SENRA Academic Publishers, British Columbia
Vol. 8, No. 3, pp. 3043-3050, October 2014
Online ISSN: 1920-3853; Print ISSN: 1715-9997

Canadian Journal of pure&applied sciences

ONTOLOGY DRIVEN ANALYSIS AND PREDICTION OF PATIENT RISK IN DIABETES

Sherimon PC[1],*Vinu PV[1], Reshmy Krishnan[2] and Youssef Saad[3]
[1]Department of Computer Science, M. S University, India
[2]Department of Computing, Muscat College, Oman
[3] Faculty of Computer Studies, Arab Open University, Oman

ABSTRACT

In any health care system, the patient medical history is crucial to help the doctors for further patient diagnosis. History of the patient is collected mainly through face-to-face interaction when the patient visits the hospital. If the medical staff is not well experienced, it results in failure of collecting the patient history. So in most of the cases, effective patient risk analysis cannot be done. This paper proposes an ontology based system to collect the patient history and to assess the patient risk factorsdue to smoking history, alcohol history, erectile dysfunction history, and cardiovascular history. According to the patient history, a total score is calculated for each of the above factors. According to the score, the ontology performs the risk assessment on a patient profile and predicts the potential risks and complications of the patient.Ontology is among the most powerful tools to encode medical knowledge semantically.

Keywords: Patient profile, risk assessment, ontology, clinical guidelines.

INTRODUCTION

The conventional health care information collection systems address the issues of interoperability, flexibility and scalability. Patient's information is an important component in any health care system. At the time of emergency, the previous medical history of the patient is crucial, so that appropriate attention and treatment can be given immediately on any place at any time (Sanjay and Anand, 2010). The main commitment for any health care system is to improve the quality and privacy of patient's information. Currently, patient information is collected mainly through static questionnaires, patient information collection systems (hard coded systems) or through face-to-face interaction when patient visits the hospital. Normally, when a patient visits a hospital, a nurse will first diagnose the patient and will record the preliminary observations such as readings of blood pressure, height, weight, body and temperature (Ahmadian *et al*., 2010). Then doctors carry out physical examination and gather further information about the patient. Here, in most of the cases, the patient medical history will be incomplete. There are many reasons for this incompleteness such as lack of experience of the medical staff in collecting relevant patient details, patient does not want to disclose certain sensitive and personal health problems, patient is not able to communicate many medical details related to family (family history) because of shortage of time, etc. Also in some cases, doctors may not be able to ask the exact questions to each and every patient. In effect, effective risk analysis cannot be done in most of the cases.

*Corresponding author email: vinusherimon@yahoo.com

In many health centers, the information is also obtained through a computerized database system, supporting the doctors. In this case, every patient is required to answer the same set of questions. Also in such systems, it is difficult to change the order of questions or to add extra questions without major programming work. So in order to avoid these shortcomings, we propose an ontology based system to collect the history of patients and to analyze and predict the patient risk factors. The proposed system generates the patient profile, apply clinical guidelines on the profile, examine the patient risk associated with different factors and predict the risk. Instead of a normal database system, a knowledge based approach is used. The intelligence of the system depends on its knowledge base and reasoning algorithm. Ontology based systems are easy to update without any additional cost or work. Ontology languages allow users to write explicit, formal conceptualizations of domain models (Vinu *et al*., 2014). They help to explicitly define the existing information about the domain and formally encode that information (Ashburner *et al*., 2000). W3C has developed a language, called OWL that can be used to describe the ontology (Huiqun *et al*., 2012). OWL is based on description logics (Sherimon *et al*., 2013a). OWL represents ontology by building hierarchies of classes that describe the concepts in a domain and the properties which relate these classes to each other (Kawazoe and Ohe, 2008). In OWL, data are represented as individuals of OWL classes and are manipulated by the inference mechanism realized by Semantic Web Rule Language (SWRL) that provides an accessible method for expression of domain information as a rule set of an antecedent-consequent pair (Kawazoe and Ohe, 2008).

MATERIALS AND METHODS

Currently available patient information gathering systems use static questionnaires. Sometimes, some of the questions in the questionnaire are irrelevant to a majority of the patients. Most of the systems are hard coded, so it is not easy to alter a question or to introduce new questions. Our approach is to use an ontology based questionnaire to gather the medical history of patients and to analyse the patient risk associated with many factors such as smoking, cardiac history, alcohol history, etc.

Ontology is a formal specification of the concepts within a domain and their interrelationships (Subhashini and Akilandeswar, 2011). It is a methodology which describes the domain knowledge structure in the area of specialty, which promotes its various kinds of data processing intended to provide systematic, semantic links among groups of related concepts (Mojgan *et al*., 2009). They are well known for many years in the Artificial Intelligence and Knowledge representation communities (Rajendra, 2009). Ontologies are used to represent knowledge. Domain knowledge is contained in the form of concepts, individuals belonging to these concepts and relationships between the concepts and, between concepts and individuals (Sherimon *et al*., 2013b). It was initially proposed to model declarative knowledge for knowledge-based systems (Haya, 2011). It is an abstract model which represents a common and shared understanding of a domain (Sherimon *et al*., 2013b). Gruber (1993) proposed the most popular definition of ontology which is defined as "…a formal, explicit specification of a shared conceptualization." The W3C has developed a language, called OWL that can be used to describe the ontology (Huiqun *et al*., 2012). It is built on W3C standards XML, RDF/RDFS and extends these languages with richer modelling primitives (Vinu *et al*., 2014).

Methodology

An ontology driven approach is utilized here to model the medical history of the patient and to assess the patient risk. Ontologies are the backbone of Semantic Web and they include the descriptions of classes, properties and their instances (Vinu *et al*., 2012). Ontology is among the most powerful tools to encode medical knowledge semantically. The ontology's structure facilitates the organization, retrieval, and analysis of the encoded knowledge, including database design and merging of databases (Andreas *et al*., 2009). Ontology Reasoners are used to checking the consistency of the ontology and to automatically compute the ontology class hierarchy (Sherimon *et al*., 2013a). The reasoner will find out any hidden relationship in the ontology (Sherimon*et al*., 2013c). The use of ontologies is well suited for applications in medicine. When certain relations are asserted in the ontology, an ontology reasoner can infer more relations, which is not explicitly asserted in the ontology (Sherimon *et al*., 2011). They are renowned for their flexible architectures, easy to share and reuse knowledge modelling structures and inexpensive maintenance operations (Kamran *et al*., 2012). So ontology based adaptive system are used to generate the profile of the patient.

Initially, a questionnaire was designed with the help of medical experts. The questions related to patient's family history, diabetic history, smoking history, etc. are included as ontological classes. Accordingly sub-classes, data properties and object properties were also defined. The user interfaces are created in Java. Jena API is used to read the questions from the ontology. The user input is analysed and the semantic profile is generated. Later the clinical guidelines are applied to the semantic profile and patient risk is analysed.

Implementation

Ontology Development

Questionnaire ontology is created in Protégé. All the main concepts are created as parent classes in Questionnaire ontology. They are *Question_Bank*, *Question*, *Answered Questionnaire*, *Patient*, etc. *Question-Bank* is an abstract class. It consists of different subclasses according to the category. These sub-classes contain instances that correspond to that particular category. For example, the instances of *Diatecian-QuestionBank* class are questions coming under nutritional history. The *Question* class consists of subclasses to hold each type of question. For example, instances of *MultipleChoiceQuestion* class are multiple choice type questions. The object property *contains* is used to associate the class *Question_Bank* with *Question* class. Data properties *questionno* and *question-part* are used to keep the question number and the question text. The object properties *hasChoice*, and *subQuestion* keeps a set of choices and one or more sub-questions, if any. The class *AnsweredQuestionnaire* is created to store the instances of patient records. The instances of the class *Patient* are used to store the risk factor score of the patient. Figure 1 represents the class hierarchy of Questionnaire ontology.

Creation of Semantic Profile

Java application developed using Jena APIs provides different interfaces for patient, nurse, doctor, lab technician and dietician. It reads the questions from the ontology and displays it to the user. Patient is required to answer a set of questions on diabetic history, family history, smoking history, alcohol history, physical-activity history, etc. Nurse will enter some vital information about the patient such as body temperature, blood pressure, height, etc. The dietician enters information about the nutritional history of the patient. The lab technician enters the results of different lab tests conducted and finally the doctor examines the patient and enters his/her observations. For every new patient, the system stores the values entered by the users. Jena aims to provide a consistent programming interface for ontology application development, independent of which ontology language we are using in programs. Figure 2 shows the interface provided for the patients to enter his/her personal information.

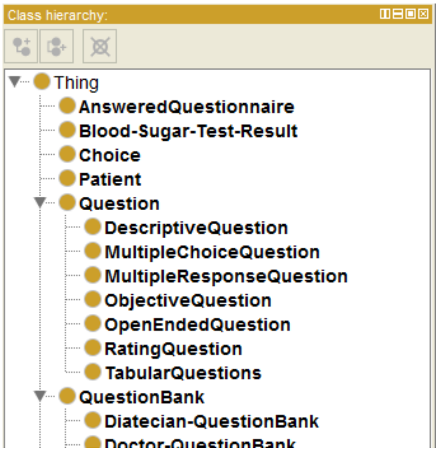

Fig. 1. Ontological Classes.

When the users complete filling the questionnaire, a new instance of *AnsweredQuestionnaire* class is automatically generated with patient ID as the instance name. For each instance, the values entered by the users are asserted into different data properties. For example, *questionnaire Answer-Cigarette-Count-Per-Day* is a data property to represent the number of cigarettes smoked by a patient per day. The domain of this class is *Answered Questionnaire* and the range is {10 or less, 11-20, 21-30 , 31 or more}. If a patient with ID 100 has chosen that he smokes 10 or fewer cigarettes a day, then the value of the property *questionnaireAnswer-Cigarette-Count-Per-Day* is “10 or less” for that particular instance.

Patient Risk Assessment

By performing the risk assessment on semantic profile, the potential risks and complications of a patient are predicted. The patient risk is assessed for five factors, according to smoking history, alcohol history, erectile dysfunction history, and cardiovascular history. It is done by checking the input provided by the users to different questions related to the above factors. A score is associated with each answer choice. So according to the user input, a total score is calculated for each category.

The clinical guidelines are hard coded in Java and the values generated are written back to the ontology. Figure 3 shows a section of the clinical guidelines implemented in Java. The risk/ risk level as per the score for each category is available in the clinical guidelines. Accordingly, the score, risk/ risk level, and suitable treatment are suggested by the system.

Here we are describing the estimation of cardiac score and the risk according to Framingham Heart Study, a risk assessment tool to predict a person’s chance of having a heart attack in the next 10 years. According to Framingham heart study, the cardiovascular risk is calculated based on factors such as Gender, Age, Total Cholesterol, HDL, Systolic Blood Pressure, Smoking habit, and Hypertension. Table 1 represents the cardiovascular risk points associated with each factor in the case of Women.

According to the user input to the questionnaire, the total cardiovascular risk points are calculated and then the cardiovascular risk is calculated according table 2.

Table 1. CardioVascular Risk Points – Women.

Points	Age, years	HDL	Total Cholesterol	SBP (Not Treated)	SBP (Treated)	Smoker	Diabetic
−3				<120			
−2		>1.56					
−1		1.30–1.55			<120		
0	30–34	1.17–1.29	<4.14	120–129		No	No
1		0.9–1.16	4.14–5.15	130–139			
2	35–39	<0.9		140–149	120–129		
3			5.16–6.19		130–139	Yes	
4	40–44		6.20–7.24	150–159			Yes
5	45–49		>7.25	160+	140–149		
6					150–159		
7	50–54				160+		
8	55–59						
9	60–64						
10	65–69						
11	70–74						
12	75+						

Table 2. CardioVascular Risk – Women.

Points	Risk, %
≤−2	<1
−1	1.0
0	1.2
1	1.5
2	1.7
3	2.0
4	2.4
5	2.8
6	3.3
7	3.9
8	4.5
9	5.3
10	6.3
11	7.3
12	8.6
13	10.0
14	11.7
15	13.7
16	15.9
17	18.5
18	21.5
19	24.8
20	28.5
21+	>30

According to the risk calculated, the treatment is suggested as per the clinical guidelines. Figure 4 represents the part of the guidelines related to cardiovascular risk. The medicine *Aspirin* is recommended by considering the risk calculated in the above step and the risk factors such as age, smoking, hypertension, etc.

RESULTS AND DISCUSSION

Patient Medical History

The class *AnsweredQuestionnaire* is created to store the instances of patient records. It consists of all patient records stored in the form of ontology. Each patient is considered as an individual (instance) of this class and is automatically generated from Questionnaire Ontology using Property Insertion (OWL 2) and Individual Insertion (OWL 1 & 2) mechanisms. The *questionnaire Answer-Typical-Result* is the object property used to link the individuals of the classes *AnsweredQuestionnaire* and *Blood-Sugar-Test-Result*. *Blood-Sugar-Test-Result* class creates instances when a patient enters his/her blood results. When a patient enters his medical history, an instance of *AnsweredQuestionnaire* is asserted in the ontology. The name of the instance will be the patient Id. The values entered by the patient are stored in the concerned sub-data properties of the main data property *questionnaireDataTypeAnswers*.

The system instantiates the Questionnaire ontology and stores the corresponding answers in it. The System processes this information and automatically generates a Patient Ontology instance in the server. Patient Medical Profile is an OWL file which encapsulates patient details as entered by the patient, nurse and other users in a Web/Mobile Application. This file is initially generated as soon as the patient enters and submits the questionnaire through the application. Later, the same is updated by the Clinical officials (e.g. nurse). Figure 5 shows the semantic profile generated in OWL.

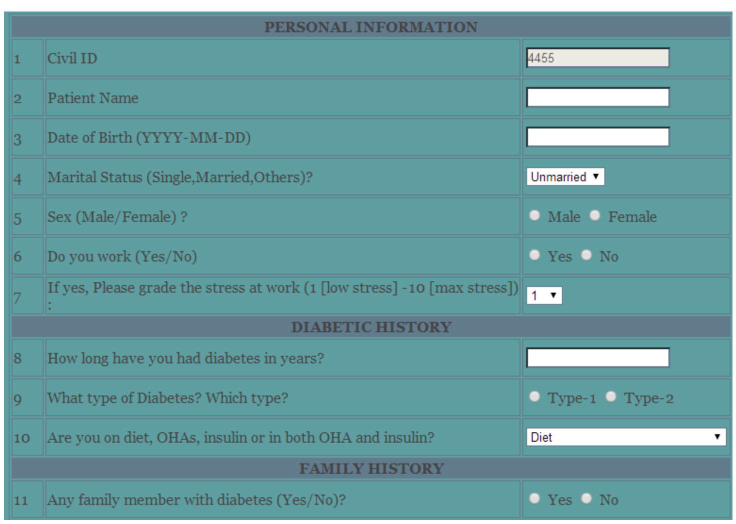

Fig. 2. Patient Interface.

```
if(uneasyGiveupTime.equalsIgnoreCase("First in the morning"))total_smoking_score++;
if(smokeSickTime.equalsIgnoreCase("Yes"))total_smoking_score++;
ontology.addDataPropertyValue("p_"+civil_id,"smoking-Score",OWL2Datatype.XSD_NON_NEGATIVE_INTEGER,
if(total_smoking_score>=0&&total_smoking_score<=3){
    ontology.addDataPropertyValue("p_"+civil_id,"smoking-Risk-Level",OWL2Datatype.XSD_STRING,"Low
    ontology.addObjectPropertyValue("p_"+civil_id, "smoking-Treatment", "Score_Level-0-3");
}
if(total_smoking_score>=4&&total_smoking_score<=6){
    ontology.addDataPropertyValue("p_"+civil_id,"smoking-Risk-Level",OWL2Datatype.XSD_STRING,"Medi
    ontology.addObjectPropertyValue("p_"+civil_id, "smoking-Treatment", "Score_Level-4-6");
}
if(total_smoking_score>=7&&total_smoking_score<=10){
    ontology.addDataPropertyValue("p_"+civil_id,"smoking-Risk-Level",OWL2Datatype.XSD_STRING,"High
```

Fig. 3. Clinical Guidelines in Java.

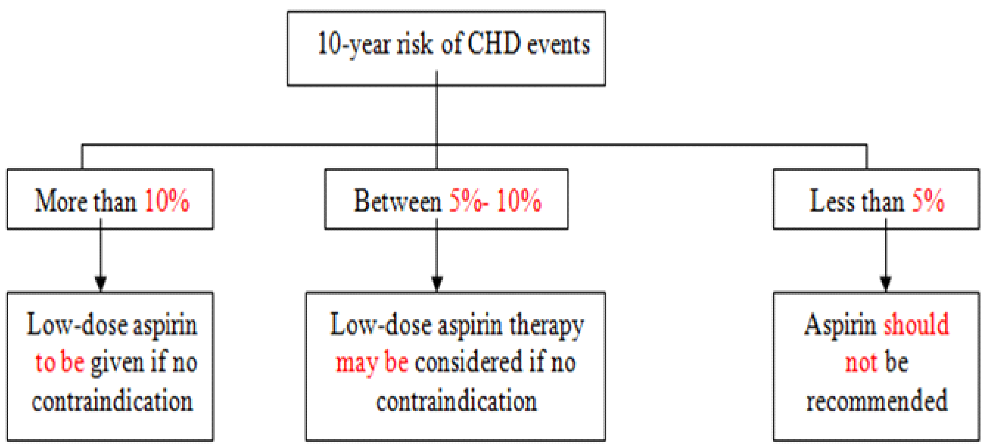

Fig. 4. Suggested Treatment as per the Risk %.

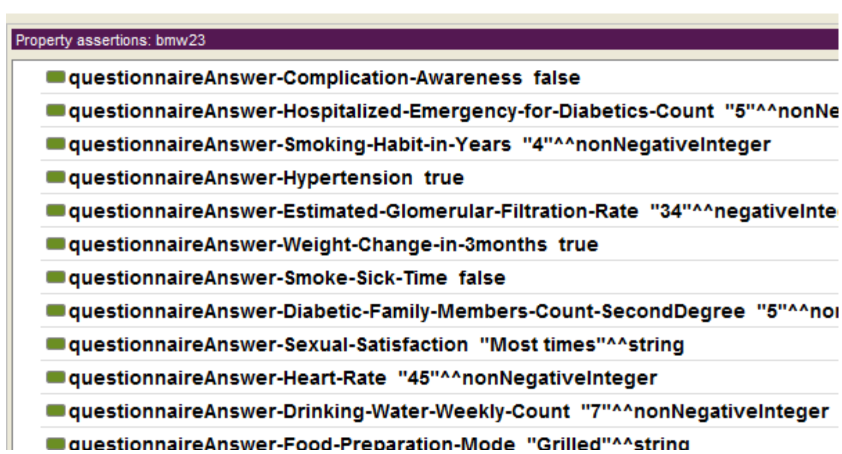

Fig. 5. Patient Profile: OWL View.

Doctor needs to view the history of the patient before suggesting treatment. The profile format created in OWL cannot be understood by a Doctor. So after the data is entered by patient, nurse, lab technician and dietician, the doctor can view the complete history of the patient in a tabular format. Figure 6 represents the view of Doctor.

Patient Risk Analysis

Apart from viewing the history of the patient, doctor can also view the risk of the patient associated with several factors. Every answer to a question carries a particular point. So according to the user input, the points will be allocated and total score is calculated for each and every

PATIENT PHYSICAL ACTIVITY HISTORY		
1	Type of Work & Related Physical Activity	My work involves definite physical effort including handling of heavy objects and use of tools
2	Physical Exercise Duration	Some but less than one hour
3	Cycling Time	Three hours or more
4	Walking Time	One hour but less than 3 hours
5	Child Care Time	Some but less than one hour
6	Gardening Time	One hour but less than 3 hours
7	Walking Pace	Slow Pace
PATIENT MEDICAL HISTORY		
1	Hospitalized for Diabetics?	Yes
2	No. of times hospitalized	5
3	Hospitalized in Emergency for Diabetics?	Yes

Fig. 6. Patient Profile: Doctor View.

cardio-Score "0"^^nonNegativeInteger
cardio-Treatment "Aspirin should not be recommended."^^string
smoking-Score "4"^^nonNegativeInteger
smoking-Risk-Level "Medium nicotine dependence"^^string
physical-Treatment "Counsel about the importance of being active, in follow up"^^string
sexual-Score "18"^^nonNegativeInteger
sexual-Risk-Level "Mild erectile dysfunction"^^string
cardio-Risk 1.5f
physical-Score "8"^^nonNegativeInteger
alcohol-Risk-Level "ZONE III"^^string
physical-Risk-Level "Moderately Inactive"^^string
alcohol-Score "16"^^nonNegativeInteger

Fig. 7. Patient Score, Risk and Treatment asserted as Data Property values.

category. For example, according to the user input to all the questions included in the Smoking History Category, a score is calculated and it provides the risk of the patient due to smoking. The system will analyze the risk of the patient according to the history in five important aspects, Smoking, Alcohol, Sexual, Cardio Vascular and Physical and displays it.

Data properties are used to store the scores, risk level and the treatment of each of the above factors. For example, *cardio-score*, *cardio-risk* and *cardio-Treatment* are the data properties related to cardiac history. As per the clinical guidelines, the score, risk and treatment are calculated, an instance of Patient class is asserted in the ontology and the above calculated values are stored as data property values of the instance. Figure 7 shows the score, risk and treatment of a patient with ID P_789, for each of the parameters discussed earlier.

PATIENT RISK ANALYSIS

Parameter	Score	Risk	Treatment
Smoking	4	Medium nicotine dependence	1. Require professional counselling. 2. May recommend pharmacotherapy if patient is assessed to be suitable. Pharmacist and/or doctor to provide more advice on pharmacotherapy. 3. Provide willpower and support from family and friends
Alcohol	16	ZONE III	Simple Advice plus Brief Counseling and Continued Monitoring
Sexual	18	Mild erectile dysfunction	Discuss with the patients the option of starting him on sildenafil
Cardio-Vascular	0	1.5	Aspirin should not be recommended.
Physical	8	Moderately Inactive	Counsel about the importance of being active, increase activity level and follow up

Fig. 8. Patient Risk Analysis.

The patient risk analysis is also displayed in the form of a table for doctors to decide about further specific treatment to the patients. The system suggests treatment as per the risk associated with each factor. It is represented in figure 8.

CONCLUSION

Knowing Patient history is vital for a doctor to perform a proper patient risk assessment and to suggest appropriate treatment. We have used ontology based approach in generating the health history of patients. These systems gather better medical history than nurse/ dietician/ other hospital staff. The questionnaire is not static, it is adaptive in nature, so only relevant questions according to patient context will be asked. System reduces the number of questions, thus saving the time of patients. Since the questions are read from the ontology, if we want to add or update or delete any questions, we need to do it only in the questionnaire ontology. The information about each patient is generated as a separate OWL file. It can be viewed from Protégé in OWL format and also from the system in the tabular format. The patient risk associated with different factors is also generated. These risk values (score) will help the doctor to understand about the current situation of a patient. Ontology based reasoning makes a way to discover new knowledge, which can lead to new directions in research. The future scope of our research is to update the rules hard coded in Java to SWRL (Semantic Web Rule Language) to implement the Clinical Guidelines and to use the reasoning power of OWL to support the doctors in suggesting specific medicines and further detailed treatment.

REFERENCES

Ashburner, M., Ball, CA., Blake, JA., Botstein, D., Butler, H., Cherry, JM., Davis, AP., Dolinski, K., Dwight, SS., Eppig, JT., Harris, MA., Hill, DP., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, JC., Richardson, JE., Ringwald, M., Rubin, GM. and Sherlock, G. 2000. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 25(1):25-29.

Gruber, 1993. A translation approach to portable ontologies, Knowledge Acquisition. 5(2):199-220.

Haya, El-Ghalayini. 2011. E-Course Ontology for Developing E-Learning Courses. Developments in E-Systems Engineering (DeSE). 245-249.

Huiqun, Z., Shikhan, Z. and Junbao, Z. 2012. Research of using Protégé to Build Ontology. In: Proceeding(s) of the IEEE/ACIS 11th International Conference in Computer and Information Science. 697-700.

Kamran, FAH., Stephen, L., Chris, E., Calum, MR. and Warner, S. 2012. An Ontology Driven and Bayesian Network Based Cardiovascular Decision Support Framework, Advances in Brain Inspired Cognitive Systems. Lecture Notes in Computer Science. 7366:31-41.

Mojgan, H., Michihiko, K., Takako, T. and Hiroshi, T. 2009. Development of clinical ontology for mood disorder with combination of psychomedical information. Journal of Medical and Dental Sciences. 56(1):1-15.

Rajendra, A. 2009. Foundations of the Semantic Web. Narosha Publishing House.

Subhashini, R. and Akilandeswar, J. 2011. A Survey on Ontology Construction Methodologies, International Journal of Enterprise Computing and Business Systems. 1(1):60-72.

Sanjay, A. and Akshat, V. 2010. Development of Ontology for Smart Hospital and Implementation using UML and RDF. IJCSI International Journal of Computer Science. 7(5):206-212.

Sherimon, PC.,Vinu, PV., Reshmy, K. and Youssef Takroni. 2013[a]. Developing Survey Questionnaire Ontology for the Decision Support System in the Domain of Hypertension. IEEE South East Conference, Florida, USA.

Sherimon, PC., Reshmy, K., Vinu, PV. and Youssef Takroni. 2013[b]. Ontology Based System Architecture to Predict the Risk of Hypertension in Related Diseases. International Journal of Information Processing and Management. 4(4)doi:10.4156/ijipm.

Sherimon, PC., Reshmy, K., Vinu, PV. and Youssef, T. 2013[c]. Exhibiting Context Sensitive Behavior in Gathering Patient Medical History in Diabetes Domain using Ontology. International Journal of Advancements in Computing Technology. IJACT. 5(13):41-47.

Vinu, PV., Sherimon, PC. and Reshmy, K. 2014. Development of Ontology for Sea food Quality Assurance System. Journal of Convergence Information Technology. 9(1):25-32.

Vinu, PV., Sherimon, PC. and Reshmy, K. 2012. Knowledge - Base Driven Framework for Assuring the Quality of Marine Seafood Export. International Journal of Artificial Intelligence and Knowledge Discovery. 2(3): 6-10.

Kawazoe, Y. and Ohe, K. 2008. An Ontology-based Mediator of Clinical Information for Decision Support Systems: A Prototype of a ClinicalAlert System for Prescriptions. Methods Inf Med. 47:549-559 doi:10.3414/ME9126.

Received: April 19, 2014; Accepted: June 26, 2014