International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA- 2013)

Building and Querying Microbial Ontology

Sreekumar Biswas[a], Sudeep Marwaha[a], P.K. Malhotra[a], S.D. Wahi[a], D.W. Dhar[b], Richa Singh[c]

[a]*Indian Agricultural Statistics Research Institute, Pusa, New Delhi, India*
[b]*Indian Agriculture Research Institute, Pusa, New Delhi, India*
[c]*AIM&ACT, Banasthali University, Rajasthan, India*

Abstract

The microbial taxonomy is based on the characteristics of microorganisms that can be objectively observed and measured. There are many scheme of microbial classification, but the latest is the three domain system and is the most accepted. Ontologies are the new form of knowledge representation that acts in synergy with agents and Semantic Web Architecture. Ontologies define domain concepts and the relationships between them, and thus provide a domain language that is meaningful to both humans and machines. The relationships in Ontology are explicitly named and developed with specification of rules and constraints so that they reflect the context of domain for which the knowledge is modelled. Ontologies can be built by using various GUI based software tools, known as Ontology editors. Among all editors Protégé is widely supported by a huge research community. For effective use of Ontology, protégé provides a query interface known as SPARQL query panel. SPARQL is a syntactically-SQL-like language for querying RDF graphs. Microbial Taxonomy Ontology is developed for the three domain system of microbes for the domain Bacteria which will be helpful for the study of Agriculturally Important Microbes (Bacteria). This ontology is built in the Protégé OWL editor from Domain to Genus level. Using this ontology, a query interface can be developed that will help detailed study of microbial taxonomy, classification of microbes as well as exchange knowledge between software agents and systems.

1. Introduction

There are many web based software developed in recent years that use ontology as their knowledge base. Example of such software are Gene Ontology [1], Plant Ontology [2], Geopolitical Ontology [3] etc. Gene Ontology was developed by Gene Ontology Consortium in 2000. They developed a tool called AmiGO for searching and browsing the gene ontology data. Plant Ontology is developed by Plant Ontology Consortium in 2002. Plant Ontology Consortium was formed in response to the need for a set of uniform terms to describe plant structures and developmental stages. Geopolitical Ontology was developed by Food and Agricultural Organization in 2002. It is an ontology describing geopolitical information. The geopolitical ontology includes names in multiple languages; maps standard coding systems; provides relations among territories; and tracks historical changes. In addition, FAO provides web services of geopolitical ontology and a module maker to download modules of the geopolitical ontology into different formats. A methodology [4] has been proposed for the conversion of taxonomies into Ontology. The proposed methodology is tested and implemented for a pilot soil ontology [5] using W3C standard Web Ontology Language and protégé 2.1 OWL plug-in. Ontology-based intelligent retrieval system for soil knowledge [6] is a system which searches the documents related to soils by using soil domain ontology. This system retrieves information like "Relationship between Laterite

soil and air pollution". Based on the concepts of the existing software, a strong need was felt for the microbial ontology should be prepared which will be helpful for students, teachers, researchers and the experts who are working in the field of microbiology.

2. Designing microbial ontology from taxonomy

The process of ontology creation from the scratch is a complex process and is expensive. As for relating each and every concept with one another, many iterations are required. But, it could be much easier if any existing source is used for the creation of ontology. Taxonomy is such an existing source that can be used for creating ontology. Taxonomy is a classified scheme, arranged in a hierarchy. An ontology generalizes this by adding additional relations among the elements of the taxonomy. Ontology engineering requires in depth knowledge of the domain as well as expertise in building knowledge representation and organization techniques, a rare combination that can be found for any domain. However, building ontologies have become significantly easierdue to the standardization of Web Ontology Language and tools such as Protégé 3.4.6. But for domain experts, especially in the fields that are not much concerned with the computers and information technology, it is still harder to use these tools and standards. On the other hand, there is another problem for the knowledge engineers, which is the problem of getting knowledge from the domain expert remains unsolved. In this section, efforts for devising an approach to create ontology from taxonomy for any knowledge domain are presented. This requires some knowledge source in the requisite domain and the existing standardized taxonomies are the right choice. Taxonomies are not just built by a single expert but are the result of sharing knowledge in the domain by many experts after investigating the field for years. Also, they provide hierarchical view of concepts of the domain that is required to build ontology. Protégé OWL Plug-in tool is used to demonstrate the implementation of generating Microbial Ontology from Microbial Taxonomy [7]. Microbial Ontology in OWL, can be easily ported to any system or to any other ontology editor having OWL support. This makes it easy to integrate with other ontologies or agent based systems that make Semantic Web a realization. For designing the microbial ontology from the taxonomy, the existing methodology [4] has been used.

3. Building microbial ontology from microbial taxonomy

For the Semantic Web technology, ontologies have come to be seen as a critical component. Protégé implements a rich set of knowledge-modelling structures and actions that support the creation, visualization, and manipulation of ontologies in various representation formats. Protégé can be customized to provide domain-friendly support for creating knowledge models and entering data. This section describes the development of Microbial ontology that acts as a knowledge base for the presented expert system and is used for querying and reasoning users' requests. By using the above approach, the Ontology on Agriculturally Important Micro-organisms in Protégé has been implemented. Moreover, this approach is generic and can be implemented with any tool that supports ontology implementation in OWL. The first step is to identify the hierarchy in which the microbial taxonomy is classified. The hierarchy level of the taxonomy is:

Domain ⟶ Phylum ⟶ Class ⟶ Order ⟶ Family ⟶ Genus

3.1 Creating Classes

- The class MicrobialTaxonomy is the topmost class created in the hierarchy. So it is placed as the subclass of the class owl:thing. Inside the class, there are three subclasses, Domain_Archaea, Domain_Bacteria and Domain_Eukarya.
- Next, the Domain_Bacteria class is populated with various Phylum classes as give in the Bergey's manual of Systematic Bacteriology.
- Then all the phylum classes are populated with its subclasses to populate the entire taxonomy.
- After creating the classes present in the taxonomy, the classes such as Basic_Characteristics, Other_Characteristics, Nutrition_type and Shape are created.
- Basic_Characteristics have several subclasses for basic characteristics of all the phylum present in the microbial taxonomy. Similarly the other classes created in the taxonomy have own subclasses.
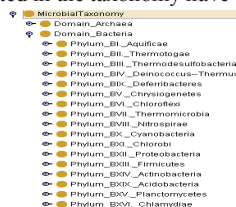

Fig. 1. Class Hierarchy of Microbial Taxonomy in Protégé

3.2 Creating Individuals

Once the classes has been created, the individuals of the classes have to be created. The individuals are the members of a class. In protégé, the individuals are created in the individual editor. All the individuals of the Microbial Taxonomy were created with the same name of their respective classes, but the letters in the names are in lower case. It is because of the reason that more than one resource in an ontology cannot share same name. As for example, the class Domain_Bacteria has an individual named domain_bacteria. Phylum_BI._Aquificae has an individual named phylum_bi._aquificae. Similarly, all other classes in the ontology have their individual named as the class, but all the letters are in lower case (shown in Fig. 3).
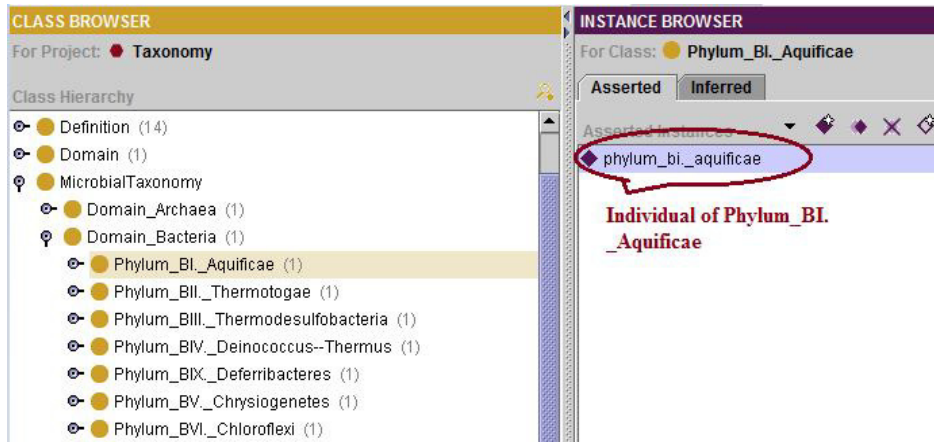

Fig. 2. Individual of a Class

3.3 Creating properties

The ontology is enriched with the classes and their individuals. Now the identified properties have to be incorporated in the ontology. The properties are created in the property browser of the protégé OWL plug-in. Here the Domain and Range of each property have to be declared. Domain refers to the class where the property have to be applied and Range refers to the class from which the property values to be picked. For example, the property hasShape is an object property and it has its Domain the Domain_Bacteria class and Range the Shape class. Similarly the other classes are created and their Domain and Range are also specified.
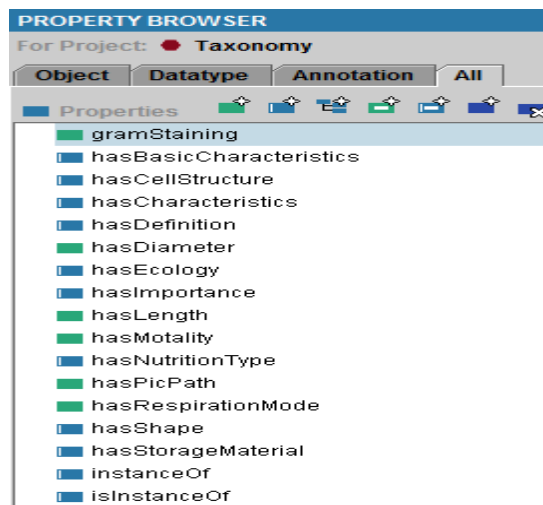

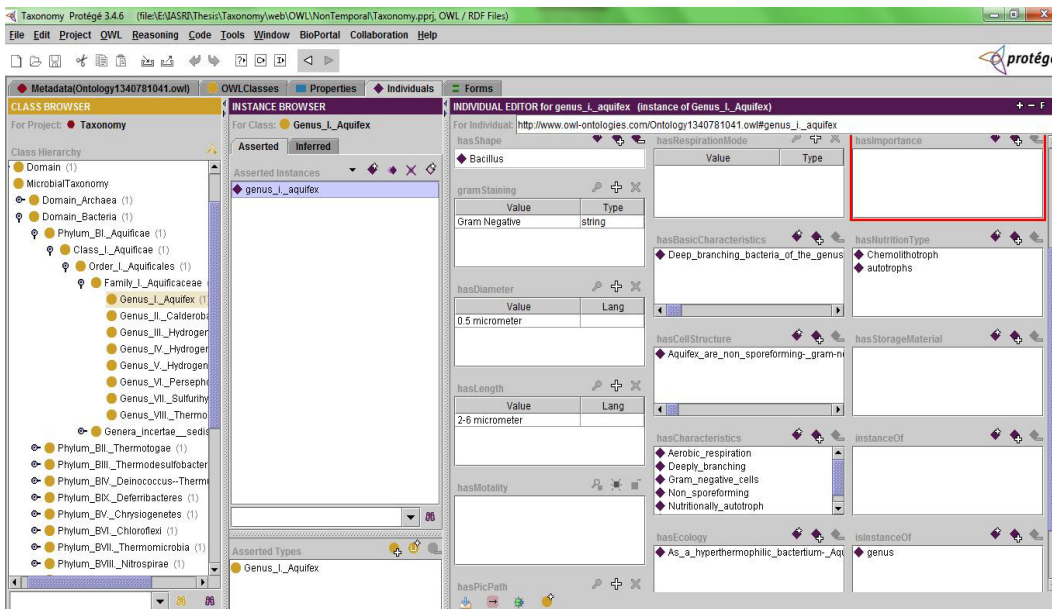Fig. 3. List of all properties of Microbial Ontology

Fig. 4. Genus_I._Aquife class with its individual and the properties

3.4 Creating Property Restriction

In addition to create properties in the ontology, it is possible to further constrain the range of properties in the context in a variety of ways. These are called Property restrictions. All restrictions are made under a tag owl:Restriction. The owl:onPropertyelement indicates the restricted property. It is easy to build Ontologies by creating restrictions in ontology editors like protégé. For example, the hasShape property is applied to the class MicrobialTaxonomy with the restriction created over its range that it has a "minCardinality 1". Similarly, all other restrictions for the classes are applied.

3.5. Querying the Microbial Ontology

The ontology is enriched with the taxonomy provided in the Bergey's Manual of Systematic Bacteriology. Now the main concern is to retrieve information from the ontology. Protégé provides a query interface known as SPARQL query panel where one can write queries to find out particular information from the Ontology.

Table 1. Some SPARQL queries used for querying the microbial ontology

S. No.	Query	SPARQL Syntax	Result
1	To find the shape of Form Genus II. Chroococcus	SELECT ?genus WHERE {:form_genus_ii._chroococcus :hasShape ?genus}	Bacillus
2	To find the characteristics of Form Genus II. Chroococcus	SELECT ?char WHERE {:form_genus_ii._chroococcus :hasCharacteristics ?char}	Photosynthetically helpful Rod shaped Aerobic respiration Gram negative cells
3	To find the ecology of Form Genus II. Chroococcus	SELECT ?eco WHERE { :form_genus_ii._chroococcus :hasEcology ?eco}	Chroococcus is known to traditionally inhabit freshwater areas, but has also been identified in water sources of higher salinity. Chroococcus has also been found in plankton inhabiting water reserves. Chroococcus is often incorrectly identified, therefore outlining a true habitat pattern is

			difficult.
4	To select the picpath of Form Genus II. Chroococcus	SELECT ?picpath WHERE { :form_genus_ii._chroococcus :hasPicPath ?picpath}	Image\Chroococcus\chroococcus.JPG

4. Design and development of software

The aspect of knowledge exchange between software agents and systems has been taken into consideration while designing the software. The tools and techniques have been chosen in accordance of the needs.

4.1. Design of the Software

While developing the software, N-tier architecture has been used. The application development environment is NetBeans IDE 7.0.1 and Protégé 3.4.6 [8], the OWL Ontology editor for creating the ontology (Microbial Taxonomy). Web development technology that has been used here is Java Server Pages (JSP), programming language is JAVA and for querying, SPARQL [9] is used. Client interface is developed with Hyper Text Markup Language (HTML), Cascading Style Sheet (CSS) and JavaScript. Third tier of software consist of database which is in MS-SQL server 2008, this make software login secured. The database is for user authentication only. Other two layers are OWL Ontology layer [10] and Semantic Web Framework layer, JENA [11]. OWL layer contains information on microbial taxonomy in the form of Ontology, which is human as well as machine readable and Semantic Web Framework layer is implemented by using JENA, a java framework for building Semantic Web applications. The system permits two types of user types, one is Normal User and other is Domain Experts (those having detailed knowledge on microbial taxonomy) during sign up to the system. Domain Experts are the users who can edit, view or add anything new to the knowledge base with prior permission of administrator. After sign in all options except "Edit Ontology" are available to users those who signed up as a normal user. "Edit Ontology" option is available to users those who signed up as a Domain Expert.

4.2. Classification of Microorganism

The users of the system will be able to study the microbial taxonomy up to the genus level for the microbes that are essentially important in agriculture under the domain bacteria. This feature is available for every type of users.By using the Advance Search option, user can classify an unknown microbe up to Genus level. This module is menu driven means the user can choose the option up to what level he/she wants the classification. User have to answer some of the questions (listed below) or select some characteristics from the drop down menu for classifying the microbe, probable results will be less with increase in number of answers given. After selecting any probable result it will give the taxonomy and detailed information of that microbe, so that user can easily compare the characteristic of the microbe he/she found with the characteristic of probable microbe generated by system. After comparison user can ensure whether the unknown microbe exactly matches with the system generated microbe or not, if not then user should move to other probable results for proper classification. The following are the questions to be asked.
1) What is the functional attribute? Or what is the importance from agricultural point of view?
2) What is the shape?
3) What are other characteristics?

The third question includes options for gram staining, cell division method, branching pattern etc.

```
<OC_Acetobacteriardf:ID=""/>
<BC_Thermodesulfobacteria rdf:ID="_0.3_to_0.6_micrometer_in_diameter_and_0.9_to_2.0_micrometer_in_length."/>
<BC_Aquifexrdf:ID="_1._Deep-branching_bacteria."/>
<OC_Aquifexrdf:ID="_16-S_RNA"/>
<BC_Aquifexrdf:ID="_2._Grows_best_in_water_between_85_to_95_degree_C"/>
<BC_Pseudomonadales rdf:ID="A_few_members_of_this_order_are_opportunistic_pathogens._Cells_are_usually_motile._
These_are_gram_negative_staining_bacteria-_usually_motile_by_polar_flagella._Some_contain_photosynthetic_pigments._Rigid-
walled_cells_of_variable_shape._In_some_species_forming_chains."/>
<BC_Entrobacteriales rdf:ID="A_large_order_of_over_100_species_of_gram_negative_and_facultatively_aerobic_bacilli
._These_are_common_commensals_of_humans_and_often_part_of_normal_gut_flora._These_are_a_major_cause_of_nosocomial_infections."
/>
```

Listing 1: Code snippet for Microbial Ontology

4.3. Editing the Knowledgebase by Domain Experts

The microbial taxonomy ontology knowledge base can be edited by domain experts if there is any wrong entry by the system developer, or any new information is available regarding a particular topic later. This can be done by using Edit Ontology option. By using this, an interface will guide the domain expert for the editing purpose. This interface asks for phylum, the domain expert want to edit. Then it provides the facility for editing the subclasses of the selected phylum along with the characteristics. There is also a panel for any suggestion by the domain expert.

4.4. Administrator Facilities

The domain expert has the ability to edit, view and add new information for the knowledge base. Because of this, the administrator must have the power to block any domain expert if he/she feels that the domain expert is a fake person. Administrator can also see all user information given at the time of sign up by the user. By seeing the information, administrator can enquire whether the person is really exists or not. If any domain expert ID is found fake then it remain blocked forever by the administrator.

5. Conclusion

Microorganisms represent by far the richest repertoire of molecular and metabolic diversity in nature. They provide the fundamental underpinning of all ecosystems; underlie basic ecosystem processes such as the biogeochemical cycles and food chains by converting an inert gas such as nitrogen to ammonia through biological nitrogen fixation (BNF), solubilize, mobilize and make available several minerals for use by crop plants, besides representing the progenitors of life on earth. Exposure to bacteria, parasites and viruses has also played an important role in human evolution. Currently, they are increasingly being explored as sources of novel genes and metabolites responsible for abiotic and biotic stress tolerance and development of transgenic crops. Ontologies define domain concepts and the relationships between them, and thus provide a domain language that is meaningful to both humans and machines. In contrast to traditional KOSs, ontologies provide conceptual abstraction and differentiated relationships. In ontologies, the semantics are developed through ensuring that each concept within the domain is uniquely and precisely defined and by specifying elaborated relationships among the concepts. The relationships in ontology are explicitly named and developed with specification of rules and constraints so that they reflect the context of domain for which the knowledge is modelled. Ontologies allow further knowledge to be inferred from the knowledge explicitly represented in ontology. The new knowledge could be derived by applying generalization or transitivity rules, the level of applicability of which is limited in a poorly defined KOS like a traditional taxonomy. Since, ontology creation from scratch is a complex process and involves huge cost, as iterations are required to relate each and every object with each other, migration from taxonomy is relatively much easier and cost efficient proposition. The web of today, the vast unstructured mass of information, may in the future be transformed into something more manageable - and thus something far more useful. The effective use of metadata among applications requires common conventions about semantics, syntax, and structure. Web Ontology Language (OWL) is designed to be used by applications that need to process the content of information instead of just presenting information to humans. Ontology is the latest way of knowledge representation, in any domain as it defines concepts and relationships between them, and provides a domain language that is meaningful to both humans and machines.

References

[1] Gene Ontology Consortium. The Gene Ontology (GO) Database and Informatics Resource. Nucl. Acids Res. 2004; 32: p. 258-261.

[2] Avraham S, Tung CW, Ilic K, Jaiswal P., Kellogg EA, McCouch S, Pujar A, Reiser L, Rhee SY, Sachs MM, Schaeffer M, Stein L, Stevens P, Vincent L, Zapata F, Ware D. The Plant Ontology Database: a community resource for plant structure and development stages controlled vocabulary and annotations. Nucl. Acids Res. 2008; 36: p. 449-454.

[3] Caterina C, Marta IS and Johannes K. Towards Interoperability of Geopolitical Information within FAO. Comp. and Informatics. 2006; p.119-129.

[4] Bedi P,Marwaha S. Designing Ontologies from Traditional Taxonomies, In the Proceedings of International Conference on Cognitive Science 2004: p. 324-329.

[5] Das M, Malhotra PK, Marwaha S, Pandey RN. Building and Querying Soil Ontology. J. of ISAS. 2012;66:3 p.459-464.

[6] Ming Z, Qingling Z, Dong T, Ping Q, Xiaoshuan Z. Ontology-based intelligent retrieval system for soil knowledge. WSEAS Trans. Info. Sci. and App. 2009;6-7. http://portal.acm.org/citation.cfm?id=1639432.

[7] Bergey DH, Harrison FC, Breed RS, Hammer BW, Huntoon FM. Bergey's Manual of Systematic Bacteriology. Baltimore: Williams & Wilkins. 1989.

[8] Gennari JH, Musen MA, Fergerson RW, Grosso WE, Crubezy M, Eriksson H, Noy NF, Tu SW. The Evolution of Protege: An Environment for Knowledge-Based Systems Development, International Journal of Human Computer Studies. 2003;58:1-89.

[9] Clark K. SPARQL Protocol for RDF, W3C Recommendation, http://www.w3.org/TR/rdf-sparql-protocol/. 2008

[10] Smith MK, Welty C, McGuinness D. OWL Web Ontology Language Guide, W3C Note, http://www.w3.org/TR/owl-guide/. 2004.

[11] Carroll JJ, Dickinson I,Dollin C, Reynolds D, Seaborne A, Wilkinson K. Jena: Implementing the Semantic Web Recommendations, In Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters, New York, NY, USA. 2004.