

# Semantic Linking and Querying of Natural Food, Chemicals and Diseases

G. Vadivu

Department of Information Technology  
SRM University, Kattankulathur, 603203

S.Waheeta Hopper

Department of Bio Informatics  
SRM University, Kattankulathur, 603203

## ABSTRACT

The biological community requires more sophisticated methods to store, integrate and query their knowledge. The semantic Web technology enables to represent, share and discover heterogeneous data through formal ontologies. Semantic Web technologies also allow to link information thus moving from document centric idea of current Web to more fine grained semantic structures. The study is made on Ontology construction of Natural Food Resources, Chemicals and Diseases. These knowledge items are linked to commonly agreed relationships. This enables the retrieval of user-specific semantic querying for topics of interest. This paper describes the works carried out on ontology-driven knowledge base and semantic querying of nutritional factors of Natural Food sources.

## General Terms

Semantic Web

## Keywords

Semantic Web, RDF, OWL, Linked Data

## 1. INTRODUCTION

The etiology of chronic diseases is due to the chemical reactions of medicines or food in the physiological processes. The balanced intake of natural foods with rich in nutritional factor has impact on curing diseases such as cancer, diabetes, asthma, and hypertension etc.,. However, the molecular reactions and mechanisms of the food intake in the digestive process are to be studied for the recommendation of correct diet.

The rise of the Internet has influenced the biologists to store and share their data. However, these distributed data are independent and often used in closed environment. Though the Web is rich of content, the users find difficulty on retrieving data from the large set, which consumes more human effort. If these data were to be linked, it would be possible for finding more inferences. The current Web is hampered with handling advanced applications such as processing, understanding and semantic interoperability of information contained in several Web documents. This is due to the simple keyword based search retrieval, which can be understood by humans but not by machines. Semantic Web is the new generation Web that tries to represent information such that it can be used by machines, not just for display purposes, but for automation, integration, and reuse across applications (Berners-Lee 2000). Furthermore, semantic Web is about explicitly declaring the knowledge embedded in many Web based applications, integrating information in an intelligent way,

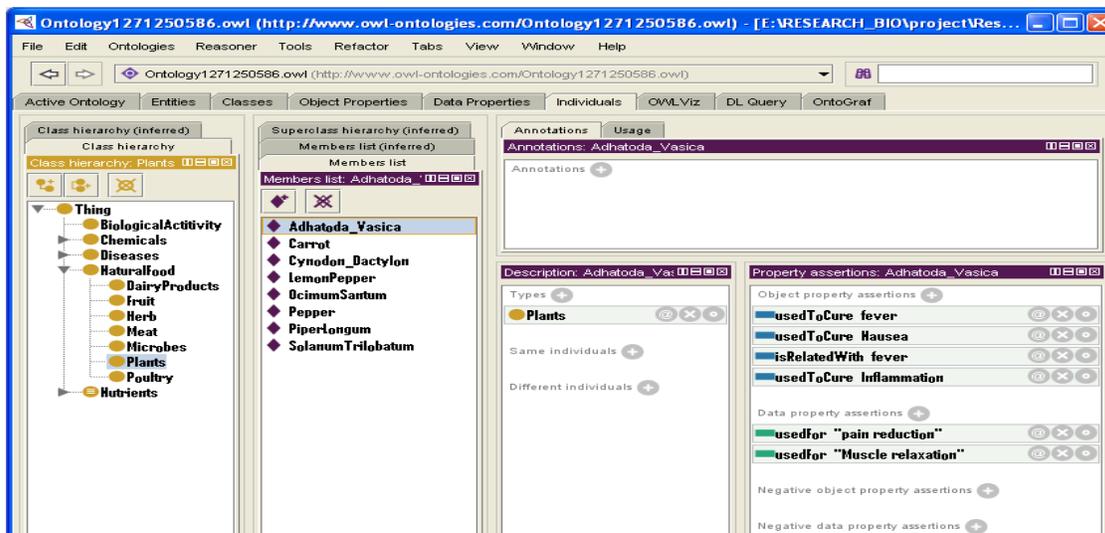
providing semantic based access to the Internet, and extracting information from texts.

Our study focuses on developing semantic relationships to the domains of Chemical Compounds, Diseases and Natural Food sources for the better querying and linking of information. The paper is structured as follows: section 2 describes about the choice of ontology language; section 3 describes Knowledge Construction; section 4 describes finding dataset similarities followed by querying linked datasets in section 5.

## 2. SEMANTIC WEB LAYER

Ontology is defined as explicit and formal specification of conceptualization. Ontology comprises a set of knowledge terms, including the vocabulary, the semantic interconnections, simple rules of inference and logic for some particular topic. Ontologies applied to the Web are creating the semantic Web. The Semantic Web architecture lists the underlying machine understandable languages for knowledge representation: XML (Extensible Markup Language), RDF (Resource Description Framework) and OWL (Web Ontology Language).

XML with XML namespace and XML schema definitions makes sure that there is a common syntax used in the semantic Web. When it comes to semantic interoperability, however, XML has disadvantages. Using XML, it is not possible to specify more relationships among the data. On top of XML is the RDF, for representing information about resources in a graph form. RDF, follows binary predicates with the format <Subject, Predicate, and Object> used to create the relationship in the form of triples. For example, “Adhatoda used to cure Fever”, <Adhatoda, usedToCure, Fever> is the triple form. RDF Schema (RDFS) defines the vocabulary of RDF model. It provides a mechanism to describe domain-specific properties and classes of resources to which those properties can be applied, using a set of basic modeling primitives (class, subclass-of, property, subproperty-of, domain, range, type). However, both RDF and RDFS are rather simple and still do not provide exact semantics of a domain. OWL is the next layer used in Semantic Web Architecture, is having more features compared to RDF. OWL is a set of XML elements and attributes, with well-defined meaning, that are used to define terms and their relationships (e.g. Class, equivalentProperty, intersectionOf, unionOf, etc.). Reasoning tasks like verification of ontology consistency, computing inferences and realizations can be easily executed with the OWL representation.



**Figure1. Class, Subclass, Individual and their Properties**

### 3. CONSTRUCTION OF KNOWLEDGE BASE

In practice, ontologies are often developed using integrated, graphical, ontology authoring tools, such as Protégé, OIled and OntoEdit. Protégé facilitates extensible infrastructure and allows an easy construction of knowledge rich domain ontologies. Protégé tool is used to develop domain ontology and querying since it has adopted the recent recommendation of the W3C, i.e. OWL standard. Protégé is also a knowledge based editor and it is open source Java tool that allows the easy construction of Bio Ontologies.

The effort is taken to provide an explicit specification of the conceptual model. Ontologies are a means to formalize explicit knowledge related to a specific domain. The ontology development focuses on developing design hierarchy towards the explicit specification of relationships. Initially, the identified components of Natural Food, Disease and Chemical compound are represented in the form of class hierarchy using protégé tool. Figure 1 shows the class, subclass, Individual and their properties creation using Protégé software. Figure 2. shows the Natural Food RDF/OWL code. The partial graphical representation of Chemical and Disease ontology is shown in Figure 3 and 4 respectively. The hyperlinks are added to navigate among html pages. The hyperlink updating is possible only for smaller set of datasets. However, it results cumbersome for larger datasets.

The semantic Web performance can be improved through linking of related entities from different datasets and establishing relationships automatically.

The ontology representation in Protégé is mapped on different dimensions. Semantic link is used to connect different datasets. The semantic linking provides more meaningful navigational paths to the users.

```
<!-- http://www.owl-ontologies.com/Ontology1271250586.owl#LemonPepper -->
<owl:NamedIndividual rdf:about="&Ontology1271250586;LemonPepper">
  <rdf:type rdf:resource="&Ontology1271250586;Plants"/>
  <hasSodium rdf:datatype="&xsd:string">Low</hasSodium>
</owl:NamedIndividual>

<!-- http://www.owl-ontologies.com/Ontology1271250586.owl#Milk -->
<owl:NamedIndividual rdf:about="&Ontology1271250586;Milk">
  <rdf:type rdf:resource="&Ontology1271250586;DairyProducts"/>
</owl:NamedIndividual>

<!-- http://www.owl-ontologies.com/Ontology1271250586.owl#Mushroom -->
<owl:NamedIndividual rdf:about="&Ontology1271250586;Mushroom">
  <rdf:type rdf:resource="&Ontology1271250586;Herb"/>
  <hasVitamin rdf:datatype="&xsd:string">Vitamin D</hasVitamin>
  <containsChemicalCompound rdf:resource="&Ontology1271250586;VitaminD"/>
</owl:NamedIndividual>

<!-- http://www.owl-ontologies.com/Ontology1271250586.owl#OcimumSantum -->
<owl:NamedIndividual rdf:about="&Ontology1271250586;OcimumSantum">
  <rdf:type rdf:resource="&Ontology1271250586;Plants"/>
  <isRelatedWith rdf:resource="&Ontology1271250586;fever"/>
  <usedToCure rdf:resource="&Ontology1271250586;fever"/>
</owl:NamedIndividual>

<!-- http://www.owl-ontologies.com/Ontology1271250586.owl#Pepper -->
<owl:NamedIndividual rdf:about="&Ontology1271250586;Pepper">
  <rdf:type rdf:resource="&Ontology1271250586;Plants"/>
  <cures rdf:resource="&Ontology1271250586;GeneralEyeDisease"/>
</owl:NamedIndividual>
```

**Figure2. Natural Food RDF/OWL Code**

### 4. LINK SIMILAR TERMS

The term ‘Linked Data’ refers to a style of publishing and interlinking structured data on the Web. The basic assumption behind Linked Data is that the value and usefulness of data increases the more it is interlinked with other data. Linked Data is simply about using the Web to create typed links between data from different sources. RDF links enable Linked Data browsers and crawlers to navigate between data sources and to discover additional linkages.

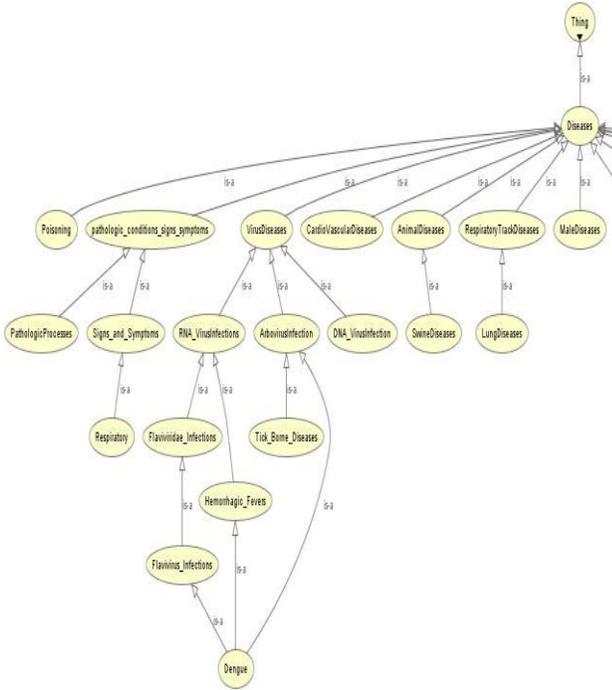


Figure3. Disease Ontology

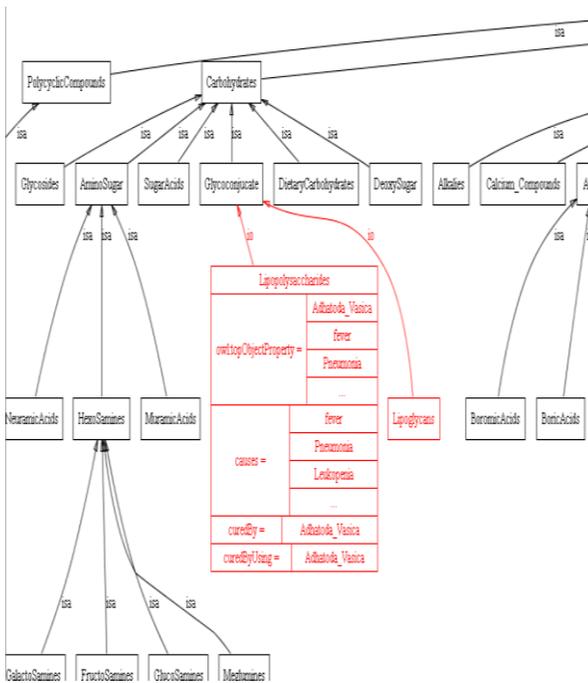


Figure4. Chemical Ontology

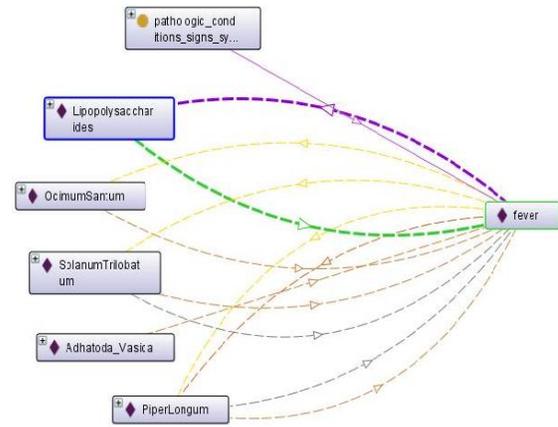


Figure5. Fever related links

The food intake which is the source of chemicals that can cure the disease can be easily found. Figure 5 illustrates the linking of Fever with chemical ontology and Natural Food ontology.

The ‘Fever’ instance of Disease can be cured with the intake of Natural Food ‘Adhatoda Vasica’ or the sources of chemicals which are rich in Lippopolysaccharides (sub-sub-class of Carbohydrates).

The linking is achieved by displaying and navigating the existing paths with the hierarchical relationships. More linkages can be established by adopting the methodology described below: Create node for all the Subjects in the RDF triple in the alphabetical order. Nodes can be created using linked list, each node with NextNode, Subject, and LinkToDirectObject as shown in Figure 6.

NextNode	Subject	LinkToDirectObject
----------	---------	--------------------

Figure6. Link Data Representation

- ‘NextNode’ points to list of all Subjects in ascending order.
- ‘LinkToDirectObject’ points to list of objects those are directly linked with the ‘Subject’ Node. ‘LinkToDirectObject’ has ‘LinkWithObject’, ‘Predicate’, ‘Object’ and ‘LinkToDirectObject’ as elements as shown in Figure 7.

LinkWithObject	Predicate	Object	LinkToDirectObject
----------------	-----------	--------	--------------------

Figure7. Dataset of ‘LinkToDirectObject’

It is advisable to make ‘LinkToDirectObject’ with minimum of five distance nodes. This list will help the system to identify whether the given values are related or not and also used to find more related nodes by following the alphabetical order of the nodes. Figure 8 shows the example to find the related nodes.

