquisition

Altan Hagicumus<sup>1</sup>, Thomas M. Gattton<sup>1</sup>, Malrey Lee<sup>2</sup>, Pradip Dey<sup>1</sup> and Arun Datta<sup>1</sup>

<sup>1</sup>School of Engineering and Technology, National University, San Diego, CA USA

<sup>2</sup>School of Electronics and Information Engineering, Chonbuk National University, Jeonju, SOUTH KOREA

**Abstract** - Automated knowledge acquisition of production rules can be performed through data mining or direct user input of case-based reasoning. However, knowledge base editors and the decision trees representing case-based reasoning in specific domains are difficult for domain experts to navigate and a domain model that provides a more intuitive interface for knowledge base editing is required. This paper describes the implementation of a hierarchical graphical user interface module that provides an intuitive interface to enter case-based knowledge and automatically generates domain-specific production rules. The knowledge acquisition algorithm and software architecture are described and the code implementation presented. The system environment is addressed for issues in distribution and security control and their solutions are specified. A knowledge acquisition scenario is presented to demonstrate the system operation and evaluated for performance. The GUI deficiencies are identified and conclusions and recommendations for further system improvements are made.

**Keywords:** Automated knowledge acquisition, production rules, medical diagnosis, healthcare.

## 1 Introduction

The reliability of knowledge bases developed through automated approaches, such as data mining, has not attained 100 percent and a means of rapidly developing total accuracy through domain expert knowledge acquisition is needed in many domains, such as healthcare. Inaccurate diagnosis and treatment in healthcare applications requires a solution that provides the accuracy of medical experts. A user-friendly interface has been developed [1] and a prototype system developed [2]. This paper describes the implementation of the prototype in a system environment and the security and distribution features of the software.

The next section describes the software architecture of the program, security and distribution features and the code implementation. Section 3 presents a simple knowledge acquisition scenario for obesity, which is evaluated for further enhancements and conclusions in the final section.

## 2 System implementation

The system is developed using C++ as the programming language. Visual C++ 6.0 is used as the compiler and the development environment. For GUI support, MFC (Microsoft Foundation Classes) library is used. The project solution is a dialog-based application with MFC support. After constructing the decision tree, all the production rules are printed to a text file called "ProductionRules.txt" which is in the same path as the executable. The diagnostic section of the program uses an interactive GUI, which enables the expert to define the domain and the related components, hierarchically. Subsequently, the knowledge acquisition continues in a procedural fashion. The basic algorithm used within the software for traversing is a Depth First Search (DFS) algorithm, which is implemented recursively.

### 2.1 Software background

The module is developed as an ActiveX object. The idea is to build a component, which is distributable and modifiable, without affecting the overall architecture of the whole solution. Being distributable is a great advantage considering Common Object Request Broker Architecture, as the distributed object may be developed in any preferred technology and may be integrated into either a localized solution or into a web environment. The preferred language for the ActiveX is C++ and the module is developed in the Microsoft VC++.NET environment with MFC (Microsoft Foundation Classes) code library support.

Within the health care system, which is designed to be a web-based application, the ActiveX will be located into a desired page and serves interactively.

### 2.2 Data Storage

The knowledge base for the decision trees is stored in MySQL, which is also the preferred database for the whole health care project. The Decision Making Module reads and writes data to a table called tbl\_DecisionMaking. This table may be related within the database and used to acquire expert data within the whole project.

The preferred connection with the database is a DSN connection, instead of DSN-less connection, because DSN-less requires the code library and the associated .dll to enable direct access to MySQL database, so the .dll file must also be downloaded with the module by the client. To avoid this overhead, the module remotely uses a system DSN in the web server.

### 2.3 Knowledge acquisition algorithm

Decision trees, as illustrated in Figure 1, can be used to generate production rules, but they are non-intuitive and difficult for domain experts to understand and navigate. Experts find a graphical representation of a system's hierarchy, as shown in Figure 2, easier to understand and build. Further, domain experts can easily associate symptomatic information with each leave of the hierarchy diagram. In the leaves of Figure 2, the symptoms associated with the failure of 3 of these components exhibits the symptom that the car won't start. Once these symptoms are associated with the individual components of the system, as represented by leaves of the tree, they become the top level branches of the decision tree, as shown in the decision tree of Figure 1. After the top level symptoms are identified, the user can input the sequence of questions that lead to a specific conclusion and recommendation. Each branch is automatically queried in the user interface and the user can select to ask another question or make a conclusion for each unique branch of the decision tree. This isolates the user from the decision tree and intuitively leads them through a set of queries that build a production rule for each branch of the decision tree. Data is kept hierarchically in a tree structure while the expert defines each rule and builds the related answers and the conclusions. The basic algorithm used for traversing within the data trees is recursive Depth First-Search-algorithm, which is implemented. The automated knowledge acquisition process is based on the knowledge acquisition algorithm, as presented in reference [1].

### 2.4 Administration and security

A web administrator who wants to make use of the module must put the associated .cab file into the server and prepare an htm file which imports the ActiveX and shows the cab file as the CODEBASE:

```
<OBJECT
ID="DecisionMaking1" WIDTH=1039 HEIGHT=621
CLASSID="CLSID:36299202-09EF-4ABF-ADB9-47C599DBE778"
CODEBASE="http://ADDRESS/
DecisionMaking.cab"
</OBJECT>
```

For the database connection, the appropriate system DSN connection must be defined in the web server. Also the associated table must be added to the database with the design structure shown in Table 1.

Table 1

Field	Type	Collation	Attributes	Null	Default	Extra
Rec_ID	int(32)			No		auto_increment
DomainName	varchar(64)	latin1_swedish_ci		No		
Rules	varchar(2000)	latin1_swedish_ci		No		
Date	varchar(10)	latin1_swedish_ci		No		

Sample screen shots and the associated actions of the knowledge acquisition are demonstrated in Section 3, using obesity domain, as an example.

### 2.5 Graphical user interface

The Active-X is graphically a dialog-based application. For the user controls, the module does not use the build-in controls, instead the needed user controls are sub-classed and drawn similar to Windows XP look & feel for the sake of a more attractive graphical user interface.

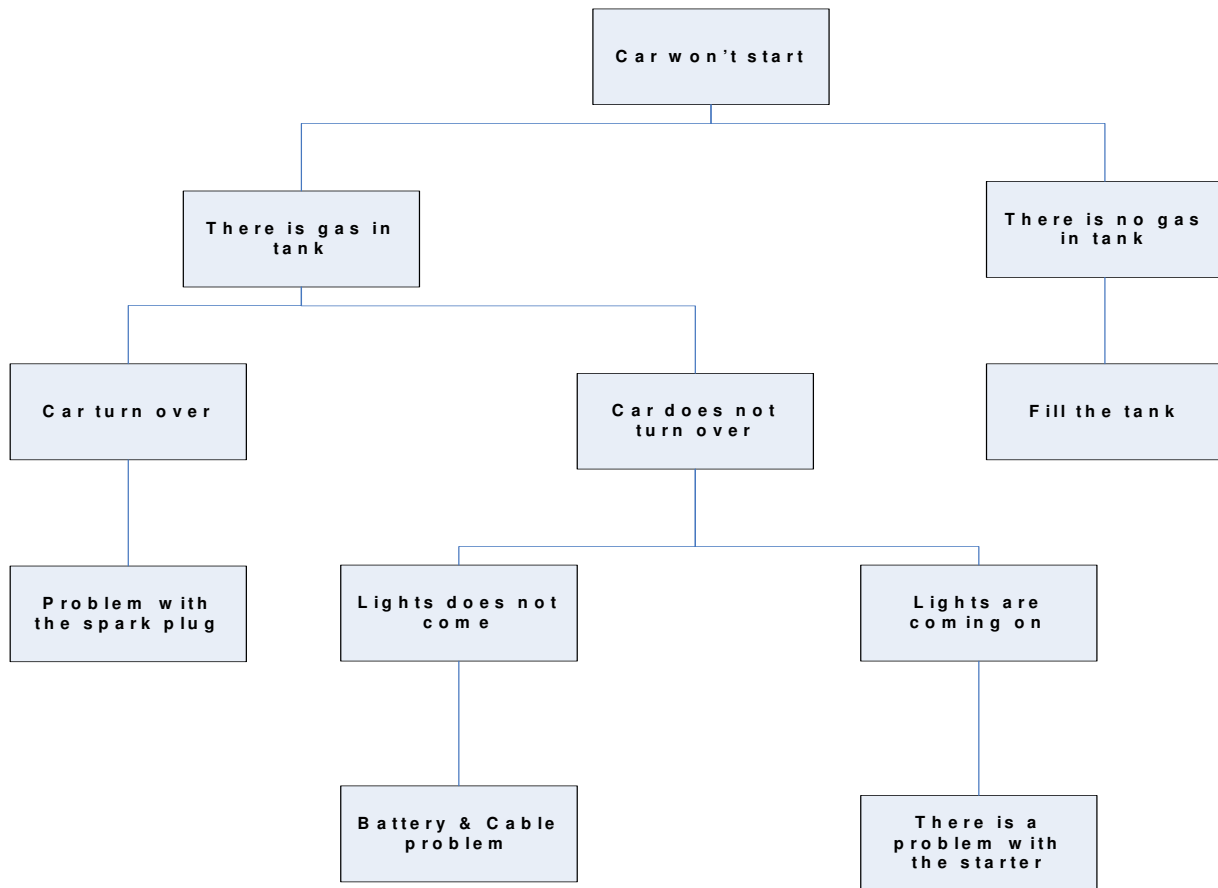
## 3 System implementation

The decision making process in a HealthCare system, consists of the diagnostic and knowledge acquisition phases. The diagnosis phase lets the expert to introduce the domain and its related branches by constructing a diagnosis tree graphically. After construction, the end leafs of the tree are considered for symptom analysis and are transferred to the next phase which is the knowledge acquisition phase. Here is an example for an obesity diagnosis tree.

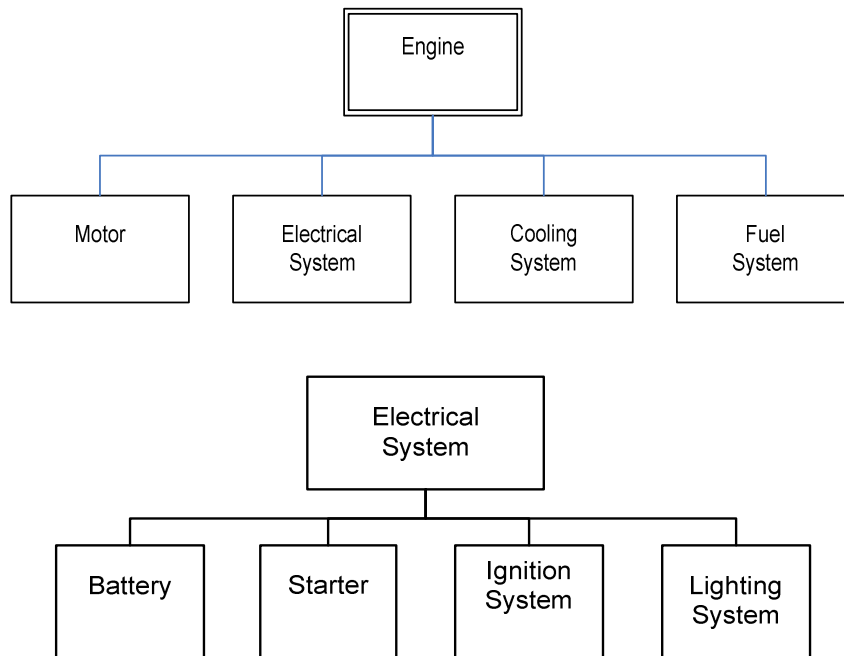
1. Obesity (domain Name)
  - a. Genetic Obesity
  - b. Behavioral Obesity
  - c. Environmental
  - d. Personal

After getting the diagnostic information, the knowledge acquisition phase considers the last open leafs of the diagnosis tree, which includes Genetic, Obesity, Environmental and Personal data. It then interactively leads the expert in the formation of the knowledge base by asking for the related symptoms, the answers for those symptoms and the conclusions. Through this knowledge acquisition process, the decision-making is defined. A decision tree with the production rules for this example could result in the following rules:

```
IF
Genetic Obesity
AND
Do relatives have it
AND
Yes
THEN
Search the family tree
```



**Figure 1. Graphical diagnostic representation**



**Figure 2a & b. System hierarchy diagrams for an automobile engine**

IF  
 Genetic Obesity  
 AND  
 Do relatives have it  
 AND  
 No  
 THEN  
 Start injection treatment

IF  
 Environmental  
 AND  
 The food is unhealthy  
 THEN  
 Relocate

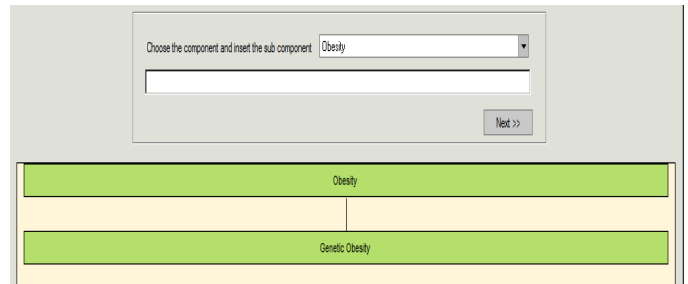
IF  
 Personal  
 AND  
 Is exercise made  
 AND  
 Yes  
 THEN  
 Check nutrition habits

IF  
 Personal  
 AND  
 Is exercise made  
 AND  
 No  
 THEN  
 Prepare regular exercise program

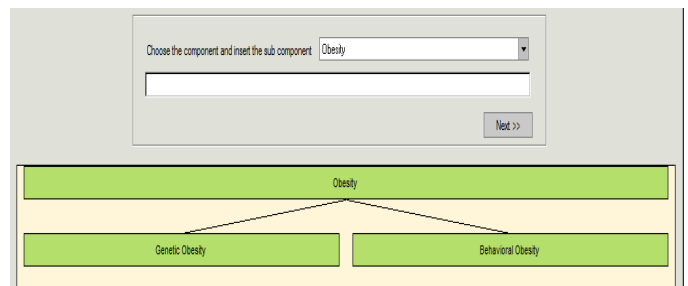
For the end user (medical expert), the screen shots with the associated actions of all knowledge acquisition are demonstrated below for the obesity domain, as an example.

**Enter Obesity as the domain name.**

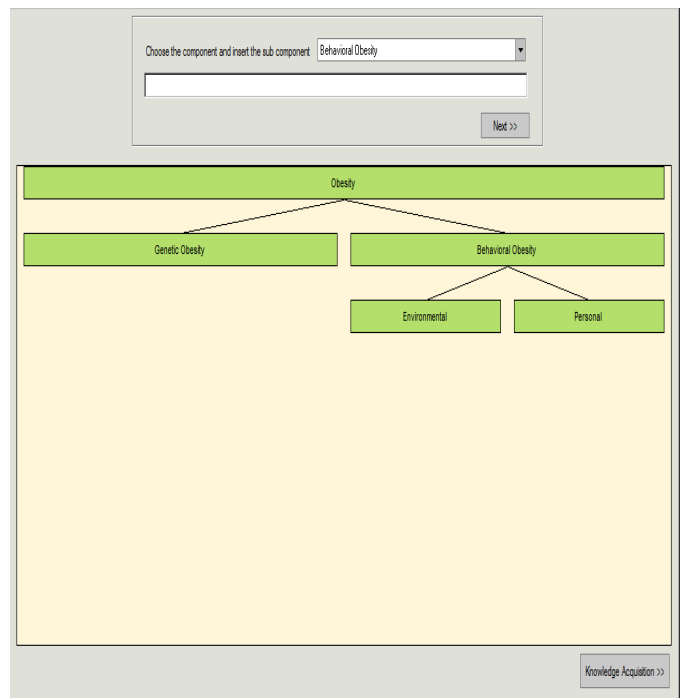
**Choose obesity from the dropdown menu to enter sub-domain.**



**Repeat and add a second sub domain for the top domain.**



**Now choose Behavioral Obesity from the combo and enter the related nodes.**



**When done with diagnosis, hit Knowledge Acquisition button and proceed.**

What is the symptom for Genetic Obesity ?



What is the symptom for Genetic Obesity ?

Do relatives have it



What is next for Genetic Obesity AND Do relatives have it Answer(s) or Conclusion ?

Answer(s)

How many answers ?

2

Yes

No

What is next for Genetic Obesity AND Do relatives have it AND Yes Answer(s) or Conclusion ?

Conclusion

What is the conclusion ?

Search the family tree

What is next for Genetic Obesity AND Do relatives have it AND No Answer(s) or Conclusion ?

Conclusion

What is the conclusion ?

Start injection treatment

The acquisition continues until all the symptoms and associated answers and conclusions are obtained from the expert, after which the decision tree is written to the database.

## 4 Conclusions and Recommendations

Expert systems represent an important set of applications of Artificial Intelligence to problems of commercial as well as scientific importance. Rule-based systems are currently the most advanced in their system-building environments and explanation capabilities, and have been used to build many demonstration programs. Most of the programs work on analysis tasks such as medical diagnosis, electronic troubleshooting, or data interpretation. The capability of current system is difficult to define. They are specialists in very narrow areas who have limited abilities to acquire new

knowledge or explain their reasoning. Technological innovations will be incorporated into expert systems as the conceptual difficulties of representation and inference in complex domains yield to standardized techniques. These will be most noticeable in the size of the computer and in the input/output of the system. Systems will use much larger knowledge bases than the few hundred to a few thousand rules now used. They will be linked electronically to large databases to facilitate inference and avoid asking questions whose answers are matters of record. Every problem area and every expert is unique. The distributed object developed for this project lets the expert's intelligence to be used within an e-HealthCare system. Because of its modularity and maintainability, it may also be used in various stages or phases of the whole solution. Further, it may be enhanced for more complicated needs and applications, in accordance with medical improvements and internet developments.

In evaluation, the software needs to allow selection of a previously identified top level symptom and recommendation. This would eliminate reentry of those nodes of the decision tree. Currently, the software allows the user to enter either another question or a conclusion, but does not show the current state of the rule being defined. The software needs to provide the logic of the rule that is being generated, so that the user can see where they are at in the decision tree.

## 5 Acknowledgements

This work was supported by the Korean Research Foundation Grant funded by the Korean government MOEHRD, Basic research Promotion Fund (KRF-2008-313-D00977)

## 6 References

- [1] Thomas M. Gatton , Malrey Lee, TaeEun Kim and Young-Keun Lee, "Diagnostic Knowledge Acquisition for Agent-Based Medical Applications," Lecture Notes in Computer Science, Springer Berlin / Heidelberg, ISSN 0302-9743, Volume 4496/2007, *Agent and Multi-Agent Systems: Technologies and Applications*, 2007, ISBN 978-3-540-72829-0, pp. 626-635.
- [2] Thomas M. Gatton, Malrey Lee, Altan Hagicumus and Yupeng Zhang, "A prototype automated knowledge acquisition system", Proceedings of the 2008 International Conference on Information & Knowledge Engineering, IKE 2008, July 14-17, 2008, Las Vegas, Nevada, USA IKE, pp.170-175, CSREA Press.