An Algorithm for Merging Two Ontologies: A Case Study

Swati Negi* and Sanjay Kumar Malik**

*M.Tech., CSE, USIC&T, Guru Gobind Singh Indraprastha University, Sector-16C, Dwarka, New Delhi-110078, India.
**Associate Professor, USIC&T, Guru Gobind Singh Indraprastha University, Sector-16C, Dwarka, New Delhi-110078, India.

Abstract

Nowadays, huge volumes of data present in various forms is accumulating on the web every day. This poses a problem in handling and in extraction of information effectively on the web. Semantic web, the future smart web, may play a significant role by providing aid in processing the vast data on the web using machine intelligence wherein information resources are in the form of concepts and relationships with the aid of ontologies. Ontologies are used for sharing information and knowledge representation i.e. conceptualization of a particular domain. There exists knowledge represented by various ontologies which are scattered and therefore makes analysis and usage a difficult task. Therefore, there is a crucial need to merge these ontologies which is abstract and a challenge. In this paper, first, an attempt has been made to explore the significance of ontology and its two most fundamental operations namely ontology creation and merging, along with a brief concerned literature survey. Second, two ontologies, namely, ‘Yoga Ontology’ and ‘Person Ontology’ have been developed as a case study in Protégé 5.2.0 tool where various graphs are presented as an analysis and information. Third, an algorithm and flowchart is proposed along with the steps illustrated with the help of screenshots for merging these two ontologies using ‘Protégé Refactor’ feature. Fourth, ‘DLQuery’ of Protégé has been implemented to illustrate certain significant results and finally, problems faced along with limitations in merging operation using protégé 5.2.0 have been discussed. This paper is a guide for students and ontologists for ontology creation and merging using certain advanced features of Protégé 5.2.0 especially focusing yoga sector of health.

Keywords: Semantic web, Ontology, Protégé 5.2.0, Creation, Merging, DLQuery, OntoGraf.

INTRODUCTION

With the growth in size of web and its technological developments, large volumes of data, is available to humans, which is becoming difficult to manage. This data is not understood by machines or software agents and the search remains keyword-based only as a bottleneck of the present web 2.0. Semantic web, the next generation intelligent web, aims at overcoming this by its relationships to be understood by machines, so that the desired information can be extracted efficiently. The term ‘Semantic Web’ was coined by Sir Tim Berners-Lee who is regarded as the father of web and is presently the Director of World Wide Web Consortium (W3C), honoured with Turing Award. It refers to Web 3.0 which primarily focuses on Resource Description Framework (RDF) and ontology to represent data [19]. The main concern is to have an intelligent search that can understand the meaning of data and can differentiate among keywords according to the context in which they are used. Ontology is one of the most integral part of semantic web applications and plays an important role in knowledge sharing or information exchange by providing a common vocabulary of a specific domain of interest in a format which is machine understandable [23][26]. It serves as a backbone for embedding semantics to the existing web. It has various issues like creation, merging, integration, mapping, etc. Ontology creation and merging are the most fundamental key issues which are the focus of this paper. For ‘creation’ and ‘merging’ operations, various tools or methodologies may be used. Protégé is one of the most widely used tool for ontology creation and merging whose latest version 5.2.0 has been used here. This paper focuses on ‘creation’ and ‘merging’ ontology operations illustrated with a case study of ‘yoga’ ontology and ‘person’ ontology.

According to Jorge Cardoso survey 2007, it was found that ontologies are most popularly used in Education domain but in recent times, ontologies have gained more popularity in the medical and health sector also. The Yoga ontology, in this paper, provides the health benefit of yoga for specific disease, which makes it desirable to be implemented. Yoga is a better way for staying healthy than the use of medicines for many diseases. This makes the selection of this domain justified, and this ontology may be used as a guide for future health ontology, which may include wider aspects of health.

In this paper, in second section, ontology has been revisited along with its various issues and stages of development process. The concerned literature survey has been presented focusing Cardoso survey 2007. In third section, a case study of “Yoga Ontology” has been illustrated with a class hierarchy and various graphs. And another case study of “Person Ontology” has been illustrated in the fourth section. In fifth section, an algorithm is proposed for merging of the two said ontologies along with a flowchart. Merging has been illustrated using “refactor” option of protégé 5.2.0 where various steps of merging have been illustrated and DLQuery has been implemented to obtain significant results.
ONTOLOGY

An ontology\(^1\) includes concepts (or classes) related to a domain which are defined with relationships among them. This makes it possible to share information based on instances (or individuals) of the concepts. Also, it can be used for knowledge representation as it shows relation among the terms in consideration with the help of classes, properties and instances.

Issues

Various issues or operations, related to ontologies are as follows as shown in Fig.1 [18].

1) **Creation**- It is the designing and development of an ontology from scratch. In this context, Ontology Engineering is significant, which refers to all things that are needed in the process of design, development and maintenance of an ontology. In other words, it includes methodologies, tools, languages etc.

2) **Merging**- It refers to building an ontology by combining knowledge from two or more ontologies into one.

3) **Integration**- It refers to the creation of a new ontology by reusing existing ontologies.

4) **Selection**- It refers to the selection of an existing ontology for reuse according to its topic coverage, popularity, semantic etc.

5) **Mapping/Matching/Alignment**- It refers to mapping classes/entities of various ontologies and matching them for further use. These can be aligned, if needed.

6) **Maintenance**- There is a problem of evolution in the existing old ontologies due to complexity of the changes it requires. Therefore, maintenance is required to ensure the consistency of the ontologies.

7) **Import**- This refers to the function of including an external ontology in current one.

8) **Export**- Its function is to use current ontology into others.

9) **Debugging**- This refers to the searching of inconsistencies in the definitions of the concepts.

Here, creation and merging issues are focussed.

Ontology Development process

Ontology development is an iterative process and there is a need to follow some crucial steps to create a good ontology. This process includes the following steps, as depicted in Fig.2.

---

Specification level
At this level, the goal of ontology needs to be completely understood, and the domain which the ontology will cover needs to be described. Key questions that can be asked are required to be determined so that the purpose of the ontology is understood clearly. When the basic knowledge about the ontology is acquired, the granularity and type of ontology may be decided [6].

Conceptualization level
At the conceptual level, classes related to domain are decided, the relations between these classes are set and the attributes are identified. Instances i.e. individuals of classes are created, which are like objects of a class. Any query related to classes will be called through objects [6].

Implementation level
At this level, actual implementation (creation of classes and properties are set) is done using a tool or methodology. The ontology is checked for inconsistencies. In case of Protégé, reasoners may be used for this purpose. Finally, the question which were asked at the specification level are validated at this level [6].

Properties of OWL
Properties of OWL represents relationships. There are basically three types of properties that are used in creation/generation of ontologies which are [6]:

1) Annotation property
   Annotation property is used to add description to the classes, individuals and object/data type properties.

2) Data type property
   Data type property is required to link individuals to an XML schema datatype or an RDF literal.

3) Object property
   Instances can relate to each other by object properties.

Related Literature Work
Gruber (1993) defined Ontology as “concepts, relationships and other distinctions that are important for modelling a domain” [10]. It takes specification of representational vocabulary which provides meaning to the vocabulary and formal constraint on its coherent use.

Borst (1997) extended Gruber’s definition of Ontology for showing its utilisation in reuse as “A formal specification of a shared conceptualization” [5].

Brickley and Guha [1999] through W3C has given the Resource Description Framework, a ontology language for encoding knowledge on Web pages to make it understandable to electronic agents for searching information ².

McGuinnes and Harmelen [2004] described OWL (Web Ontology Language). When the information in a document is needed to be processed by an application, then OWL can be used, which is opposite of the condition where information is to be presented only. It can be used to provide the meanings to the terms in vocabularies explicitly and the relationships between those terms. OWL may express meaning and semantics more efficiently than XML, RDF, and RDF-S, and therefore it goes forward with the ability to represent machine understandable content on the Web. OWL is an extension of the DAML (DARPA Agent Markup Language) and OIL (Ontology Interchange Language) web ontology languages [5].

According to Jorge Cardoso’s survey of 2007 [7], ontology is primarily used for understanding the structure of information by people and software agents, as also presented in Fig.3:

---

methodology to design and develop an ontology.

Another finding of Cardoso’s survey was that OWL is the popularly used ontology language, as presented in Fig.5 [7].

Cardoso, through his survey [7], also deduced that most users prefer Protégé tool as an ontology editor, as shown in Fig.6.

Figure 4: Usage of type of Methodology  
[Source: Cardoso, 2007]

Figure 5: Usage of Ontology Language  
[Source: Cardoso, 2007]

Figure 6: Usage of Ontology Tools [Source: Cardoso, 2007]

Khondoke[16] also concluded that Protégé is the most dominant and domain-independent tool which is used by 75% of the respondent of the survey conducted by him. Fig.7 shows his result for 32 respondents out of which 24 respondents used Protégé tool.

Figure 7: Usage of Tools [Source: Khondoker et al., 2010]

Yoga Ontology Creation using Protégé 5.2.0

Protégé

Protégé\(^3\) has been developed at Stanford University as an ontology construction tool which is very popular for its user-friendly graphical user interface to create and edit ontologies. It provides a number of functionalities with the help of plug-ins like creation of graphs (e.g. Ontograf, OwlViz etc.), multimedia support, querying (with the help of DLquery and SPARQL) and reasoning engines (e.g. FaCT++ , HermiT, Pellet, jcel etc).

---

\(^3\)https://protegewiki.stanford.edu/wiki/Main_Page
problem solving methods, etc. It can be used for additional services like it enables its users to create, manipulate, visualize and query ontologies in various representation formats like RDF/XML, OWL/XML, turtle, OBO etc. It is the most widely used tool which enables for ontology development according to the Cardoso’s Survey of 2007 [11] and as inferred by Rahamatullah [12].

Yoga Ontology creation

First ontology named as ‘YOGA’ contains ‘YOGA_POSES’, ‘CLEANSING_TECHNIQUES’, ‘PRANAYAMA’ and ‘DISEASE’. YOGA ontology establishes a relationship between the yoga poses that are beneficial for a specific disease. Its graph has been depicted using Ontograf, which is a plug-in of Protégé 5.2.0, as shown in Fig. 8 and its class hierarchy has been shown in Fig. 9.

In YOGA ontology, the DISEASE class is very important and contains various instances like liver, thyroid, arthritis, diabetes etc. These instances are important for applying queries onto a particular disease. Fig.10 shows instances of DISEASE class through an OntoGraf.

YOGA ontology has YOGA_POSES as class which is further divided into sub-classes like Backward_bends, Twists, Side_bends, Stretches, Forward_bends, meditative_poses, Inversions, Relaxing_poses and Balancing_poses. It also contains many instances for various Yoga poses like chakrasana, tadasana, halasana, bhujangasana, garudasana etc. These instances are important for applying queries for yoga poses. Fig.11 below is showing instances of DISEASE class through OntoGraf.
Table 1: Inferences/usefulness of YOGA Ontology

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Inferences/Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Step by step instructions have been provided for creating an ontology which is quite useful for new ontology developers.</td>
</tr>
<tr>
<td>2.</td>
<td>Step by step instructions have been provided for merging two ontologies which makes it easier to understand for new ontology developers.</td>
</tr>
<tr>
<td>3.</td>
<td>Queries have been presented to suggest Yoga Poses to cure a particular disease.</td>
</tr>
<tr>
<td>4.</td>
<td>Diseases which are cured by a yoga pose are found through queries.</td>
</tr>
<tr>
<td>5.</td>
<td>Person suffering from a particular disease can be identified using query.</td>
</tr>
<tr>
<td>6.</td>
<td>The resultant YOGA ontology can be used to extend health ontology on a large scale.</td>
</tr>
</tbody>
</table>

**Creation of PERSON Ontology**

The second ontology named as ‘PERSON’ contains ‘Person’ and ‘Disease’ as its classes. The graph for PERSON ontology has been depicted in Fig.12 using Ontograf. This ontology establishes a relationship between the person and the disease they have.
Merge Operation: An algorithm and flowchart proposed

Ontology Merging refers to building an ontology by combining knowledge of two or more ontologies into one. The algorithm for merging two ontologies is given in Table 1 and the flowchart is given in Fig.13.

The second ontology i.e. ‘PERSON’ Ontology is merged onto the first ontology i.e. ‘YOGA’, to get the resultant merged YOGA ontology. This resultant merged YOGA ontology will have ‘YOGA POSES’, ‘CLEANSING TECHNIQUES’, ‘PRANAYAMA’, ‘DISEASE’ along with ‘person’ as classes.

Table 2: Algorithm for merging two Ontologies

<table>
<thead>
<tr>
<th>Input: First Ontology(YOGA), Second Ontology(PERSON)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output: Resultant MergedOntology (YOGA Ontology with PERSON Ontology)</td>
</tr>
</tbody>
</table>

1. Open Second Ontology (Ontology to be merged) in tool(Protégé 5.2.0).
2. Open First Ontology (Ontology in which merging has to be done) in the same window in Protégé 5.2.0.
3. Check for similarities.
4. Select ‘Merge Ontologies’ in Refactor Menu.
5. Then, select ‘Merge into existing ontology’ radio button.
6. Select First Ontology as Target Ontology.
7. Resolve inconsistencies in Resultant Merged Ontology by changing the Full URIs of conflicting classes and individuals.
8. Check for consistency of Resultant Merged Ontology using a reasoner.
5.1 Steps for Merging (‘YOGA Ontology’ and ‘PERSON Ontology’)

**STEP 1:** Open the ontology which is required to be merged i.e. PERSON ontology. Then, Open the Yoga ontology in which merging has to be done. The Yoga ontology must be opened in the same window [3] [4].

**STEP 2:** Select Refactor and choose “Merge ontologies”. Here, from YOGA ontology and PERSON ontology, select YOGA ontology, as depicted in Fig.14 [3].

**STEP 3:** Select Merge type. Merge can be performed with new ontology or existing ontology, as depicted in Fig.15 [4].

**STEP 4:** Select the Target ontology. In this work target ontology is Yoga. After merge operation, Yoga ontology will be modified to carry data of PERSON ontology along its own data, as depicted in Fig.16 [4].
STEP 5: A merged ontology can now be seen with classes of merged class, as depicted in Fig.17 [4].

Figure 17: Screenshot for partially merged ontology-YOGA

STEP 6: Start a Reasoner to sort out inconsistencies in the merged ontology. However, reasoner does not sort out the inconsistencies completely [4].

STEP 7: As two different classes of “DISEASE” and “disease” representing the same thing are present in the resultant ontology, there is a need to change the name of one of the class and rename it to the name of the other class (suppose disease renamed as DISEASE), to avoid any conflict.

STEP 8: Change URI of “disease” class by selecting “Rename Entity” in Refactor menu, as depicted in Fig.18.

Figure 18: Screenshot for changing entity URI

STEP 9: It be noticed that Individuals in merged ontology is redundant as they belong to two different existing ontologies, like asthma and liver/liver_disorder, as depicted in Fig.19.

Figure 19: Screenshot for showing redundant individuals

STEP 10: Again, change the URI of individuals which are of the same type by selecting Rename Entity in Refactor menu, as depicted in Fig.20.

Figure 20: Screenshot for removing inconsistency of individuals
The code snippet of the merged ontology is as follows in Table 3:

### Table 3: OWL Snippet (merged ontology)

```owl
```

**Queries**

In Protégé, DLQuery may be implemented to get important results which can be applied to utilise this YOGA Ontology [20]. Some are as follows:

**QUERY-1**: Yoga Poses suggested in a particular disease. For example, as shown in Fig.21, yoga poses recommended in Back ache have been given through DLQuery.

**DLQuery**: YOGA_POSES and recommendedFor value back_ache

![Figure 21](https://via.placeholder.com/150)

**Figure 21**: Query for selecting YOGA_POSE which are beneficial for back_ache

**Inference**: Query have been presented to suggest Yoga Poses to cure a particular disease.

**QUERY-2**: Person who is suffering from a particular disease is depicted in Fig.22 through a DLQuery.

**DLQuery**: PERSON and sufferingFrom value thyroid

![Query](https://via.placeholder.com/150)

**Figure 22**: Query for selecting person who is suffering from a specific disease.

**Inference**: Person suffering from a particular disease can be identified using this query.

**QUERY-3**: Diseases which are cured by a yoga pose is found by a query as shown in Fig.23.

**DLQuery**: DISEASES and curedBy value bhujagasana

3) When two ontologies having a class with the same name are merged, that class will appear twice in the resultant ontology. This problem needs to be manually addressed.

4) The entity URI needs to be changed after checking the ‘Show full IRI’. If the full IRI is not changed, the change might not apply.

5) When two ontologies are merged, if there is a same instance in both the parent ontologies, it will appear twice in the resultant ontology.

6) To resolve the conflict of two same name classes in the merged ontology, there is a need to change the IRI of the class from the ontology that is to be merged onto the other ontology.

CONCLUSIONS & FUTURE WORK

The paper focuses on ontology creation and merging using Protégé 5.2.0 and tool available in it. This work uses case study of Yoga which includes class hierarchy and OntoGrafs. Yoga Ontology is useful in health sector and is a new concept in ontology. An algorithm and flowchart has been proposed in concern to a yoga ontology for merging. "DLQuery" has been implemented for efficient retrieval of yoga beneficial information to cure specific disease that can be processed by machine automatically. This work is beneficial for the students attempting to create and merge ontology and a guide to Ontologists to enhance the case study and an algorithm and flowchart proposed may benefit in this direction.

In future, different tools or methodologies may be used for creation and merging of two ontologies. Other issues of ontology like integration, mapping, matching etc may be explored. The enhancement of the said mini YOGA ontology may be taken further in depth and a full-fledged enhanced ontology on YOGA may be developed covering different aspects of yoga like Pranayama etc and this may be integrated in future with other ontologies of health sector.

REFERENCES


