# A Methodology for Ontology Building

| M.Gawich | A.Badr | A.Hegazy | H.Ismail |
|----------|--------|----------|----------|
| French University | Cairo University | Arab Academy | Modern University |
| Shorouk,Cairo | Guiza | Sheraton,Cairo | Maadi,Cairo |
| Egypt | Egypt | Egypt | Egypt |

## ABSTRACT

Ontology is generally defined as an explicit specification of conceptualization which involves the exploration of concepts and its relationships in the domain of interest. Ontology is used to share knowledge across semantic web services, agents and information systems. Therefore the ontology engineer should follow a methodology during the ontology construction to ensure its reliability. This paper presents ontology building methodologies such as Uschold and king [3], Viral Hepatitis Ontology design Methodology [4] and Amaya Berneras et al [9].Moreover the paper presents Rheumatoid and Osteoarthritis ontologies which belong to the medical domain and a proposal for a new methodology applied to build the two medical ontologies. What distinguishes the proposed methodology is the execution of knowledge representation which involves the use of conceptualization and inference rules.

## General Terms

Ontology, Ontolgy building methodologies

## Keywords

Ontology Building, Knowledge Base, Medical Ontology, Rheumatoid, Osteoarthritis

## 1. INTRODUCTION

Ontology is defined as formal explicit specification of shared conceptualization [1]. Where " formal" indicates that the ontology should be machine readable and "shared" means that it is confirmed by a group or community.

Ontology is used by many dynamic applications [20] such as semantic web search engine, agents and information system. First for the semantic web search engine, the user should ask a question such as "find me the main treatment of Rheumatoid disease". To find the suitable answer, the semantic web search engine deals with agents. Agents are autonomous software entities, one of agent system components which help the agent to find the reliable answer is the knowledge base which can be represented by the ontology. Second for Information System [19], Ontologies can be used in IS as a repository used to store explicit and captured tacit knowledge in a specific domain.

The goal the ontology building methodology is to assure the reusability, the integration and the reliability of the ontology. Each methodology is based on the experience of ontology building in a specific domain. First for, Uschold and King Methodology has focused on the reusability and the integration of the built Enterprise ontology .Second for, Amaya Berneras et al methodology has focused on the application "KACTUS project" which the ontology is created for. Finally for, the viral Hepatitis

ontology methodology has focused on the intervention of experts in the early stage of ontology building and the integration. The proposed methodology is based on the building of Rheumatoid and Osteoarthritis ontologies in the medical field. It has focused on the intervention of experts, the use of explicit sources and the creation of inference rules which are not applied by other methodologies. Moreover the methodology explains the techniques which are applied to capture the tacit knowledge from experts.

This paper presents in section (2) the related works which shows a survey ontology building methodologies and investigates the Rheumatoid and Osteoarthritis diseases. Section (3) introduces a proposal for a new methodology for ontology building and section (4) presents a proposal framework for comparative analysis between Ontology building methodologies.

## 2. RELATED WORKS

## 2.1 Ontology Building Methodologies

The following methodologies are proposed by few research groups for developing ontologies building a specific ontology or based on individual experience of building a specific ontology.

### 2.1.1 Uschold And King Methodology

This methodology is based on the experience [3] of building "Enterprise "ontology .Phases for developing ontologies are:

### 2.1.1.1 Identify purpose

It refers to the motivation of the ontology building, benefits and its users. Whether the ontology will be shared and refined by a large community or a small group.

### 2.1.1.2 Building the Ontology

This phase will be outlined through the following steps:

### 2.1.1.2.1 Ontology Capture

a. Identification of concepts and its relationships.

b. Extraction of precise unambiguous text definitions for such concepts and relationships.

c. Identification of terms which refer the concepts and relationships.

### 2.1.1.2.2 Coding

It involves the representation of the captured concepts and its relationship in a formal language.

### 2.1.1.2.3 Integrating with Existing Ontologies

During the capture and coding of the ontology, the reusability of existed ontology is considered.

### 1.1.1.2 Evaluation and Documentation

Ontology should reflect the requirement specifications and competency questions. The established guidelines for documentation are based on the type and the purpose of ontology.

### 1.1.2 The Viral Hepatitis (VH) Ontology design Methodology

The methodology [4] is based on the building VH ontology in biomedical domain. The established phases are:

### 1.1.2.1 VH ontology Extraction Phase

The required knowledge is captured from several medical sources such as domain experts, medical book[5] and trusted websites[6,7,8]. After the knowledge extraction , the Hepatitis viruses (A,B,C) are studied. The symptoms , signs and laboratory finding are determined. The extraction phase had been passed through two steps:

### 2.1.2.1.1 Step 1: Mapping between the VH diseases and their symptoms/signs/laboratory findings

In this step each Viral Hepatitis is mapped to its corresponding symptoms, signs and laboratory findings.

### 2.1.2.1.2 Step 2 : Classifying the VH diseases and symptoms/ signs / laboratory findings.

It means the classification of  Viral Hepatitis into three classes: Symptoms, signs and laboratory findings .the symptoms class is classified to other subclasses according to the VH disease type .

### 1.1.2.2 VH Ontology Validation Phase

The domain experts have been consulted to validate the results of both VH diseases. The symptoms, signs and laboratory findings are also reviewed by them. The mapping and the classification steps are also reviewed. Also there are classes which are added to be suitable for the Ontology of Biomedical Reality (OBR) framework..

### 1.1.2.3 The VH Ontology representing in OWL phase

It involves the transformation of Viral Hepatitis ontology in OWL . The Viral Hepatitis classes and relations are represented as classification tree and representing the classes and relations of the merged Viral Hepatitis Diseases classification tree with OBR . The protégé –OWL is exploited to implement this phase.

### 2.1.3 Amaya Berneras et al

The methodology [3] is based on the use of KACTUS project[9]. One of its objectives is to investigate the feasibility of knowledge reuse in complex technical systems and the role of ontology is to support it. So every time an application is built, the ontology that represents the required knowledge for this application is also built.

The methodology is outlined through the following phases :

### 2.1.3.1  Specification of the application

It provides the application context and a view of the components that the application tries to model.

### 2.1.3.2  Preliminary design based on relevant top-level ontological categories

It involves the list of terms and tasks developed in the previous phase and can be an input to obtain several views of the global model.

The design process provides ontologies developed for other applications and can be refined and adapted for using it in the new application. Finally,the ontology refinement is used to achieve  a definitive design.

## 2.2 Medical Ontologies

Ontologies [2] are applied in the medical field to represent and reorganize medical terminologies. Physicians provide the terminology expressed by their own language to the ontology engineer in order to construct the medical ontology which helps them to share their captured knowledge.

The medical and healthcare information systems need to share complex specific and multilingual medical concepts. These issues require inference of meaning and deep understanding.

The construction of medical ontology can present reliable information; it enables the patient to share and access the required knowledge through the medical information systems.

The information system and knowledge base system which focus on the patient diagnosis should deal with the ontology which contains inferences rules concerning disease diagnosis.. Therefore Rheumatoid and Osteoarthritis Ontologies are built. Rheumatoid and Osteoarthritis symptoms seem similar, the recently graduated doctors can't decide if the patient has attacked by Rheumatoid or Osteoarthritis. Both of Rheumatoid and Osteoarthritis are chronic diseases which have not a real treatment

### 1.1.3  Rheumatoid

"Rheumatoid arthritis [10] (RA) is an inflammatory disease that affects the major  joints of the  human body, especially which are covered with synovium" The Rheumatoid usually attacks human who suffers from surplus of immunity".

 Rheumatoid usually attacks the pregnant woman and adult whom its age exceeds twenty nine. The recommended analyzes used to check the Rheumatoid are the CRP(C - reactive protein), Rheumatoid factor and Anti-CCP (Anti-cyclic citrullinated protein). The patient suffers from morning stiffness, arthritis hand joint and polyarthritis joints.

The treatments which are used to eliminate the Symptoms are the anticytokine, the DMARDS (disease-modifying antirheumatic drugs) and NSAIDs (Nonsteroidal Anti-inflammatory Drugs).

### 1.1.4  Osteoarthritis

"Osteoarthritis [11] is an immune disease that causes severe pain and swelling of joints. It can attack any joint, but usually it affects hands, knees and hips. Osteoarthritis can cause erosion of the joint cartilage"

Osteoarthritis usually attacks the pregnant woman and adult whom its age exceeds twenty nine. There is no specific

analysis used to diagnose. The patient suffers from stiffness, pain in the neck, polyarthritis joints and muscle weakness. The treatments which are used to eliminate the Symptoms are the Anticytokine and NSAIDs (Nonsteroidal Anti-inflammatory Drugs).

# 3. NEW PROPOSED METHODOLOGY FOR ONTOLOGY BUILDING

Since the ontology is defined as formal explicit specification of shared conceptualization where formal implies that the ontology should be machine-readable and shared that it is accepted by a group or community. Moreover it is a method of knowledge representation; the new proposed methodology will be based on the knowledge capture process.

As Fig.1 shows the established phases to build the Rheumatoid and Osteoarthritis ontologies are:



**Fig.1 Ontology building methodology**
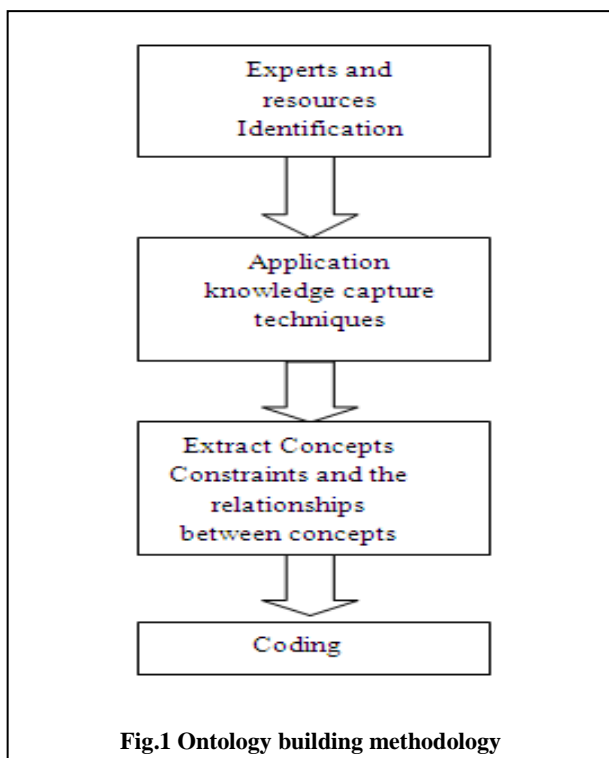
## 3.1 Experts and Resources Identification:

There exist two approaches to identify experts, the first one is to choose a single expert, and another one is to choose more than one expert "multiple expert" in the field of interest. Each approach has its advantages and its drawbacks [12], the second approach was chosen, because it is more suitable to the ontology which its validation must be established by a group. Concerning the resource there is a trusted software program created by Novartis called "Up-To-Date". It is used to capture general diseases symptoms which are also reviewed by the experts.

## 3.2 Application of Knowledge Capture Techniques:

### 3.2.1 Interview

The interview is commonly used in the early stages of tacit knowledge capture [15]. The interview includes three types of questions.

The first is the structured questions which are used to capture specific information. Fig.2 shows a sample of this kind of questions which is the dichotomous (Yes/No).

The second is the unstructured questions; it is used to explore an issue or concepts. Example of these questions:

- What are the symptoms of Rheumatoid?
- What are the symptoms of Osteoarthritis?
- What is the difference between Rheumatoid and Osteoarthritis?
- What is the treatment of Rheumatoid?
- What are the rheumatoid complications?

The third is the semi-structured technique; it enables the knowledge developer to ask predefined questions and allows to the experts some freedom to express his answer. Example of these questions:

- What are the major symptoms of Rheumatoid?
- How you make sure that this person suffers from severe Rheumatoid?



**Fig.2 Unstructured Questions**
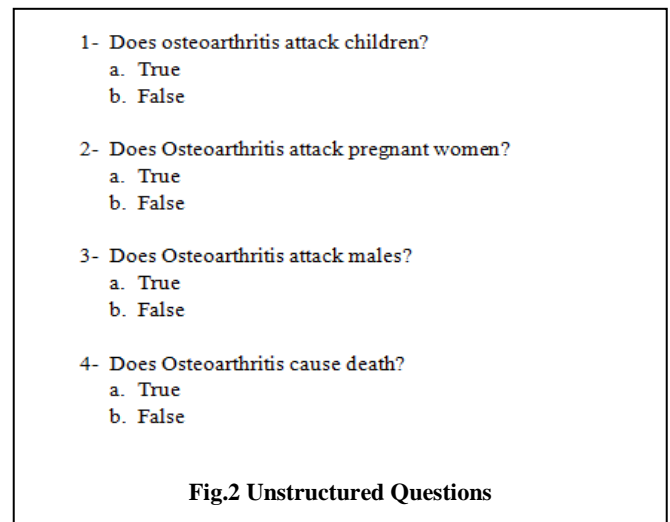
### 3.2.2 Protocol Analysis

It enables [14] the expert to think aloud when he solves a problem, the knowledge engineer should capture the strategies the expert follow to solve the problem.

Example of these questions:

- How to avoid Rheumatoid complications?
- How do you know if that person has attacked by osteoarthritis or rheumatoid?

### 3.2.3 Delphi Method

It involves [13] series of questions which are used to pool the expert' responses in order to solve a difficult problem. Each expert's contributions are shared with the rest of experts by using the results from each series of questions to build another one.

Example of these questions:

- If glucocorticoids cause unbearable pain to the Rheumatoid patient, how can you solve this problem ?

• If a woman is suffering from Kidney failure and during her pregnancy, she was attacked by the Rheumatoid , how can you solve this problem? What is the most suitable treatment in this case?

### 3.2.4  Concept,Axioms extraction and codification.

The concepts are extracted from the knowledge capture techniques. The relationships between concepts are created, and formally codified by the use of Protégé ontology editor as Fig.3 and Fig.4 show. On the left panel, the concepts are defined as classes and the relationships between classes are defined as object property. The Fig.3 presents also the Rheumatoid ontology ,while the Fig.4 presents    the Osteoarthritis ontology .

Concerning the axioms are used  or writing and storing inference rules"SWRL" in the ontology. The Fig.5 shows an inference rule that explains the patient is attacked by the rheumatoid if the analysis results show that "Anti_CCP" and "Rheumatoid_factor" exist . Through these  rules the knowledge base system can identify whether the patient is attacked by the Rheumatoid or not.

## 4.  PROPOSAL FRAMEWORK FOR COMPARATIVE STUDY

Points of comparison that are used to point out differences and similarities between Uschold & king methodology, VH ontology design methodology, Amaya Berneras methodology and the proposed methodology:

### 4.1 Experts' consultation

It involves if the  built ontology based on the extract of explicit knowledge which is found in  websites, books or software programs or  based on capturing tacit knowledge from one or more consultant .

### 4.2 Application Dependency

It involves if the goal of the ontology building is to be usable by a specific application and cannot be reusable by other applications .

### 4.3 Modes of development

There are three modes of development [16], the top level which reflects the determination of general concepts and moving to the specific concepts.
The bottom-up method deals with the subclasses identification and then grouping them into global classes, the middle-out method[17] defines most principal concepts and after that knowledge engineer decide to apply the top-down method or the bottom-up method[6] .

### 4.4 Inference rules use

 Since the ontology is a tool to represent knowledge, knowledge in the ontology is represented not only in the form of concepts sub concepts and relations hierarchy but involves also several facts and rules "If... Then …" [18] Those facts and rules are encoded by the SWRL (Semantic Web Rule Language) language.

**Table1. Comparision Between Methdologies**

| | Uschold & King | VH | Amaya Berneras | Proposed methodology |
|---|---|---|---|---|
| Experts consolidation | X | √ | X | √ |
| Application dependency | X | X | √ | X |
| Development modes | Middle-out | Bottom-up | Top-down | Middle-out |
| Inference rules (Axioms) | X | X | X | √ |

The above table shows first the viral hepatitis ontology design methodology and the proposed methodology mention the intervention of experts during the ontology building while others methodologies didn't mention. Second for the application dependency, the Amaya berneras methodology is based on an ontology which its motivation is to be usable by the Kactus project only, while others methodologies are used to build independent ontologies which can be used by any ontology tools or agents. Third for the development modes, the middle out mode applied by the "Uschold and king" and proposed methodology, the middle out involve the implementation of top-down mode and the bottom up after. Finally for the use of inference rules, only the proposed methodology implements inference rules.
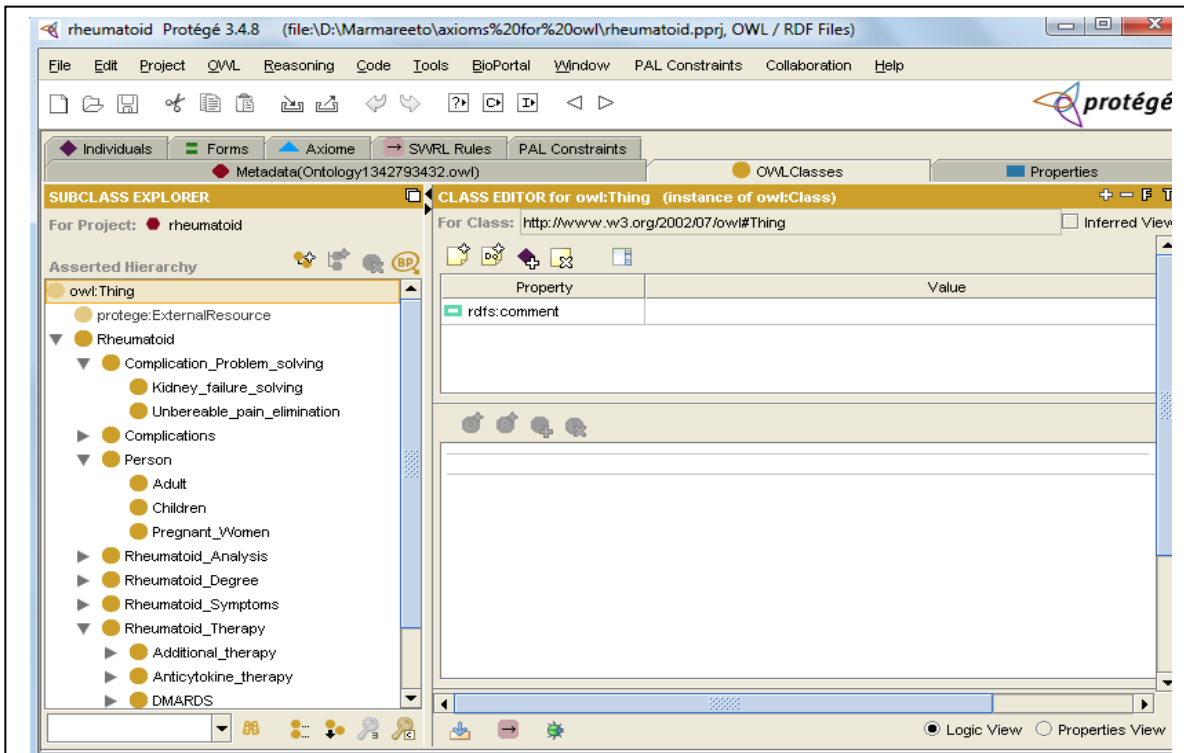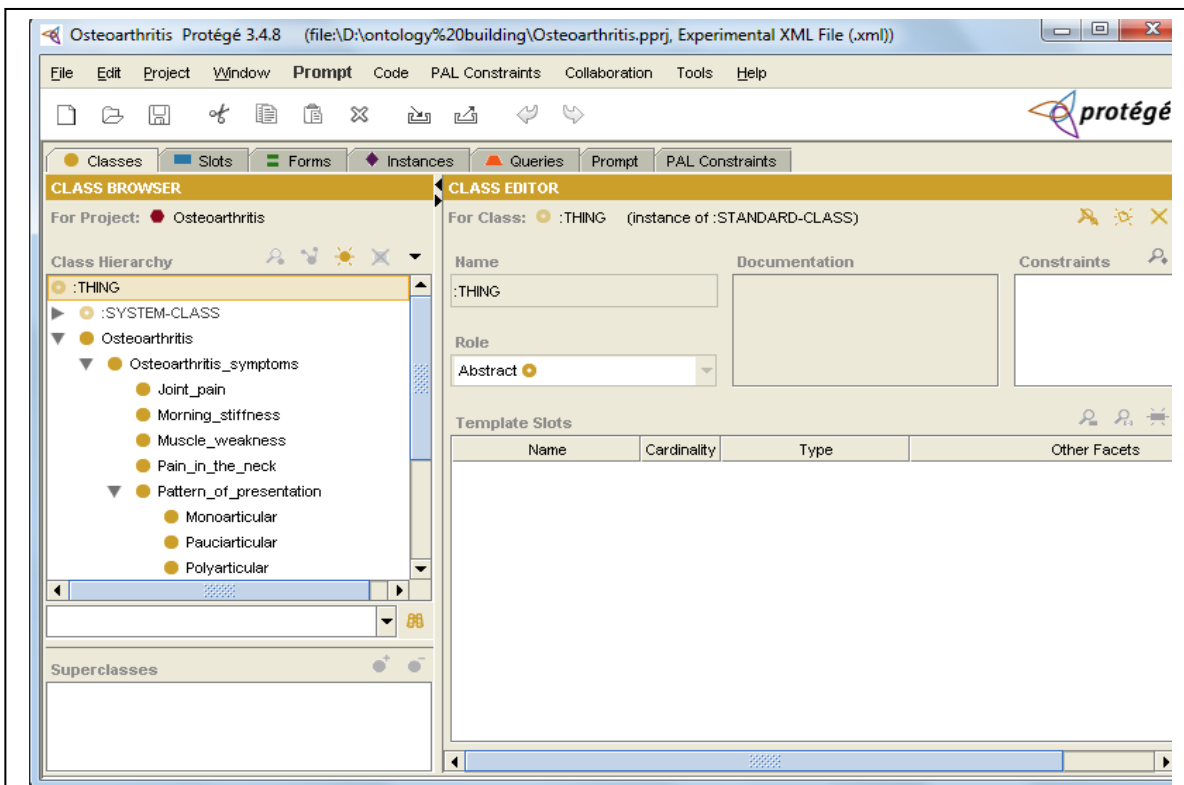
**Fig.3 Rheumatoid Ontology**
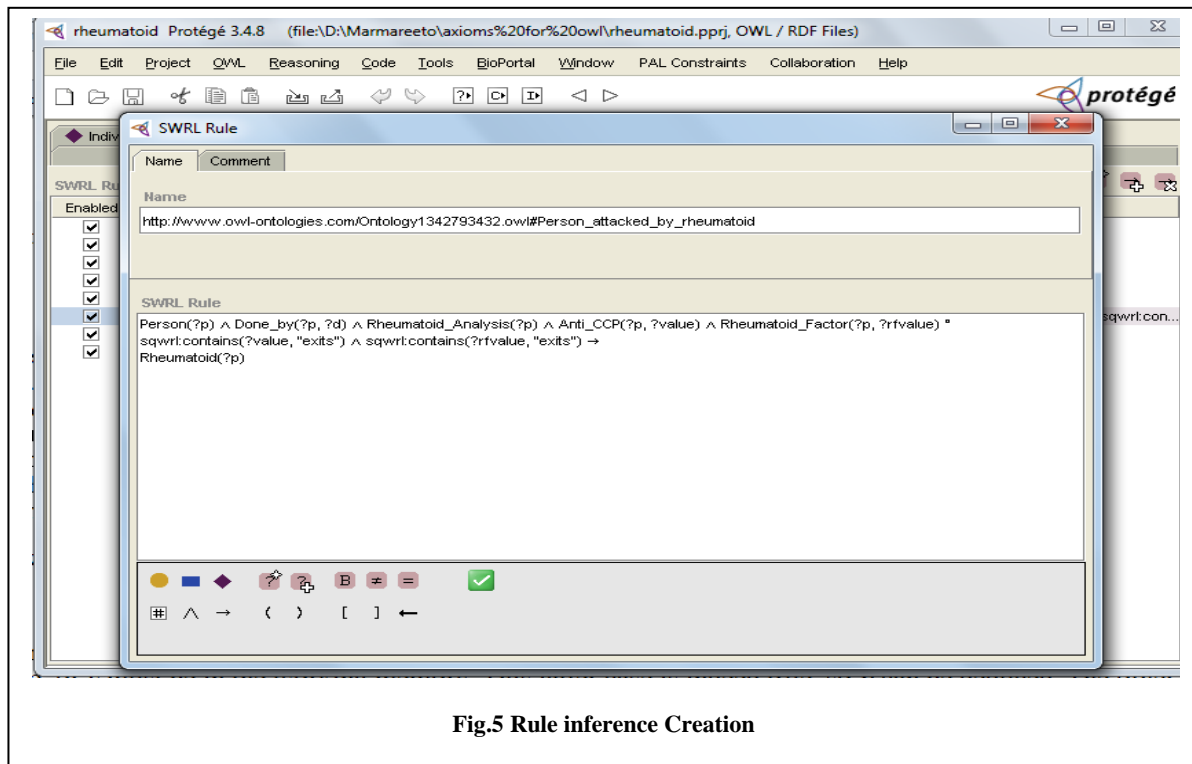


**Fig.4 Osteoarthritis Ontology**

**Fig.5 Rule inference Creation**

## 5. CONCLUSION

There is no one correct way to model a domain of interest, the best one depends on the motivation which the ontology is created for. ontology should be updated to be suitable for the web applications which are dynamic. Ontology concepts should reflect the classes and relationships used in the domain of interest. The proposed methodology involves expert's consolidation; explicit resources and the application of knowledge capture techniques increase the ontology reliability. Also the use of inference rules in the built ontologies reflects its integration with knowledge base system. Rheumatoid and Osteoarthritis ontologies are built to be used by any tool in the ontology domain and by the agents.

## 6. ACKNOWLEDGEMNTS

## REFERENCES

[1] T.Gruber, Towards principles for the design of ontologies used for knowledge sharing.Int.J.of Human and Computer Studies, 43:907-928,1994.

[2] http://www.openclinical.org/ontologies.html# definition . Retrieved on February 2012.

[3] F.lopez,M. Overview of methodologies for building ontologies. Fernández López, M. Laboratorio de Inteligencia Artificial. Facultad de Informática .1999.

[4] I.F.Moawad,G.Almarzouki,A.Salem. Building OBR-Based OWL Ontology for Viral Hepatitis. Egyptian Computer Science Journal,ECS,Vol.36 No.1, January 2012.

[5] Nicolas A.Boon,Nicki R.Colledge&Brian R.Walker.(2006).Davidson's Principles &Practice of Medicine.Pages 935-994,20th edition.

[6] http://www.emedecinehealth.com/healthtopics/article_em.html,10 OCt,2010.

[7] http://diseasesviruses.suite101.com/article.cfm/hepatitis-c---prevention-symptoms-and-treatment,22 Oct,2010.

[8] http://hepatitis .emedtv.com/hepatitis-d/hepatitis-d-symptoms.html ,27 Oct,2010 .

[9] G. Schreiber, B.Wielinga, W.Jansweijer. The KACTUS View on the 'O' Word. In IJCAI Workshop on Basic Ontological Issues in Knowledge Sharing, 1995.

[10] Arthritis foundation, http://www.arthritis.org/disease-center.php?disease_id=32, retrieved on May 2012.

[11] Medline Plus. A service of the U.S National library of medicine, http://www.nlm.nih.gov/medlineplus/osteoarthritis.html , retrieved on May 2012.

[12] E.M.Awad,H.M.Ghaziri,KnowledgeManagement,Chapter 5: Capturing tacit knowledge,P.129P130,2004.

[13] E.M.Awad,H.M.Ghaziri,Knowledge Management,Chapter 5:Capturing tacit knowledge ,P.170-P171,2004.

[14] E.M.Awad,H.M.Ghaziri,Knowledge Management,Chapter 5:Capturing tacit knowledge ,P.161-P163,2004.

[15] E.M.Awad,H.M.Ghaziri,Knowledge Management,Chapter 5: Capturing tacit knowledge,P.138-P142,2004.

[16] Minyoung Ra, Donghee Yoo, Sungchun No, Jinhee Shin, and Changhee Han .The Mixed Ontology Building MethodologyUsing Database Information. Proceedings of the International MultiConfernce of Engineers and Computer Scientists 2012 Vol1 , IMECS 2012, March 14-16,2012, Hong Kong.

[17] M. Grüninger, M. S. Fox, "Methodology for the design and evaluationof ontologies," In Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing held in conjunction with IJCAI-95,Montreal, Canada, 1995.

[18] E.M.Awad,H.M.Ghaziri,Knowledge Management, Chapter 4: Knowledge Creation and Knowledge Architecture,P.107,2004.

[19] B.Yilidiz, S.Miksch. Ontology Driven Information Systems :Challenges and Requirements . International Conference on Semantic Web and Digital Libraries. Indian Statistical Institute Platinum Jubilee Conference Series 2007 35–44 (2007).

[20] J.Euzenat, P.Shvaiko.Ontology matching,pages 9-9,2007.