International Journal of Computer Science and Business Informatics

IJCSBI.ORG

An Ontology Based System for Predicting Disease using SWRL Rules

Mythili Thirugnanam
Assistant Professor (Senior), School of Computing Science and Engineering
VIT University, Vellore, Tamilnadu, India

Tamizharasi Thirugnanam
Assistant Professor , School of Computing Science and Engineering
VIT University, Vellore, Tamilnadu, India

R. Mangayarkarasi
Assistant Professor (Senior), School of Information Technology and Engineering
VIT University, Vellore, Tamilnadu, India

ABSTRACT

This paper basically provides information about the various diseases to the user with the help of ontology. The developed ontology consists of disease and its relationship with symptoms and SWRL rules (Semantic Web Rule Language) are designed for predicting diseases. The developed ontology system has contains two stages. The first stage is defining the class hierarchy and defining the object and data properties. The second stage is executing rules which extract the disease details with symptoms based on the rule specified. Finally the inferred axioms reflected in the ontology. The various testing shows the successful execution of ontology. The analysis of the results obtained followed by their discussion gives the final risk value to the user of the system.

Keywords

Ontology, Disease, Semantic Web Rule Language (SWRL)

1. INTRODUCTION

In today's world everyday new diseases are identified and diagnosed. Everyday normal people become patients and get diagnosed. Generally diseases are identified by the symptoms that our body creates. A cause for a disease may differ from one another. But a common cause for most of the diseases is ignorance, which itself is caused by the lack of knowledge about the symptoms indicated by human body. To address this problem we need a common knowledge sharing tool that can help us to share knowledge about diseases through the symptoms of them. Ontology is a technology created for sharing the knowledge in machine understandable format. With the help of ontology an information system about diseases and symptoms is created, some relationships between diseases and symptoms are identified and rules are written using those relationships in order to diagnose a particular disease out of given symptoms. Those rules are written in SWRL (Semantic Web Rule Language) format. The choice of information system over a database is to provide precise and relevant information for user query. Presently similar disease information systems using ontology is available but they do not deal with human diseases [4].The motivation behind the development of this system is to spread knowledge about the diseases and their symptoms to the various kind of people. Lack of knowledge about the diseases is also a cause for diseases. This emphasizes the essential of developing a disease information system. Another important thing is that this ontology can be reused to create similar system and thus it is very efficient.

2. BACKGROUND STUDY

MajaHazic et al., [9] had proposed ontology based information system for bio medical field. Ontology based grid middleware was developed for human disease research, to resolve the medical issues and disease factors. The proposed system fails to concentrate for improving the usability of the system. Maja Hadzic et al., [10] introduced Ontology based multi-agent system to support human disease study and control. A new ontology called generic human disease ontology (GHDO) was designed to represent knowledge about the human disease. Knowledge representation is included the four attributes such as disease types, symptoms, causes and treatments. The security and interaction of the developed system is not appreciable. Maja Hadzic at el., [11] had enhanced their study on implementing Ontology based Disease System. They developed the GHDO (Generic Human Disease Ontology) to represent the disease knowledge and the information of the diseases is organized into four dimensions such as Disease types, Symptoms, Causes and treatment. In addition to this, the system was aimed to support the study of complex disorders caused by many different factors simultaneously. Illhoi Yoo et al., [7] studied various document clustering approaches for MeSH ontology to improve the clustering quality. The results obtained from the developed biomedical ontology MeSH, enhanced the clustering quality on biomedical documents and decent document clustering approaches performance was improved. Hongyi Zhang et al., [6] introduced a method for obtaining biological functions of gene by using GENE ontology method. They were able to differentiate the relation between parent’s gene and their children’s genes of all types of genes. From the results it was possible to construct probabilistic gene regulatory networks with the method of coefficient of determination (CoD). Akifumi Tokosumi et al., [1] evaluated existing medical ontologies and proposed future directions of medical knowledge repository system with three knowledge repository such as localized nature of knowledge, collective acquisition of knowledge and usable knowledge repository. The locality of knowledge was suggested to be used in medical ontologies. Tran Quoc Dung et al., [15] developed Ontology based health care information extraction system – VnHIES. Elements extracting algorithm and new semantic elements were used to extract the health care semantic word.

Document weighting algorithm was applied to get summary of health information. The result of the proposed system is more accurate. Tharam S. Dillon et al., [14] had carried out work on ontology in bio medical information storage and processing. The importance of ontologies in representing bio medical information storage and processing knowledge models are discussed and also explained the uses of ontologies in semi-automatic and automatic tasks. Dahua Xu et al., [5] implemented pest and disease information system based on WEBGIS named as Diseases and Pest Information system (DIPS). A DIPS was a pest and disease control system and warning system designed for crop pests. The system architecture consists of different components and sub-systems such as WEBGIS component – the interface component, DIPS component which was used for data access, and Component model management sub-system. Component model management Sub-system manages the centralization between DIPS and DIPS and manages its components like mathematical models, pest components and interface component. The COM + component technology was used in order to incorporate object oriented technology, thus improved the scalability, reusability of the proposed system and enabled distributed computing. Shastri L. Nimmagadda et al., [13] proposed ontology based data warehouse modeling and managing ecology of human body for disease and drug prescription management. The system was focused on introducing the concept of ontology based warehouse modeling and representing human body system in ontological representation. The proposed system was yet to be put into practice. Antonio J. Jara et al., [3] implemented ontology and rule based intelligent information system to detect and predict myocardial diseases. The developed system was used in pre-hospital health emergencies, remote monitoring of patients with chronic conditions and medical collaboration through sharing health-related information resources. Rule based system was designed to predict the illness by applying chronobiology algorithm. Ontology trees were constructed to in order to provide knowledge base of diseases. To avoid the observation periods in the hospital the system was used to send the information about the detected symptom or disease. Though this system was useful the chronobiology algorithm was not based on the diagnosis and improvements in the artificial intelligence layer were required. Ali Adeli et al., [2] had developed a Fuzzy Expert System for Heart Disease Diagnosis based on the V.A. Medical Center, Long Beach and Cleveland Clinic Foundation database. It had 13 input fields and one output field. The input fields were attributes of heart disease such as chest pain, resting electrocardiography (ECG), etc. The output field was an integer value ranges from 0 to 4 to denote different levels of the heart disease. The system showed 94% accuracy in classifying the heart disease. Ersin Kaya et al., [5] had developed a Diagnostic Fuzzy Rule-Based System for Congenital Heart Disease. They retrieved medical dataset of patients from Pediatric Cardiology Department at Selcuk University, from years 2000 to 2003. They classified the medical dataset into 4 groups for fuzzy classifications and then the fuzzy rules were created based on various attributes in the data set. These attributes includes 8 Conditional attributes, 4 Decision attributes. After classifying fuzzy rules, they weighted it two different methods such as weighted vote method and single winner method and compared the results. They increased the accuracy of Classification of Congenital Heart Diseases. Yip Chi Kiong et al., [16] had developed Health Ontology System to store clinical databases into a shared cumulative ontology so that it can be intercepted by machines. Such system was built upon their previous work that is Ontology generator, Ontology Distiller and Ontology Accumulator. These are software tools used in the system generate ontology. Ontology generator generates ontology from a database. Ontology Distiller does the reverse process by storing ontology into a database. The Ontology accumulator does the integration of similar types of ontology. Integration of these tools helped to convert small databases to complex database tables. Lynn Marie Schriml et al., [8] created disease ontology that provides a backbone for disease semantic integration. The developed disease ontology contains a knowledge base of 8043 human diseases. It was designed as a web interface designed for high speed, efficiency and robustness through the use of graph database. It supported querying of disease name, synonym, and definition of diseases. This work was not extended to relations among symptoms, causes and diseases. Above literature clearly shows, there is no specific domain ontology is not yet developed for human disease. Hence, there is need for developing such a domain ontology which focuses only human diseases to help the people to get to know the information without any difficulty. Therefore, this work focused to develop ontology based system for predicting the human diseases.

3. PROCEDURES ADOPTED FOR DEVELOPED SYSTEM

Mythili et al., [12] proposed the ontology based disease information system. The developed system adopted the procedure from the proposed work .The developed system consists of three different phases such as Knowledge acquisition phase, Rule engine phase and Query processing phase. Creation of ontology is done in knowledge acquisition phase. Writing the SWRL rules are done in the rule engine phase. Query processing phase get the query from the user and responds to his or her query.

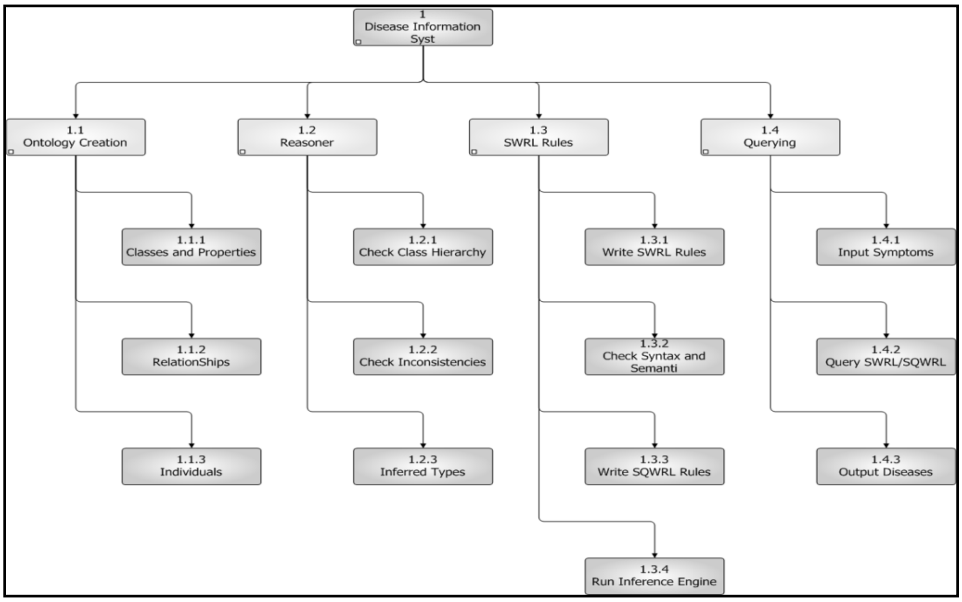

Figure 1. Components of the developed ontology based disease information system

3.1 Knowledge Acquisition Phase

Ontology of a domain is useful to share and reuse the explicit knowledge of that domain. Creating medical related ontology's concepts are very much in need. Creating ontology and identifying relationships in them can be done with the help of a tool called protégé. Building ontology consists of different modules such as creating ontology, creating classes and creating or properties, identifying relationships and adding records. Sample ontology class construction is shown in Table1 and protégé implementation is shown in the Figure2.

Table 1 Sample ontology classes and their properties.

	Class 1	**Class 2**	**Class 3**
Class name	Heart disease	Chest_Pain	Fever
Object property	has_symptom	is_a_symptom_of	is_a_symptom_of
Data property	has_name	has_id	Has_Temperature

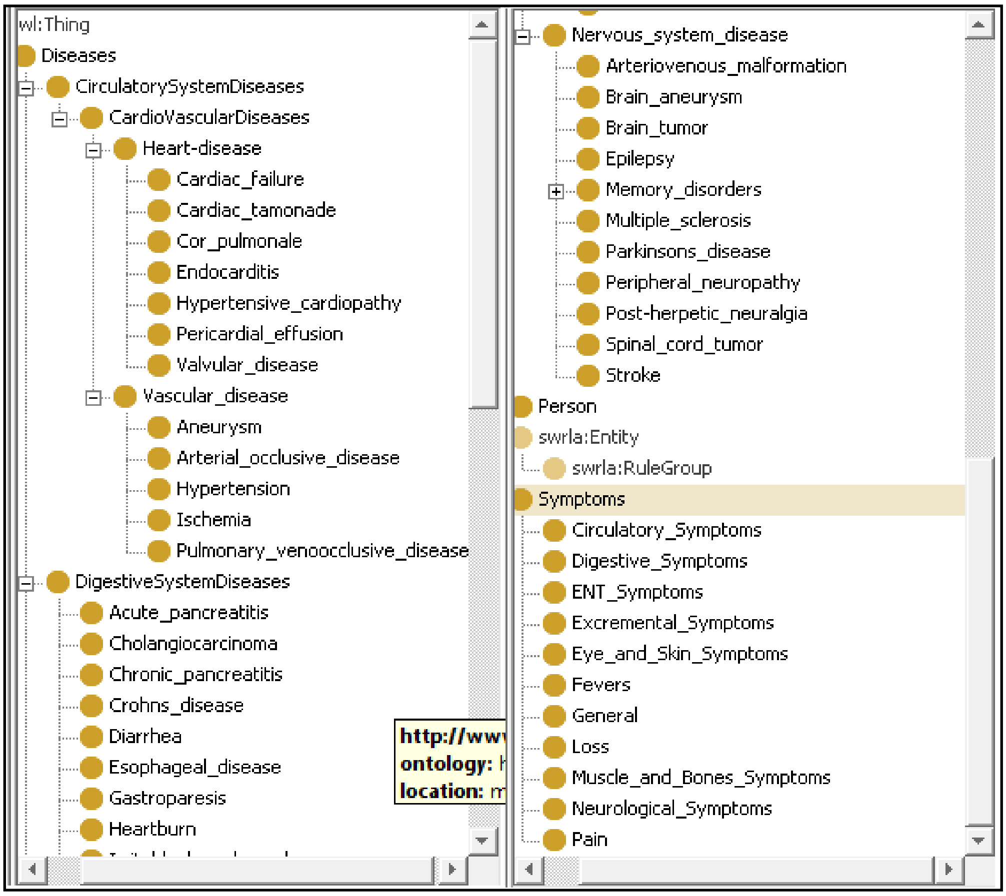

Figure 2. Class Hierarchy (using Protégé).

Here the required knowledge for the system is collected and organized in required format. The knowledge is collected from various resources such as World Wide Web, Physicians, etc. The Detailed description of diseases, symptoms are stored in a datasheet. As already mentioned, the medical sciences field is very vast. So incorporating every disease is not possible.

3.1.1 Creating Ontology

Creating ontology is done with the help of Protégé tool. The required ontology created and saved in local storage. The actual code for the ontology is created by protégé where the code is in XML format. The extension of created ontology is given as .owl where OWL stands for Ontology Web Language. With this file we can create desired classes and their properties. Creating ontology is simple step while using protégé. Protégé simply creates the required XML file and store it in OWL format so that the file can be accessed as ontology. The implementation of same is shown in the Figure 3.

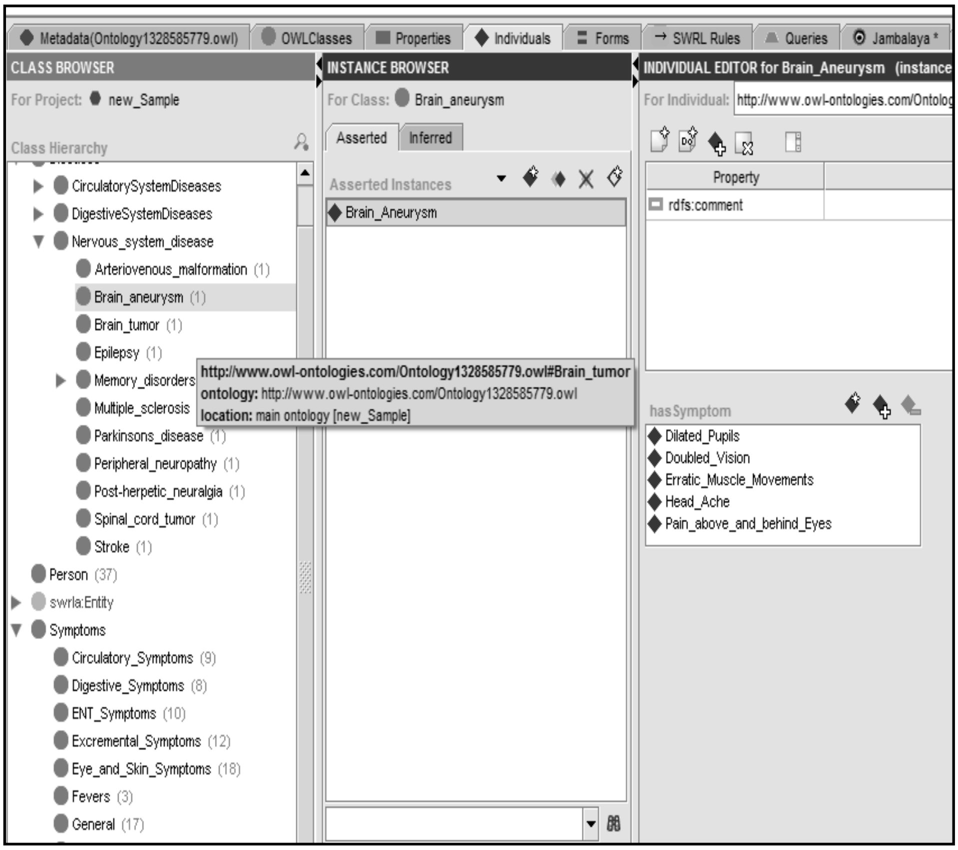

Figure 3. Creating ontology using protégé.

3.1.2 Creating Classes and Properties

The system classifies the diseases into human body system wise diseases. Hence each disease system is considered as a subclass of super class called diseases. So we create a subclass of symptoms under the super class. So that it serves as a repository of a symptoms and they related with any disease system, where disease systems classes and symptoms are siblings of disease. Finally a class called people is created in order to relate diseases and symptoms. This people class is a disjoint class from symptom. The implemented class architecture is shown in the Figure 3. As it is strongly recommended to identify and add the properties of a class or subclass in order to make the classes more understandable, properties are created. The main data properties of all classes are "*hasName*" and "*hasID*". The main object properties of diseases class, Symptoms class and person class are hasSymptom, isSymptomof and hasDisesae respectively. The creation of data and object properties using protégé is shown in the Figure 4.

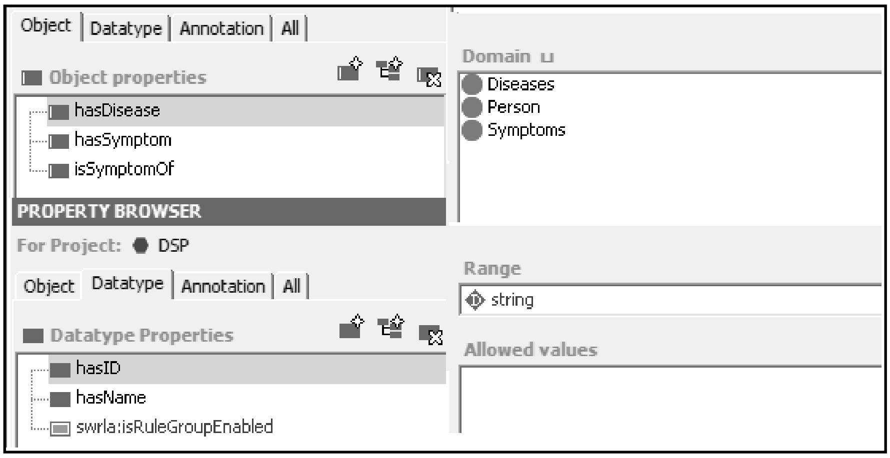

Figure 4. Data Properties and Object Properties

3.1.3 Identifying the Relationships between Classes

The relationship between a system disease and a symptom can be "hasSymptom" relationship. For example "stroke hasSymptom dizziness". We can create a vice versa relationship "isSymptomOf". For example: "dizziness is a symptom of stroke". Finally the person classes have two important relations with these two classes, such as it as relationship with diseases as "hasDisease" and with symptoms class as "hasSymptom". For example: "Person-1 hasSymptom Dizziness, Fainting, etc. Sample snapshot for identifying the relationship between classes are show in the Figure 5.

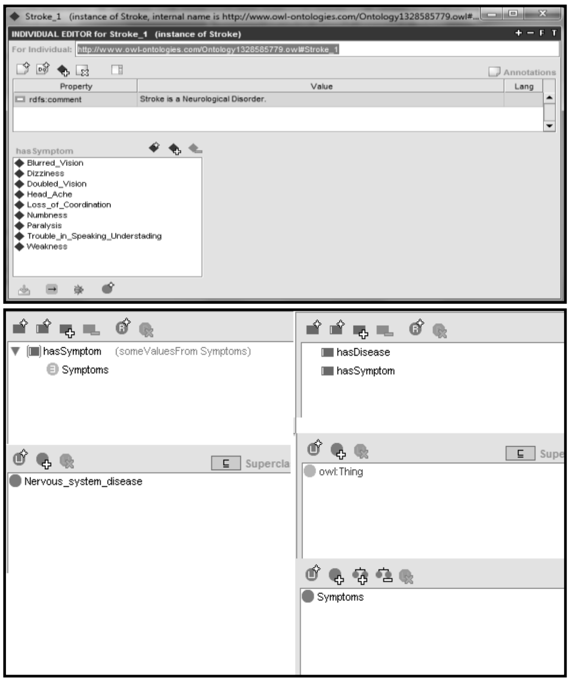

Figure 5. Relationship between classes.

3.2 Rule Engine Phase

The Rule engine phase of the system consists of a semantic reasoner and SWRL Rules. The semantic reasoner is used for checking the consistency of the relationship among the classes and their properties (Figure 6). This is done in order to validate the SWRL rules. SWRL rules are created from valid relationships to detect disease from the given symptoms. For example if a person has a symptom chest pain, shortness of breathing, pain in arms, dizziness, eye color red and fastest heartbeat we should say that the person has coronary heart disease. The following rule is a sample SWRL rule created using Protégé and SWRLTab. The following figures show some the rules in graphical format. There is another format to write these rules that is known as SQWRL which provides SQL like querying functionalities. Those rules also show in graphical format.

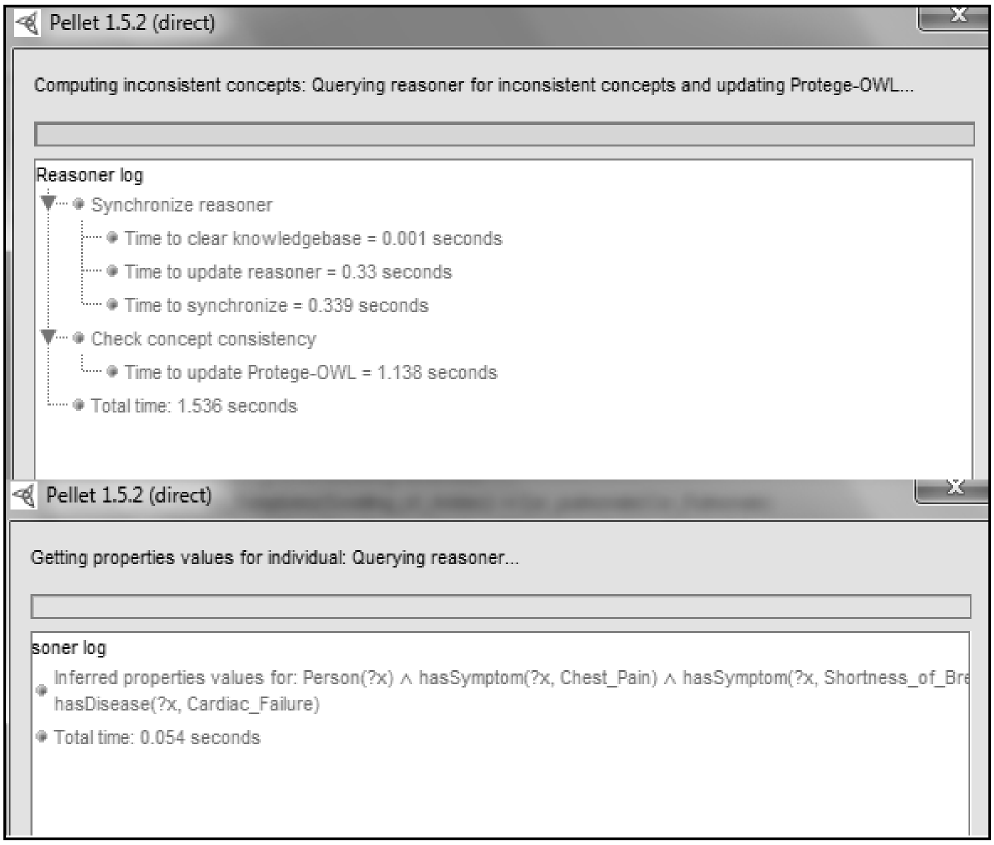

Figure 6. SWRL Reasoner Dialogue - checking consistency

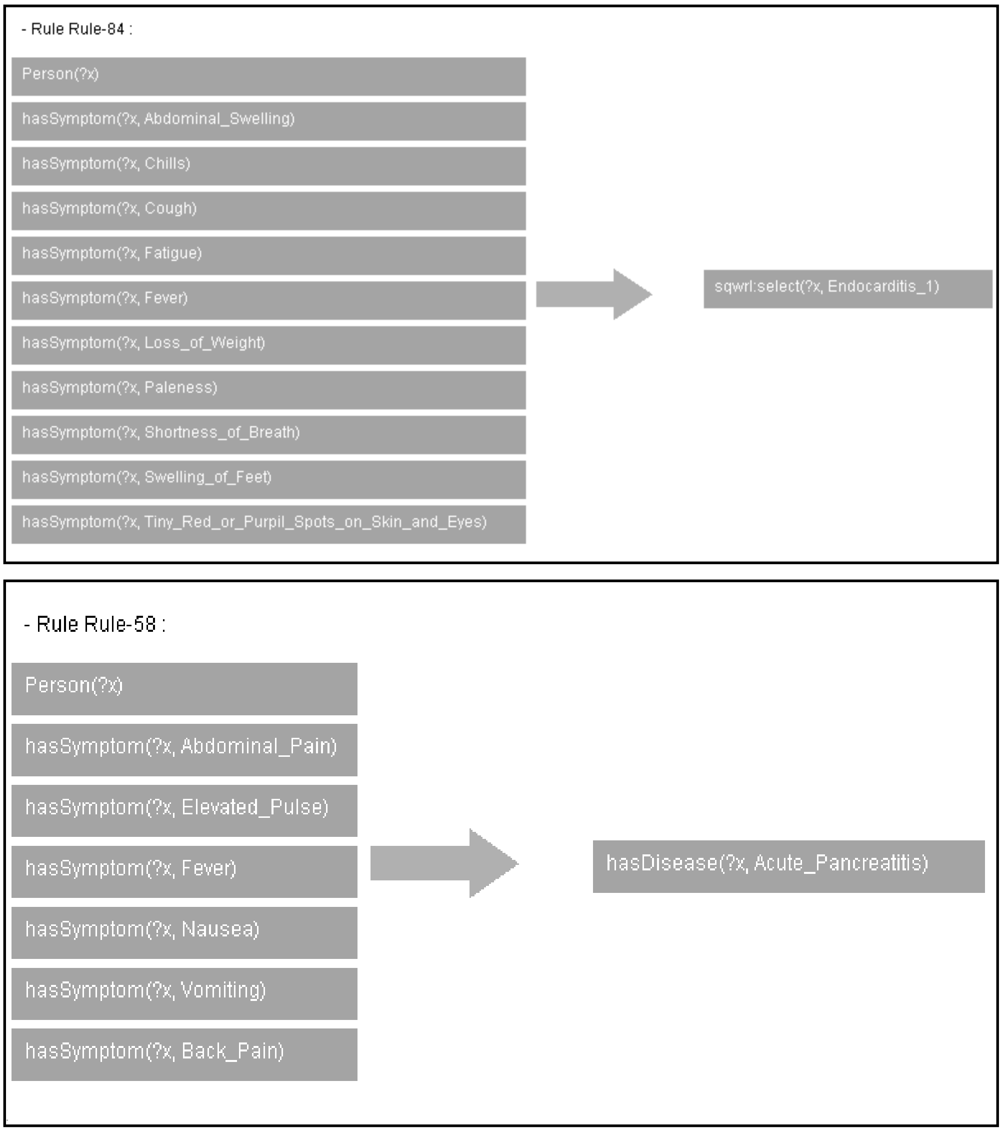

Figure 7. SQWRL and SWRL Rules in graphical format.

3.3 Query processing phase

In query processing phase user interaction is managed. If user enters some symptom he or she has, the system request the query processor. The query processor checks with SWRL rules for relations between the diseases. It returns the diseases associated with symptoms entered. The query processor then displays the output to the user. In protégé these querying is done with the help of JessTab.

4. RESULT & DISCUSSION

The output SWRL and SQWRL rules are the main form of output in order to retrieve information out of ontology. After checking the consistency of the class hierarchy the SWRL rules are executed in jess inference engine. The output of the jess inference tab is shown in the following Figure 8.

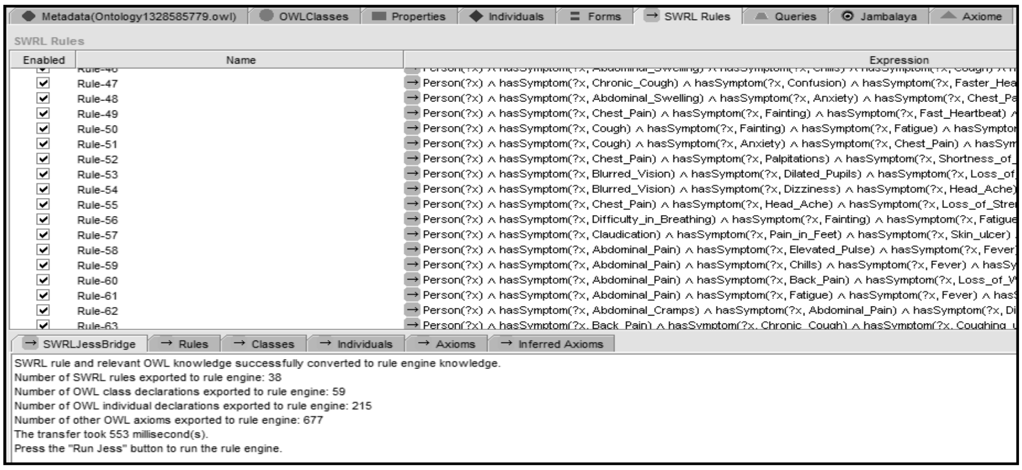

Figure 8. SWRL output (Jess Inference process)

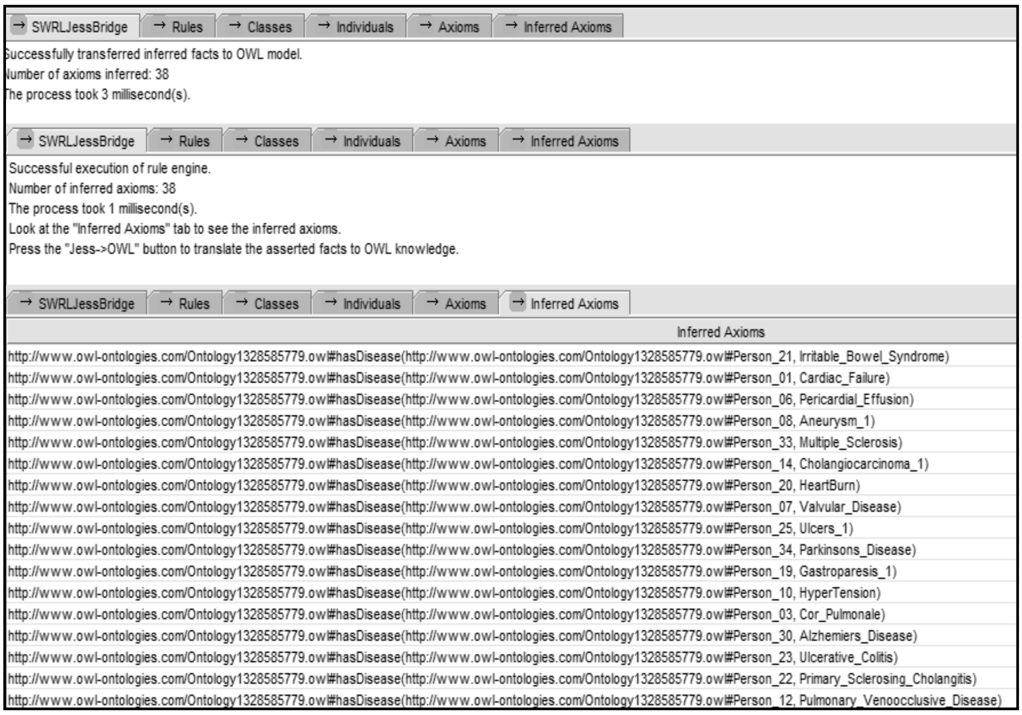

Figure 9. SWRL output (Inferred Axioms).

4.1 Result Analysis

The results are successfully obtained from the jess inference engine and SQWRL Tab. In figure 10.part of the inferred axioms from the ontology is shown. The jess result is in the form of "person-x has Disease Name"; The SQWRL output is in the form of "Person – Disease". Both results are shown in Figure 10 and Figure 11.

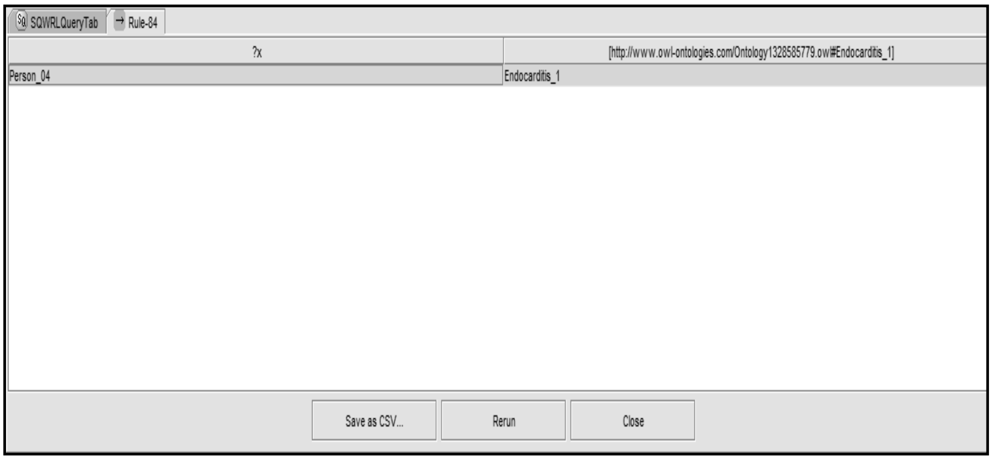

Figure 10. Output of SQWRL

These results are stored in the properties of each instance of class. An instance of class person who has symptoms for example palpitations and Shortness of breath may have the disease hypertensive cardiopathy. This is defined in the SWRL rule and the output is stored into the ontology.

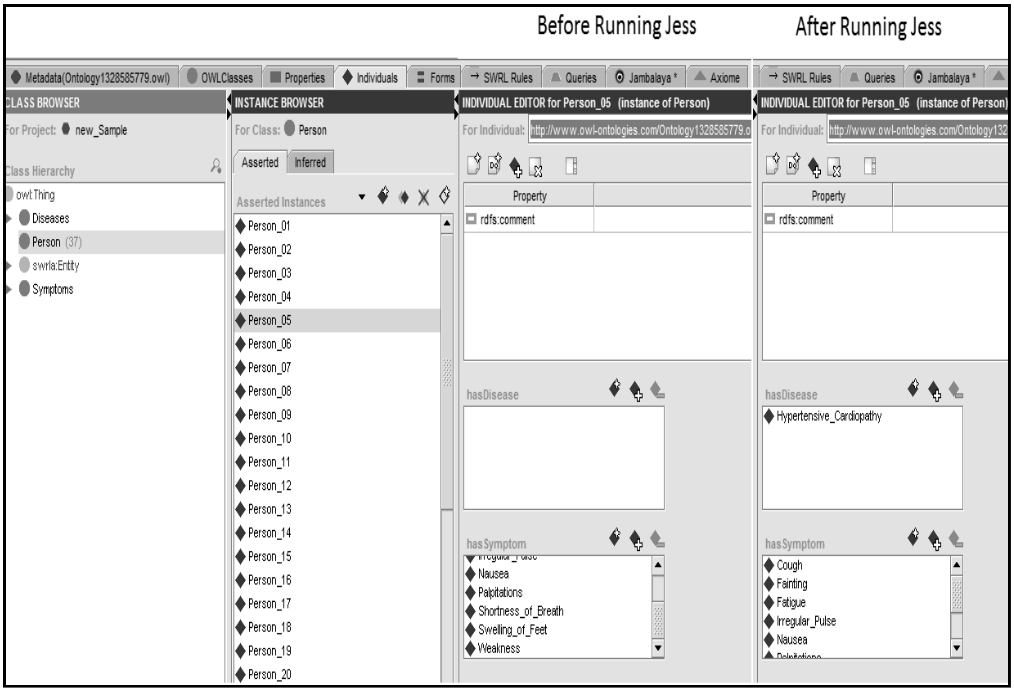

Figure 11. Replication of output in ontology after using Jess.

These results can be viewed using visualization tools. Jambalaya is a visualization tool that provides options to query the ontology. Hence we customized that tool so that it will work as per this project's requirements. Some of the visualizations are shown in Figure 12.

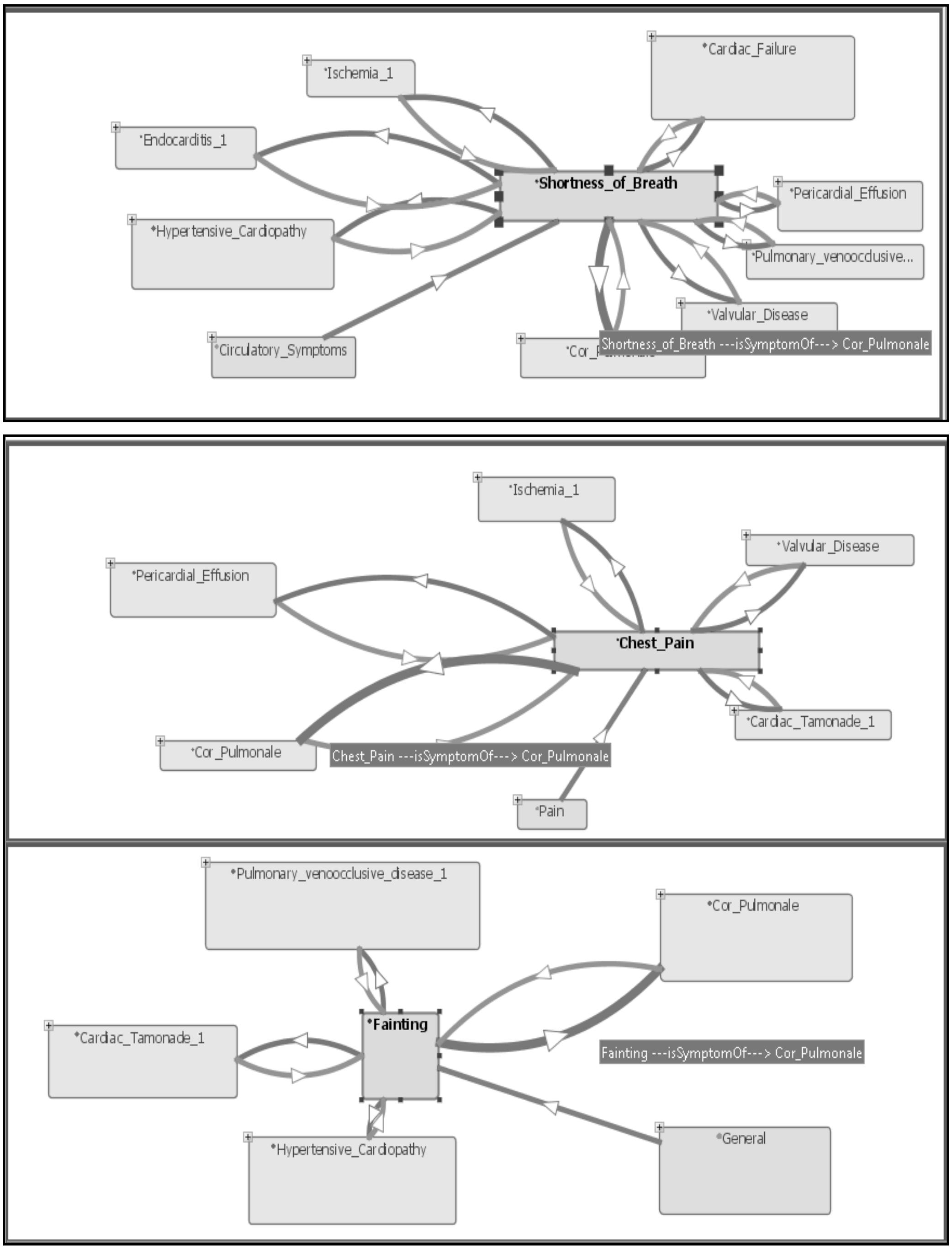

Figure 12. Visualization using Jambalaya.

4.2 Discussion

This system depicts the use of ontology as a knowledge acquisition method and successfully proved it in the disease – symptoms domain. This is just an initation or example of how powerful the semantic web technologies can be used in knowledge sharing and envisage the resuablity of ontologies.

5. CONCLUSION AND FUTURE WORK

This paper provides an approach to create disease information system about diseases and symptoms with the help of ontology. The developed disease information system will help the users to be aware of diseases and their symptoms and help them to take viable actions. Thus this work points out the importance of creating a disease information system and demonstrated a successful one. The future developments to system include inclusion of other diseases, inclusion of cause-disease relationship, inclusion of Treatment-disease and presenting the ontology in along with easily accessible user interface.

REFERENCES

[1] Akifumi Tokosumi, Naoko Matsumoto, and Hajime Murai, Medical Ontologies as a Knowledge Repository, *Proceedings of IEEE/ICME international conference on complex medical engineering*, 2007, pp.487-490.

[2] Ali. Adeli, Mehdi. Neshat,, A Fuzzy Expert System for Heart Disease Diagnosis. *Proceedings of the International Multi Conference of Engineers and Computer Science*, Vol I, IMECS 2010 ,Hong Kong , March 17-19 ,2010.

[3] Antonio J.Jara, Francisco J Blaya, Miguel A. Zamora and Antonio F.G. Skarmeta, An ontology and rule based intelligent information system to detect and predict Myocardial Diseases, *Proceedings of the* 9th *International Conference on information technology and Applications in Bio-medicine*, 2009.

[4] Dahu aXu, Yinsheng Zhang, Huilin Shan , Develop of diseases pest information system based on WEBGI , *Proceedings of International Seminar on Future Bio Medical Information Engineering* ,pp.461-464.

[5] Ersin Kaya, Bulent Oran and Ahmet Arslan , A Diagnostic Fuzzy Rule-Based System for Congenital Heart Disease, *World Academy of Science, Engineering and Technology*, vol 54, 2010 ,pp.253-256.

[6] Hongyi Zhang, Jiexin Pu, Junying Zhang, Construction of gene regulatory networks based on gene ontology and multivariable regression, *Proceedings of International conference on mechatronics and automation* , 2006, pp.1324-1328.

[7] Illhoi Yoo, Xiaohua HU , Biomedical Ontology MeSH improves Document Clustering Qualify on MEDLINE Articles :A comparison study, *Proceedings of the* 19th *IEEE Symposium on computer –based medical systems*,2006, pp. 577-582.

[8] Lynn Marie Schriml, Cesar Arze, Suvarna Nadendla, Yu-Wei Wayne Chang, Mark Mazaitis, Victor Felix, Gang Feng and Warren Alden Kibbe, Disease Ontology: a backbone for disease semantic integration , *Nucleic Acids Research'*, Vol 40 ,2011,pp.947-956.

[9] Maja Hadzic, Elizabeth Chan, Pronpit Wongthongtham, Robert Meersman, Disease Ontology based Grid Middleware for Human Disease Research Study, *Proceedings of 30th Annual Conference of the IEEE Industrial Electronics Society*, vol no. 1 , 2004, pp.480- 486.

[10] Maja Hadzic, Elizabeth Chang. Ontology-based Multi-agent System support Human Disease Study and Control, *Proceedings of the conference on Self-Organization and Autonomic Informatics* 2005, pp.129-141.

[11] Maja Hadzic, Elizabeth Chang, Ontology-based Support for Human Disease Study, *Proceedings of 38thHawaii International Conference on System Sciences*, 2005,pp.1-7.

[12] Mythili Thirugnanam, Mangayarkarsi, Paatabiraman, Sivakumar, Ontology based disease information system , *Proceedings of International conference on Modeling Optimization and Computing* ,2012, pp. 3235-3241.

[13] Shastri L. Nimmagadda, SashiK. Nimmagadda and Heinz Dreher, Ontology based data warehouse modeling and managing ecology of human body for disease and drug prescription management, *Proceeding of second IEEE International Conference* on *Digital Ecosystems and Technologies*, 2008, pp.212-220.

[14] Tharam S. Dillon, Elizabeth Chang, Maja Hadzic, Ontology Support for Biomedical Information Resources, *Proceedings of 21st IEEE International Symposium on Computer-Based Medical Systems*, 2008, pp.7-16.

[15] Tran Quoc Dung, Wataru Kameyama,, A Proposal of Ontology-based Health Care Information Extraction System: Vn HIES, *Proceedings of IEEE international Conference in research , innovation and vision for the future* , 2007, pp. 1-7.

[16] Yip Chi Kiong, Sellappan Palaniappan, Nor Adnan Yahaya, Health Ontology System, *Proceedings of 7th International Conference on IT in Asia* (CITA), 2011.