An Expert System for HIV and AIDS Information

A.N. Masizana-Katongo, T.K. Leburu-Dingalo, D. Mpoeleng

Abstract— **We propose to develop an expert system that will provide general information on HIV and AIDS to the public of Botswana. This project is motivated by the Microsoft sponsored project conducted at the Department of Computer Science named IHISM for which the authors are members. IHISM aims to contribute to the digital divide by developing an HIV and AIDS public information portal accessible through mobile phones. The case study is based on Botswana, one of the leading countries hard hit by the HIV and AIDS pandemic where the government is working hard to facilitate the provision of education and raising awareness concerning the pandemic. Whilst IHISM is concentrating on the digital divide and mainly concerned with semiliterate people in rural settlements this alternative is for those who have access to the internet and hence can use the system wherever they are. The proposed system will act as an online 'expert' and is aiming at supporting the initiative of providing accurate information to the general public.**

Index Terms— **Expert Systems, HIV AND AIDS**

I. Introduction

In January 2005 the Department of Computer Science, University of Botswana received funding of $43,000 for the IHISM project from Microsoft, USA, as part of their Digital Inclusion initiative [1] which aims to bridge the gap between relatively well-developed and less-developed regions in the world. The project termed, IHISM [2] (Integrated Healthcare Information System through Mobile telephony) aims to explore the use of mobile phones as an access technology to a variety of HIV AND AIDS related information to the required by the general public. The disease has affected every segment of Botswana society and according to UNAIDS estimates [3], one-third of Botswana's sexually-active population between the ages of 15 and 49 (out of a total population of 1.5 million) are infected with the virus that causes AIDS, the highest rate in the world.

However, Botswana is also one of the leading countries at the front line of the war against AIDS and the government has demonstrated a very high level of political commitment to addressing the HIV and AIDS epidemic The government has not only formed various partnerships to combat HIV and AIDS [4 - 7] but is also committed to inform the nation about the pandemic by introducing various education and communication programs [8, 9] These programs aim at raising awareness of the diseases and offer relevant information to the general public on the dangers and management of HIV and AIDS pandemic. This response has seen Botswana exhibit high prevalence over HIV and AIDS infection as the general awareness of the dangers of the disease has been raised among the population. The government also provides public handouts which are distributed by the ministry of health as well as the hosting of public events such as HIV/AIDS fairs and pageants to try to educate and sensitize the public about the disease.

The programs described give a clear indication that the complexity and incurability of HIV/AIDS requires that the government in addition to providing clinical support, needs to also put an emphasis on the dissemination of information about the disease, prevention methods, therapeutic methods and psychological support. In this age it is evident that one of the best ways to do this is through the utilization and provision of ICT technologies.

To support this initiative, the IHISM project has pledged to develop a general HIV and AIDS information portal which would be queried by the users using mobile phone technology. The information service portal would allow the general public to request for information on topics related to HIV and AIDS such as descriptions, infection, testing, counselling and support, opportunistic diseases and paediatric care etc. The portal represents this information in the form of Frequently Asked Questions (FAQ) service where the user inputs a query on any of the subjects. The project is on-going and has achieved thus far a milestone of deliveries and has influenced a number of research projects within the department.

This project has emerged as a by-product of the IHISM concept by developing a system which provides similar services in the form of an expert system to the public with internet access. The system is meant to act as an online 'expert' in HIV and AIDS information such that that some information may have to be derived through inference as opposed to simple data retrieval. The system will accept as input an FAQ from the user and provide the most relevant answer to the question. Dissemination of accurate information about HIV is an essential element of the national AIDS prevention strategies and thus information service provided by this system has to be accurate. The challenge is that users may ask the questions differently in pursuit of the same answer and the system should be able to systematically

Dr A.N. Masizana-Katongo Computer Science Department is a Lecturer at the University of Botswana, Private Bag 0022 Gaborone, Botswana, 267 355 2865 fax: 267 318 5097; e-mail: masizana@ mopipi.ub.bw.
T.K. Leburu-Dingalo is an MSc student with the department; e-mail: leburub@ mopipi.ub.bw.
Dr D. Mpoeleng is also a Lecturer at the department, e-mail: mpoeleng@ mopipi.ub.bw.

analyse the questions and provide an appropriate answer. The main challenge is to determine the various forms in which a typical FAQ question could be mapped to the relevant answer. Another challenge is to determine the degree to which the answer is relevant to a particular question. To address the former we derive various keywords for any input question and to address the latter we employ the use of certainty factors offered within the Corvid Exsys Expert system developer [10] used to develop the system.

II. HEALTH EXPERT SYSTEMS

The application of IT research and development to support health and medicine is an emerging research area with significant potential. Major initiatives to improve the quality, accuracy and timeliness of healthcare data and information delivery are emerging all over the world. The Agency for Healthcare Research and Quality (AHRQ) [11], of the US Department of Health Services (HHS), awarded grants and contracts to promote the use of health information technology. Computerized systems including expert systems have been used to carry out efficient and effective data processing on complex problems to support various problem domains since the 1970's. Jackson[12] defines an expert system as a "…computing system capable of representing and reasoning about some knowledge–rich domain, such as internal medicine or geology, with a view to solving problems and giving advice.". Since the advent of artificial intelligence in the 1970's which saw the birth of expert systems, various domains have taken advantage of this technology. The most popular application has been in the area of health and medicine. MYCIN [12], developed in 1970 at the Stanford University, is one of the most popular medical expert system used to assist to diagnose and treat blood diseases. MYCIN was the pioneer in demonstrating how a system can be used to successfully perform medical diagnosis.

Another Early expert system is the PACE(Patient Care Expert System (Pace) [13] which was conceived in 1977 with the purpose being to make "intelligent selections" from the overwhelming and ever changing information related to health in order to facilitate patient care. The system started off as an educational system for the nursing profession. Throughout the years the system evolved and went through many development generations to a point where it became an advanced clinical management system capable of supporting the entire health care field to diagnose and care for patients with pulmonary diseases.

Another expert system called, MITIS system, was developed in 2004 at the National Technical University of Athens. The MITIS system was developed to assist in the management and processing of obstetrical, gynaecological, and radiological medical data [14]. The concept behind this system is to record and store information from experts in medical departments of gynaecology, radiology and obstetrics to provide a centralized mechanism for managing patient information within and outside a hospital.

HIV and AIDS information is required to be accurate, timely and easy to understand. As it is essential that HIV and AIDS information distributed should be from an expert source, expert systems are a preferential way to undertake this task. The advent of HIV and AIDS has prompted the development of knowledge base systems in this problem domain. Most of these systems are still at research and developmental levels. In the recent years, researchers are developing Question and Answer systems on HIV and AIDS to assist mainly medical practitioners. Most of these systems differ in the way they were developed and take different forms of HIV and AIDS expertise delivery to the intended users including online. Atalay, B. *et al.* [15] developed a system that diagnoses HIV-patients and prescribes the correct drug regimen for them. Temesgen *et al.* [16] developed a Questions and Answers support system named CHESS which provides answers to questions logged in by people living with HIV and AIDS infection. Sloot *et al.* [17] employs grid technology to develop an HIV expert system which delivers expert knowledge accessed from distributed databases of various infectious diseases.

Another system is the Customized Treatment Strategies for HIV (CTSHIV) [18], a knowledge-base system for management of HIV-Infected Patients. CTSHIV expert system was developed at the University of California Irvine as a tool that "…recommends individualized treatment strategy for HIV patients…" The concept behind the system is to analyze the strains of HIV in a patient and find out what antiretroviral agent the strain is resistant to. Based on this information the system is then able to determine which antiretroviral drugs are appropriate for that particular patient. Although CTSHIV is aimed at physicians as a diagnostic tool it does have the advisory capabilities in that it refers users to more information in respect to recommendations being generated.

This project proposes a more general version of the CTSHIV expert system which would provide HIV and AIDS information to the public of Botswana based on the FAQ logged in by the public.

III. HIV AND AIDS EXPERT SYSTEM

In describing the basic concept of an expert system, Giarratano Riley [19] describes three components being the user, who supplies facts and information to the expert system or receives expert advice from the system, the knowledge base which contains knowledge, and finally the inference engine that uses rules and the knowledge base to draw conclusions in response to a user's query.

The proposed expert system consists of basically the same components. The system consists of three main components; the user interface, inference engine and knowledge base. The user interacts with the system through the interface that consists of graphical screens that allow the user to type in a question and display responses from the system. The variables that are used at the user interface are the user question and answer. The user variable is used with the question logic block to ask a question while the answer is used by the results block to publish the response to the user.

The answers evaluated for this task are stored in the knowledge base. The knowledge base is built from mainly Frequently Asked Questions (FAQ) and answers manual of a local HIV/AIDS information call center, Ipoletse, [9]. Ipoletse in Setswana means 'Ask for yourself', meaning if you need information just pick the phone and have the information given to you. The call center was set up by Botswana government in 2002 with a mission to provide free information on HIV and AIDS to the public. The manual consists of over 200 questions and answers on HIV/AIDS related information. This knowledge base also acts as a store for rules used by the inference engine in the extraction of keywords and a meta block that holds the answers and their ranks.

In order to extract the keywords for each question and answer pair, an online survey was set up using a web survey tool called Limesurvey [20]. A sample of a 100 students at the University of Botswana were provided with the questions from the manual and asked to provide another way of asking the same questions so that a sample of different words could be mapped to a particular question and mapped onto a relevant answer. As an example in asking the question,

Q_1'what is HIV'. A_1: HIV is......

The input from the sample users offered variations of the same questions such as 'what do you '***understand***' about HIV; ***'describe'*** HIV, What is ***meant*** by HIV etc. The statistical analysis of the extracted words was performed in order to derive the most frequently used words (FUW). In the case of this question the FUW set include ***define, do, does, understand, describe, and explain*** etc. with the given percentages as in **Fig. 1**. The chart shows that for this particular question, HIV is pretty much the main keyword and a default word followed in some capacity by the verbs like do, does and adverbs like describe, define, understand and describe.

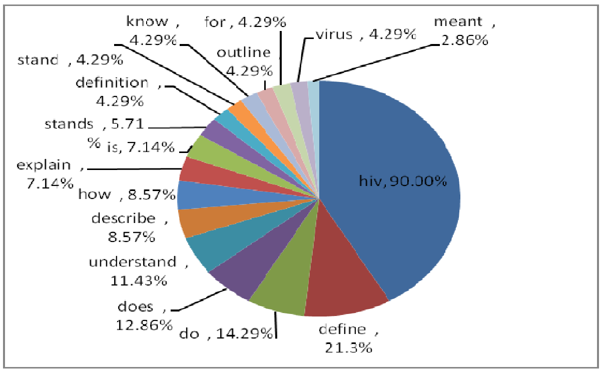

Fig 1: Percentage FUW for Q_1

By employing significant FUW's we derive phrases that are in fact variations of the possible combinations of the words that can be accepted as the question. A keyword would comprise of one or more combination of question default words, tags (HOW, WHAT) and FUW. In order to provide flexibility, each question is allowed to have several keywords to map to the correct answer.

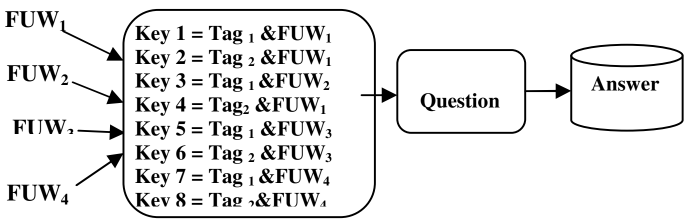

Fig 2: Keyword Extraction

However, particular keywords may be relevant to a question and the answer A_1 may provide a degree of explanation but that would not map directly to the essence of the answer and hence differing in a degree of relevance. Hence in constructing keywords the answer retrieved must be attached with a degree of relevance to the query or question entered. **Fig.2** shows how keywords could be derived from four FUW and two question tags resulting in eight keywords.

The inference engine can be considered the brain of the system. It is responsible for the "reasoning" of the system. In the proposed expert system, the inference engine extracts keywords from input by the user and processes them to check the validity of the question asked. The system is rule based and thus implements inferencing by utilizing the IF THEN rule to draw conclusions as to which answer is to be retrieved for a relevant query or question. To reach a recommendation the system uses the encoded information of keywords to determine if they qualify for any answer in the system. To make this assessment the weight process logic block utilizes confidence variables to rate the relevancy of keywords combination to any particular input or query.

IV. DEVELOPMENT OF THE SYSTEM

The system was developed using Exsys CORVID development tool [10] l. This product provides an advanced environment to develop Knowledge Automation Expert Systems in various domains and has good features to allow web applications. The inference engine uses the problem-solving logic to emulate the decision making of domain experts. It allows the domain experts to easily describe their decision-making steps in a logical manner using tree-structured logic diagrams described as rules. People interact with the system as if they were talking to the expert and in this case an expert in HIV and AIDS knowledge. The rules of the system are defined, organized and structured using one or more logic blocks in tree-structured diagrams. The logic may be a complex branching tree that systematically covers all possible input cases, or a simple diagram that correlates with a few rules. Logic Blocks are created and maintained in a visual, intuitive development environment using the IF/THEN representation.

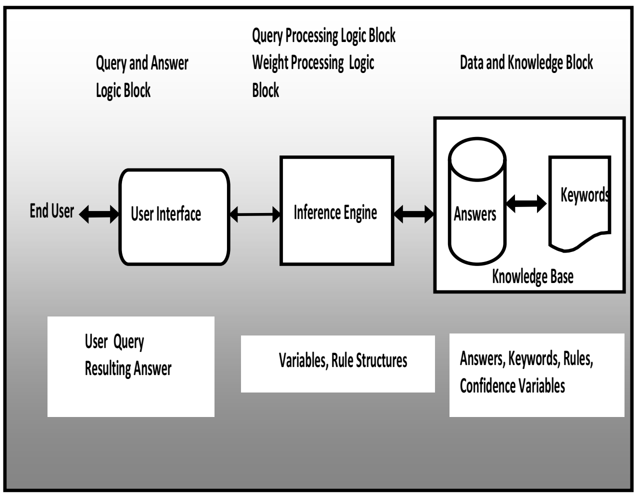

Fig 3: Components of the Expert System

Fig.4 below demonstrates the main processes of user input, processing and system response. First the user types in a question at the system interface. The system uses rules to assess if the question's keywords match those in the system. If a match is not found the system outputs a no answer response. If a match exists in the system the keywords are assigned weights referred to as confidence variables. The weights for all identified keywords are then summed. This sum is used to measure how accurate the question is, relative to predefined real life user questions.

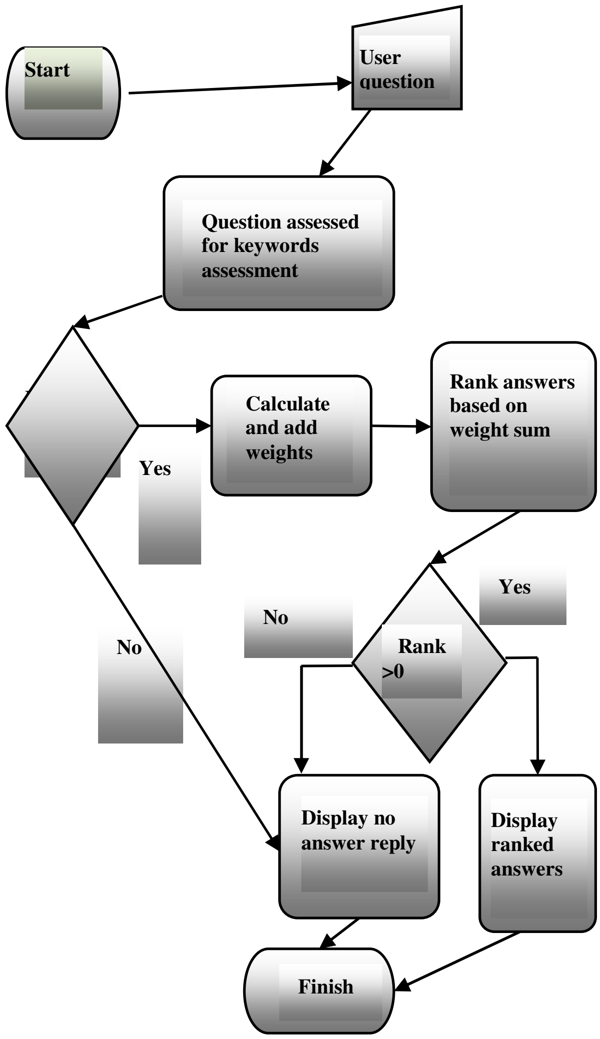

Fig 4: System Flowchart

Therefore if a combination of keywords closely resembles a predefined question then the sum for that combination would be high. The system then accesses answers that are related to the predefined question and ranks the answers based on the weighted sum. If none of the answers get a ranking of over 0 then the system outputs a no response reply. Otherwise the system outputs a list of answers ranked according to their relevancy to the question.

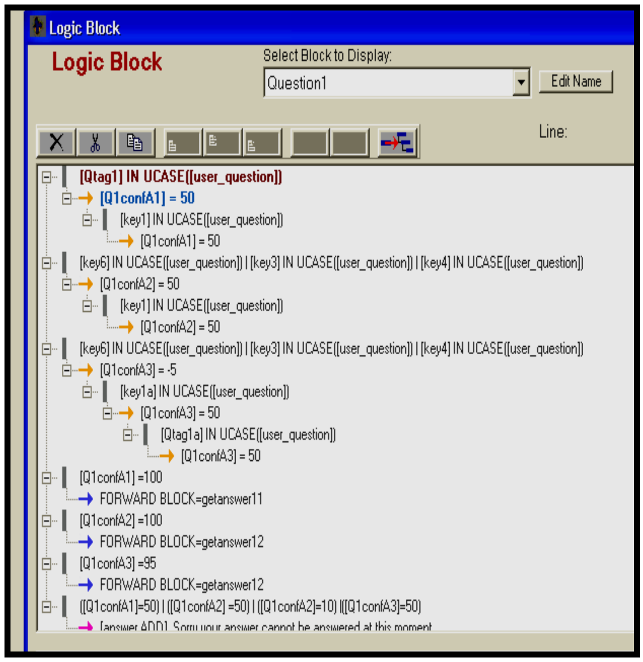

Fig 5: A Typical Logic Block Showing Question 1 Variations

V.SYSTEM INTERACTIVTY

A sample of twenty students was requested to interact with the system to verify its efficiency and accuracy. The diagram on **Fig. 6** shows the user interface that will be presented when the user launches the system. User interaction screens are designed using HTML templates that allow the full power of HTML to be utilized to create highly interactive but easy to use interfaces. The diagram shows a sample of the interface where the user enters the query and clicks on the submit button. The user is allowed to enter the long question as they understand it and the system is then supposed to extract the keywords from it.

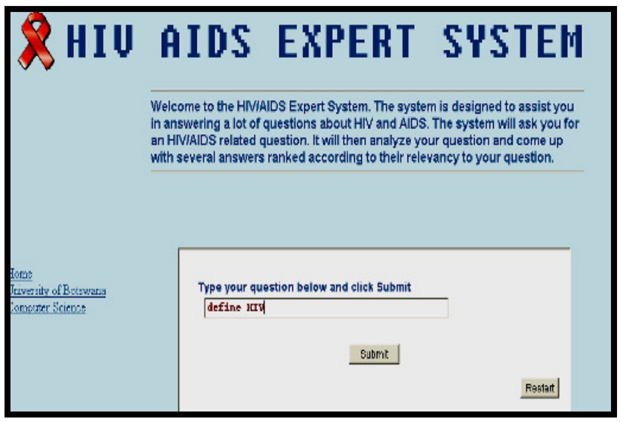

Fig 6: User Input Interface

As shown in **Fig. 7**, the retrieved answer is sometimes the selection of a single answer or a list of possible answers attached with a degree of relevancy to the question asked arranged in their order of certainty. The program on request, can explain how it arrives at its conclusion. The input question is ‘Define HIV’ and on analysing it, using **Fig. 1**, we realise that these are amongst the FUW for ***A1***. Hence the system can convert these two words to become a keyword formed from the two FUW’s, Define (21%) & HIV (90%), perhaps the most relevant keyword for this question. Therefore the first answer A1 is a perfectly relevant one with 100% certainty as shown in **Fig. 7**. The next retrieved answer has a measure of relevancy to the input question but less than ***A1*** and so on.

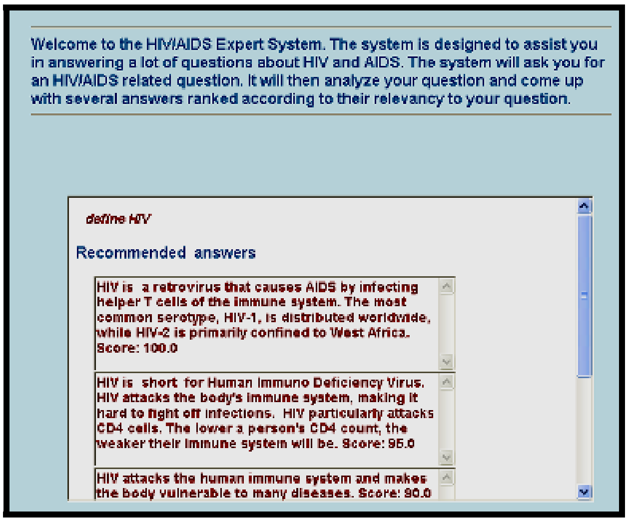

Fig 7: Sample User Interface Output Screen

More than 90% of the participants agreed that the expert system was not only easy to use but was also able to almost always provide a relevant answer to the input question. They also felt comfortable asking questions and felt the system could be very useful in their own HIV and AIDS health care support. The results not only give more information to the input question but also show the relevancy of the answer to the question and hence asking a question can lead to other relevant important question that you did not ask for. Most importantly the system indicates that an HIV and AIDS question can be derived using keywords which the question represented accurately in most of the cases.

VI. CONCLUSION

An Expert System has been developed and a sample of participants used to build the data and test the effectiveness of the system. The results show that the system has been developed successfully and found to be a good system to support the public with timely information on HIV and AIDS. However, we cannot rule out the possibility of some input data being outside the collected data, thus rendering the knowledge base incomplete and more likely not to provide the expected feedback to users in some cases. As a result, there is a need to simulate other variations of the FUW’s not provided by the sample. We therefore recognize the need to develop a procedure to derive further guesses which could result in a great deal more guesses for expanding the knowledgebase. The system is also intended to be exhaustive and thus the knowledge base is going to expand phenomenally over time as various other sources of HIV information are used. We therefore recognize the potential research areas which include designing a more formal or algorithmic keyword extraction method for the questions.

REFERENCES

[1] Microsoft Digital Inclusion Program http://research.microsoft.com/ur/us/fundingopps/rfps/DigitalInclusion_2005_RFP.aspx accessed 17 February 2009.
[2] IHISM Project website www.ub.bw/ihism accessed 18 February 2009.
[3] UNAIDS/WHO 2006 Report On The Global AIDS Epidemic.
[4] Botswana National Strategic Framework for HIV AND AIDS", NACA, 2003.
[5] The NACA Project, http://www.naca.gov.bw/about.htm Accessed 18 February 2009.
[6] African Comprehensive HIV AND AIDS Partnerships (ACHAP), http://www.achap.org/profile.html Accessed 21 February 2009.
[7] Botswana Harvard Partnership http://www.hsph.harvard.edu/bhp/ Accessed 21 February 2009.
[8] MASA - The National Antiretroviral Therapy Programme of Botswana, Botswana Ministry of Health, http://www.moh.gov.bw/index.php?id=192 , Accessed 21 February 2009.
[9] IPOLETSE Call Centre for HIV AND AIDS, Botswana Ministry of Health, http://www.mri.co.bw/html/call-ipoletse.htm, Accessed 21 February 2009.
[10] EXYS CORVID Development Environment http://www.technology-reports.com/company.asp?id=163, Accessed 21 February 2009.
[11] Agency for Healthcare Research and Quality (AHRQ), http://www.ahrq.gov/RESEARCH/hitfact.htm, Accessed 20 January 2009.
[12] Jackson, P. “MYCIN: Medical Diagnosis using Production Rules”. Introduction To Expert Systems. Addison-Wesley Publishing Company, (1986):(102-114)
[13] Evans, S. “PACE-Then and Now” The Pace System: An Expert Consulting System For Nursing, Springer,(1996): (26-32)
[14] Matsopoulus, G. K, et al. MITIS: a WWW-Based Medical System for Managing and Processing Gynecological-Obstetrical-Radiological Data. Computer Methods and Programs in Biomedicine, Volume 76, Issue 1,(October 2004): Pages 53-71
[15] Atalay, B. Potter, W.D. Haburchak, D, HIVPCES: a WWW-based HIV patient care expert system, Computer-Based Medical Systems, 1999. 12th IEEE n, 214-219.
[16] Temesgen, Zelalem, et al. A Comprehensive Health Enhancement Support System (CHESS) for People with HIV Infection, Springer, AIDS and Behavior, Volume 10, Number 1, January 2006 , pp. 35-40
[17] Sloot, A. V. Boukhanovsky, Keulen, C. A. Boucher, A grid-based HIV expert system, IEEE Computer Society, Volume 1 , 471 - 486 .
[18] Pazzani, Michael, et al. CTSHIV: A Knowledge-Based System For the Management of HIV-Infected Patients, University of California (Irvine),(1999), http://ieeexplore.ieee.org/iel3/5136/13957/00645169.pdf?tp=&arnumber=645169&isnumber=13957 , Accessed 20 January 2009.
[19] Giarratano, J.C., Riley, G.D. “Introduction To Expert Systems”. Expert Systems 4th Edition. Thomson Course Technology, (2005):(19,29,32,34).
[20] Limesurvey, a survey application, http://www.limesurvey.org/ Retrieved 17 January 2009