The Development of Web-Based Expert System For Diagnose Diabetes Disease Using Forward Chaining Method

Mostafa Salama R  
Dept. Communication Information System  
msr19782003@yahoo.com

Abstract - Diabetes Mellitus (DM) is a disease that affects many people who are overweight or could be a descendant of this disease. There are several classifications of diabetes, namely: DM Type1, type2 DM, other specific, and gestational diabetes mellitus. Society needs a technology to diagnose early diabetes from symptoms. In this expert system uses forward chaining method that is built using the PHP programming language and MySQL database. Knowledge Representation is useful in representing phenomena into a mathematical rule that will be stored in a database. The database consists of several tables, including: symptoms, symptoms_disease, disease, question. Knowledge Base materials sourced on the internet and medical books on diabetes. The expert system life cycle consists of feasibility analyst, conceptual design, knowledge acquisition, knowledge representation, and validation. System administrators can have a policy to make additions, deletions and changes to the data.

Keywords: PHP, MySQL, Expert System, Diabetes, Disease, Symptoms, Knowledge Base.

Almost 16 million Americans have diabetes, but about one-third of them are not aware of their condition. With the proper treatment and lifestyle changes, many of the possible complications, such as blindness, amputations, heart disease, kidney failure, and premature death, can be prevented or delayed.

Many people need a tool that can assist them quickly to diagnose the illness according to symptoms that everyone felt. Obviously, this requires a truth result of symptoms that included by patients. According to previous survey, users consider that this application will provide a good benefit in helping them, especially for people who live away from doctors and desperately need information as quick as possible. People need an easy tool for them so to build this Web-based information system is appropriate.

Web sites around the world are available and free of access with millions of applications and features. Web site can be built using html and web programming such as PHP, ASP, JSP, JavaScript, Flash, Flex and also the database management can be implemented using MySQL, Oracle, or other database management system. PHP has many advantages among other programming language, it is ease, free of use and powerful to deal with any algorithm, while MySQL is the database management system that is also easy, free of use, and reliable to manage millions of data. There for this research is dedicated to develop an application using PHP and MySQL to help users around the world to diagnose themselves to diabetes mellitus. Diagnose process is handled by the Expert System using forward chaining method, because this method will reach and find the goal from a lot of antecedent and suitable for this case.

Problem Statement
The problem for this research is how to make a web-based application to diagnose user’s possibility to diabetes mellitus. Detailed problems are: (1) Conducting Knowledge Acquisition. (2) Knowledge Representation. (3) Validation.

Objectives of the Project
The objectives of this research are: (1) Developing a Knowledge Base. (2) Developing Knowledge Representation using Database System. (3) Validating the prototype.

LITERATURE REVIEW

Diabetes Mellitus
Diabetes mellitus also called DM is a group of diseases characterized by an
elevated blood glucose level (hyperglycemia) resulting from defects in insulin secretion, in insulin action, or both. DM is not a pathogenic entity but a group of etiological different metabolic defects.

Common symptoms of DM are lethargy from marked hyperglycemia, polyuria, polydipsia, weight loss, blurred vision and susceptibility to certain infection. Severe hyperglycemia may lead to hyperosmolar syndrome and insulin deficiency to life-threatening ketoacidosis. Chronic hyperglycemia causes long-term damage, dysfunction and failures of various cell, tissues and organs.

There was several classification systems established for DM by the WHO Expert Committee on DM (1980, 1985). The current WHO classification system has been established in co-operation with the National Diabetes Data Group (USA). It is mainly based on the etiology of DM. The detailed classification are : (1). Type 1 DM. (Immune mediated, Idiopathic), (2) Type 2 DM (3) Other specific types of DM are : (a) Genetic defects of islet B-cell function, (b) Genetic defects of insulin action, (c) Diseases of exocrine pancreas, (d) Endocrinopathies, (e) Drug- or chemical- induced DM, (f) Infections, (g) Other genetic syndromes Uncommon forms of DM, (h) Uncommon forms of DM. (4) Gestational DM

Expert System

An expert system is software that uses a knowledge base of human expertise for problem solving, or to clarify uncertainties where normally one or more human experts would need to be consulted. Expert systems are most common in a specific problem domain, and are a traditional application and/or subfield of artificial intelligence (AI). A wide variety of methods can be used to simulate the performance of the expert; however, common to most or all are: 1) the creation of a knowledge base which uses some knowledge representation structure to capture the knowledge of the Subject Matter Expert (SME); 2) a process of gathering that knowledge from the SME and codifying it according to the structure, which is called knowledge engineering; and 3) once the system is developed, it is placed in the same real world problem solving situation as the human SME, typically as an aid to human workers or as a supplement to some information system. Expert systems may or may not have learning components.

Expert systems were introduced by researchers in the Stanford Heuristic Programming Project, including the "father of expert systems" Edward Feigenbaum, with the Dendral and Mycin systems. Principal contributors to the technology were Bruce Buchanan, Edward Shortliffe, Randall Davis, William vanMelle, Carli Scott, and others at Stanford. Expert systems were among the first truly successful forms of AI software.

Forward Chaining Method

Forward chaining is one of the two main methods of reasoning when using inference rules (in artificial intelligence) and can be described logically as repeated application of modus ponens. Forward chaining is a popular implementation strategy for expert systems, business and production rule systems. The opposite of forward chaining is backward chaining.

Forward chaining starts with the available data and uses inference rules to extract more data (from an end user for example) until a goal is reached. An inference engine using forward chaining searches the inference rules until it finds one where the antecedent (If clause) is known to be true. When found it can conclude, or infer, the consequent (Then clause), resulting in the addition of new information to its data.

Inference engines will iterate through this process until a goal is reached.

For example, suppose that the goal is to conclude the color of a pet named Fritz, given that he croaks and eats flies, and that the rule base contains the following four rules:

- If X croaks and eats flies - Then X is a frog
- If X chirps and sings - Then X is a canary
- If X is a frog - Then X is green
- If X is a canary - Then X is yellow

This rule base would be searched and the first rule would be selected, because its antecedent (If Fritz croaks and eats flies) matches our data. Now the consequents (Then X is a frog) is added to the data. The rule base is again searched and this time the third rule is selected, because its antecedent (If Fritz is a frog) matches our data that was just confirmed. Now the new consequent (Then Fritz is green) is added to our data. Nothing more can be inferred from
this information, but we have now accomplished our goal of determining the color of Fritz.

Because the data determines which rules are selected and used, this method is called data-driven, in contrast to goal-driven backward chaining inference. The forward chaining approach is often employed by expert systems, such as CLIPS.

One of the advantages of forward-chaining over backward-chaining is that the reception of new data can trigger new inferences, which makes the engine better suited to dynamic situations in which conditions are likely to change.

PHP

PHP: Hypertext Preprocessor, it is a server-side scripting language, like ASP. PHP scripts are executed on the server, itsupports many databases (MySQL, Informix, Oracle, Sybase, Solid, PostgreSQL, Generic ODBC, etc.) PHP is open source software

PHP is free to download and use. PHP is a powerful tool for making dynamic and interactive Web pages. PHP is widely-used, free and efficient for any kind of web pages.

PHP was conceived in 1994 and was originally the work of one man, Rasmus Lerdorf. It was developed by other talented people and has gone through three major rewrites to mature it. As of January 2001, it was used on nearly five million domains worldwide, and this number is growing rapidly. It is shown by www.php.net/usage statistic of PHP usage of 2007 has reached 20,917,850 domains with 1,224,183 IP addresses.

PHPMyAdmin

phpMyAdmin is an open source tool written in PHP intended to handle the administration of MySQL with the use of a Browser. It can perform various tasks such as creating, modifying or deleting databases, tables, fields or rows; executing SQL statements; or managing users and permissions.

Tobias Ratschiller, then an IT consultant and later founder of the software company Maguma, started to work on a PHP-based web front-end to MySQL in 1998, inspired by MySQL-Webadmin. He gave up the project (and phpAdsNew, of which he was also the original author) in 2000 because of lack of time.

By that time, phpMyAdmin had already become one of the most popular PHP applications and MySQL administration tools, with a large community of users and contributors. In order to coordinate the growing number of patches, a group of three developers registered The phpMyAdmin Project at SourceForge.net and took over the development in 2001.

PHP Designer

Php Designer 7 is more than just a powerful and lightning fast PHP IDE and PHP EDITOR - it’s also a full-featured HTML-, CSS- and JavaScript editor boosted with features so you can get your work done -- for both beginners and professional developers.

It assists you with everything from editing, analyzing, debugging to publishing websites powered by PHP, HTML, CSS to JavaScript plus other languages.

This tool supports not only working with your favorite PHP frameworks but also popular JavaScript frameworks jQuery, Ext JS, YUI, Dojo, MooTools and Prototype.

MySQL

MySQL is official software developed by Swedish company called MySQL AB, whose name was ToX Data Konsult AB. At the beginning, MySQL was using the name "mSQL" or "mini-SQL" as the interface used, apparently using mSQL was experiencing a lot of obstacles, which is very slow and inflexible. Therefore, Michael Widnies ("Monty"), hir nickname, try to develop the interface and finally MySQL were developed. At the moment, MySQL was distributed specifically for noncommercial purposes and it was free, while for commercial purpose it required to pay the license. Since version 3.23.19, MySQL has categorized as GPL licensed software, which can be used without payment. (John, 2008)

MySQL can be executed on different operating systems such as Linux, Unix, Windows. The advantage of MySQL is the speed of access, cost, configuration, availability of MySQL source code under Open Source License and being the popular open source database. MySQL is a
relational database software (usually called Relational Database Management System or RDBMS), as other database Oracle, PostgreSQL, ASP, etc., which has fast execution of SQL in multithreaded and multuser environment. Therefore, because of its high potential to be used as a reliable database, all supporting features must develop continually to optimize the usage of MySQL.

CONCEPTUAL DESIGN

Knowledge Representation

Knowledge Representation (KR) research involves analysis of how to accurately and effectively reason and how best to use a set of symbols to represent a set of facts within a knowledge domain. A symbol vocabulary and a system of logic are combined to enable inferences about elements in the KR to create new KR sentences. Logic is used to supply formal semantics of how reasoning functions should be applied to the symbols in the KR system. Logic is also used to define how operators can process and reshape the knowledge. Examples of operators include negation, conjunction, adverbs, adjectives, quantifiers and modal operators. Interpretation theory is this logic. These elements--symbols, operators, and interpretation theory--are what give sequences of symbols meaning within a KR.

Suppose that $P_1, P_2, ..., P_n$. Where is $P_1$ ($\sum_{i=1}^n i$ ) are the symptoms of diabetes mellitus type 1. These facts can be represented by logical implication of discrete mathematics,

\[ \text{IF } (P_1 \land P_2 \land P_3 \land ... \land P_n) \rightarrow \text{Diabetes Mellitus Type 1 (Conclusion)}. \]

Database System

Traditionally, knowledge representation systems have been implemented using main memory as the primary run-time storage medium. As a result, the size of a computer's internal memory has placed an upper bound on the size of the knowledge bases that could be loaded. Database management systems, on the other hand, use external storage (disk) as the primary run-time storage medium. In contrast to knowledge representation systems, the upper bound on the size of a database is not limited by the size of main memory, but rather by the size of external storage. This allows databases to grow to nearly unbounded size given the cost-space trade-offs of disk vs. main memory.

Knowledge representation systems, however, have strengths unavailable to normal database systems. They allow a complex structural representation of the data (knowledge) that allows inferencing and complex query evaluation to be performed. In addition, relations between data can be recorded as "abstract" relations, allowing knowledge about functional relationships, relations between groups of entities, and other such higher-level connections between the "raw" data elements. MySQL is used while recovering the advantages of of DBMS's, in particular the efficient storage and retrieval of massive amounts of data.

MySQL a high performance knowledge representation system that deviates from the norm by using a database-management system to provide run-time storage advantages. As a KR system, MySQL supports the same representational functionality as its memory-based predecessor.

Through the use of a hybrid KB/DB, the user can issue various complex and interesting queries against a database in a manner much simpler than that required to issue queries of the same meaning using a stand-alone DBMS. In particular, the complex inferencing capabilities of MySQL can be utilized to formulate interesting and potentially useful conjunctive queries. In addition, data mining, the search for knowledge in databases, the data-mining system can find relevant relationships in the DB using taxonomical and other semantic knowledge present in the KB.

Database design will consist of four tables: (a) symptoms, (b) disease, (c) symptoms disease, (d) questions.

Forward Chaining Method

Inference engines will iterate through this process until a goal is reached.

For example, suppose that the goal is to conclude which type of diabetes illness, given that he has high blood sugar and
diabetes historical family, and that the rule base contains the following rules:

If X has high blood sugar - Then X is a diabetes illness Type 1

If X has diabetes history - Then X is a diabetes Type 2

If X has diabetes - Then X is high blood sugar

If X has diabetes - Then X is maybe has historical family

This rule base would be searched and the first rule would be selected, because its antecedent (If has high blood sugar) matches our data. Now the consequents (Then X is a diabetes illness Type 1) is added to the data. The rule base is again searched and this time the third rule is selected, because its antecedent (If X has diabetes) matches our data that was just confirmed. Now the new consequent (Then X is high blood sugar) is added to our data. Nothing more can be inferred from this information, but we have now accomplished our goal of determining the type of diabetes.

One of the advantages of forward-chaining over backward-chaining is that the reception of new data can trigger new inferences, which makes the engine better suited to dynamic situations in which conditions are likely to change.

Managing the Knowledge Base

A Knowledge Base (abbreviated KB) is a special kind of database for knowledge management, providing the means for the computerized collection, organization, and retrieval of knowledge. Also a collection of data representing related experiences, their results are related to their problems and solutions. KB consists of three type : (1) Machine-readable knowledge bases, (2) Human-readable knowledge bases, (3) Knowledge base analysis and design.

In this research, collected information related to the symptoms of the disease and the relationship between diseases that occur. Information about the symptoms of the disease is obtained from a medical book or from doctors who are experts in diabetes. Furthermore, the data obtained will be stored in the table, which has several tables, among others: (1)symptoms, (2)disease, (3)symptoms_disease, (4)questions.

RESEARCH METHODOLOGY

Expert systems are a piece of software. Expert systems are similar to conventional software in some respects, but different in others. To build expert systems, we must merge new ideas with traditional ones. Often expert systems are only one part of a larger software system. Thus the development of the expert systems has to be coordinated with other software development efforts.

Thus some software engineering methods also apply to expert systems and features of the software lifecycle can also be observed in expert systems. We can draw analogies between the phases in expert systems construction and corresponding phase in software development.

Feasibility Analysis.

The domain in which the expert system is to operate and the task which will be performed by the expert systems are studied and analyzed by the expert systems builder.

Identification of an appropriate task is a vital step in the development of any expert systems. The vague sense that an expert system might be useful in a particular knowledge domain is not, in itself, sufficient reason for justiﬁcation of the efforts required in building an expert system. The decision about whether to build an expert system should be based on a specification of what tasks will be performed by the expert
system and whether or not they meet appropriate needs.

**Conceptual Design**

The conceptual structure of the system is defined, along with a specification that describes the way in which the expert system will carry out the task.

In general, the next step, namely, knowledge acquisition, is facilitated by having a good idea of the overall structure of the system. A good conceptual design tells the knowledge engineer what to look for and can be used to decide which issues are important and which are not.

**Knowledge Acquisition**

The knowledge required for performing the task is acquired from human experts, case histories, references source, etc.

This phase deals with the task of obtaining knowledge and formalizing it so that it can be included in the expert-system knowledge base. Since expert systems rely heavily on the quality of the knowledge they process, knowledge acquisition is a crucial part of the expert system construction process.

Knowledge Acquisition for diabetes mellitus taken from several books and consulted with the doctor that has experiences on diabetes mellitus.

<table>
<thead>
<tr>
<th>Code</th>
<th>Disease</th>
<th>Name of the Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td>Type 1 DM</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>Type 2 DM</td>
</tr>
</tbody>
</table>

As an example when the application started, it will provide some question to the user, i.e.:

- **Are you smoking? Yes/No.**
- **Please write down your weight:**
- **Please write down your height:**
- **How old are you?**
- **etc**

These questions depend on the symptoms of the disease.

**Knowledge Representation**

The knowledge is formalized and represented within a symbolic program so that it is executable by the inference engine.

Knowledge must be expressed in the knowledge representation method and the language of the expert systems tool used for building the expert system. This task is facilitated when the form of the knowledge that is captured from the expert correspondence closely to the knowledge structure that is available in the expert system tools.

**Validation**
User’s views, expert opinions, or operational criteria are used to determine whether the expert system has achieved and acceptable degree of success.

Intuitively, the aim of a validation effort for any system is to answer the following question: “Does the system works?” In the case of complex systems, however, one should first ask the following question: “Does the system works?” the basic axiom relating to the validation of complex system states: “Complex systems cannot be evaluated or validated by using simplistic criteria”. The expert system should thus be evaluated using a range of validation guidelines.

KNOWLEDGE ACQUISITION

Expert Knowledge of Diabetes Mellitus

The diagnosis of diabetes carries considerable consequences and should therefore be made with confidence. If the patient has classical symptoms (such as increased thirst and urine volume, unexplained weight loss, pruritus vulvae or balanitis) or drowsiness or coma, associated with marked glycosuria, the diagnosis can be readily established by demonstrating fasting hyperglycaemia. If the fasting blood glucose concentration is in the diagnostic range shown in table below, an oral glucose tolerance test (OGTT) is not required.

Selecting the right Knowledge Representation

When applied in the computer domain, knowledge representations range from computer-oriented forms to conceptual ones nearer to those present in our internal world models. Five knowledge levels can be established using this criterion:

- Implementational
  - this is the more computer aware level. It includes data structures such as atoms, pointers, lists and other programming notations.

- Logical
  - symbolic logic is inside this level. Thus, symbolic logic propositions, predicates, variables, quantifiers and Boolean operations are included.

- Epistemological
  - a level for defining concept types with subtypes, inheritance, and structuring relations.

- Conceptual
  - the level of semantic relations, linguistic roles, objects and actions.

- Linguistic
  - the more computers distant level, it deals with arbitrary concepts, words and expressions of natural languages.

Knowledge representation for the application of expert system on this research is using rule-based knowledge (logic), it is according to several research conducted before and the method we are using. Based on this knowledge base and the problem set, forward chaining method using rule base for knowledge expert system is efficient, it can be clearly seen that a set collection of symptoms has been stored to database before has a unique pattern to solve the problem.

From every symptoms question has been answered by user, rule base take an important part, it operate premis/antecedence each symptoms into mathematical logic using the rule on discerete mathematics.

EXPERT SYSTEM DESIGN

After doing knowledge acquisition, knowledge will be store at database, on this research we use MySql as database system. MySql database will act as a high performance knowledge representation system that deviates from the norm by using a database-management system to provide run-time storage advantages. As a knowledge representation system, MySql supports the same representational functionality as its memory-based predecessor.

Rule Table
admin administrator. At user level, will use several function describe below:

**DM Forward Chaining**
Forward chaining method located on file Try.application.php. on these process, user have to answer some question, from DM general symptoms until specific question, they have to answer by selecting the answer 'yes' or 'no' and cannot both of them.

From the answer, system will redirect to the next question properly. Sure, the question shows will related to several answer previously.

There are several function used for admin:

**Adding data**
To add disease and symptoms of the disease.
(a) Symptoms, this function located on symptoms.management.php. Function will receive admin data about new symptoms and add to database, located on symptoms table.
(b) Disease, this function located on disease.management.php. Function will receive admin data about new disease and add to database, located on disease table.

**Deleting data**
Deleting data consist of deleting on disease and symptoms tables.
(a) Symptoms, this function located on symptoms.management.php. Function will receive the $_GET parameter to delete id_symptoms from passing parameter that had been choosen.
(b) Disease, this function located on disease.management.php. Function will receive the $_GET parameter to delete id_disease from passing parameter that had been choosen.

**Updating data**
Updating data consist of symptoms and disease.
(a) Symptoms, this function located on symptoms.management.php. Function will receive the $_POST parameter to update with key id_symptoms from passing parameter that had been choosen.
(b) Disease, This function located on disease.management.php

Function will receive the $_POST parameter to update with key id_disease from passing parameter that had been choosen.

**Table of Rule**

<table>
<thead>
<tr>
<th>Field</th>
<th>F.Type</th>
<th>F.size</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule no</td>
<td>Numeric</td>
<td>2</td>
<td>Unique, Primary Key</td>
</tr>
<tr>
<td>P1</td>
<td>Text</td>
<td>100</td>
<td>Symptoms</td>
</tr>
<tr>
<td>P2</td>
<td>Text</td>
<td>100</td>
<td>Symptoms</td>
</tr>
<tr>
<td>P3</td>
<td>Text</td>
<td>100</td>
<td>Symptoms</td>
</tr>
<tr>
<td>P4</td>
<td>Text</td>
<td>100</td>
<td>Symptoms</td>
</tr>
<tr>
<td>P5</td>
<td>Text</td>
<td>100</td>
<td>Symptoms</td>
</tr>
<tr>
<td>P6</td>
<td>Text</td>
<td>100</td>
<td>Symptoms</td>
</tr>
<tr>
<td>P7</td>
<td>Text</td>
<td>100</td>
<td>Symptoms</td>
</tr>
<tr>
<td>P8</td>
<td>Text</td>
<td>100</td>
<td>Symptoms</td>
</tr>
</tbody>
</table>

**Designing the Process**
Process designing will explain about how system works to manage input data to the output data with several function provided before. It can be clearly seen that the application will be used by two kind of user; there are administrator and the common user.

For Administrator privileges, there are some menu like : add, delete, and update.

**Function of the System**
In the system application, it uses a lot of different function for user and
User Interface Design

User interface is the way to communicate with user and administrator. There are two kinds of user interface: (1) For general user, this kind of user will find the information from the system, like the disease according to the symptoms that had been inserted. (2) For administrator, this kind of user responsible to the content of system, from storing knowledge base until conclusion of application.

IMPLEMENTATION

Disease Management

Symptoms Management

Mapping Disease and Symptoms

Rule Management

Try Application

CONCLUSION

This research focusing on Conducting Knowledge Acquisition, Knowledge Representation and Validation. Expert system using rule base with forward chaining, the inference engines will iterate through this process until a goal is reached. Knowledge Base consists of three type: (1) Machine-readable knowledge bases, (2) Human-readable knowledge bases, (3) Knowledge base analysis and design. Knowledge Acquisition for diabetes mellitus taken from several books and consulted with the doctor that has experiences on diabetes mellitus. Knowledge must be expressed in the knowledge representation method and the language of the expert systems tool used for building the expert system. Knowledge representation for the application of expert system on this research is using rule-based knowledge (logic), it is according to several research conducted before and the method we are using.

REFERENCE


Fuecks, Harry., 2003. The PHP Anthology, Sitepoint, Australia


Mira, Ali., 1998. Methodology and Tools in Knowledge-Based Systems, 11th International Conference IEEA, Spain
