Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 66(3) 2012 459-464

Building and Querying Soil Ontology

Manoranjan Das[1], P.K. Malhotra[1*], Sudeep Marwaha[1] and R.N. Pandey[2]
[1]*Indian Agricultural Statistics Research Institute, New Delhi*
[2]*Indian Agricultural Research Institute, New Delhi*

Received 17 January 2012; Revised 21 June 2012; Accepted 11 July 2012

SUMMARY

Soil Taxonomy is based on soil properties that can be objectively observed and measured. There are many soil classification systems but USDA Soil Taxonomy is most accepted worldwide. Ontologies are the new form of knowledge representation that act 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 human beings and computing 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 modeled. Ontologies can be built by using various GUI based software tools, known as Ontology editors. Among all editors Protégé [Gennari *et al.* (2003); Golbeck *et al.* (2003)] 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 [Clark (2008)]. Soil ontology developed for USDA soil taxonomy will be helpful for study of soil taxonomy and classification of new soils. Soil Ontology is built in the Protégé OWL editor from Order to Sub group level. Using this soil ontology, a query interface can be developed that will help in detailed study of soil taxonomy, classification of new soil as well as exchange knowledge between software agents and systems.

Keywords: RDF, OWL, Protégé, SPARQL, Ontology.

1. INTRODUCTION

There are many types of soils that are formed by the interaction of different soil-forming factors and processes, possessing different properties and characteristics. We need to classify the soil for organising and establishing a common understanding about each and every soil worldwide. Soil Taxonomy is based on soil properties that can be objectively observed and measured. Soil Taxonomy makes use of nomenclature which gives definite connotation of the major characteristics of soils. There are many soil classification systems but United States Department of Agriculture (USDA) Soil Taxonomy [USDA, NRCS (2010)] is most accepted worldwide. Ontologies are the new form of knowledge representation that act in synergy with agents and Semantic Web Architecture [Berners-Lee *et al.* (2001)]. 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 modeled. Ontologies can be built by using various GUI based software tools, known as Ontology editors. Among all editors Protégé [Gennari *et al.* (2003); Golbeck *et al.* (2003)] is widely supported by a huge research community.

There are many web based softwares which use ontologies as their knowledge base like Gene Ontology (GO) and Plant Ontology (PO). Gene Ontology is

*Corresponding author : P.K. Malhotra
E-mail address : pkm@iasri.res.in

developed by Gene Ontology Consortium (2000). They developed a tool named as AmiGO for searching and browsing the Gene Ontology database. Plant Ontology is developed by Plant Ontology Consortium (2002). Plant Ontology Consortium was formed in response to the need for a set of uniform terms to describe plant structures and developmental stages. Bedi and Marwaha (2004) proposed a methodology for the conversion of taxonomies to ontologies. The proposed methodology is tested and implemented for a pilot soil ontology using the IEEE standard Web Ontology Language (OWL) and Protégé 2.1 OWL plug-in. OWL is the W3C recommendation for describing Ontology [Dean *et al.* (2003)]. Ontology-based intelligent retrieval system for soil knowledge [Ming *et al*. (2009)] 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".

Presently the USDA Soil Taxonomy is available in text form. There is hardly any web based software for detailed study of the Soil Taxonomy and for classification of newly found soil according to USDA Soil Taxonomic classification rule. So the students, researchers and experimenters face many problems during their study and to classify a newly found soil. In this paper, efforts for devising an approach to create ontology from soil taxonomy are presented. The knowledge base of this software is an ontology (Soil Ontology) which is built by using Protégé editor [Gennari *et al.* (2003); Golbeck *et al.* (2003)].

2. DESIGNING ONTOLOGY FROM TAXONOMY

Ontology creation from the scratch is a complex process and is expensive, as many iterations are required to relate each and every concept with one another. Migration from taxonomy is relatively much easier and cost efficient proposition as taxonomies are already standardized. 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, due to the standardization of Ontology Web Language [Smith *et al.* (2004)] and tools such as Protégé 3.3.1, building ontologies have become significantly easier. 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, for the knowledge engineers, the problem of getting knowledge from the domain expert remains unsolved. In this paper, 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 Soil Ontology from Soil Taxonomy. Soil 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. List of steps for designing the ontology from domain taxonomy are as follows:

1. Study the taxonomy and list the major entities in which the concepts are classified. Put them in the hierarchical format.
2. These entities will be the classes in the ontology in such a fashion that the first classification term will be the top class under *owl:thing* and the second term will be its first sub class and next heading will be the next sub class of the first sub class and so on.
3. Create top level classes of the concepts that appear under the top level classification term as a separate hierarchy under *owl:thing* class. Similarly, create other classes in a hierarchy corresponding to the concepts as they appear in the taxonomy hierarchy.
4. List different properties that are applicable to the instances or individuals of the classes defined in step 2. These properties are assigned appropriate values at the time of creation of individuals. By default, these properties are inherited to the individuals of the classes that are derived from them.
5. Decide the range and data type of values for each property of the class. In OWL, one can have two types of properties Data type Property and Object

property. The Data type property can have data types defined by XML Schema such as Boolean, Float, Integer, String, and Symbol values while Object property can point to other class or individual of the class.

6. Also, establish the *rdfs:subClassOf* relationship as a necessary condition between the concept classes defined in step 3 and corresponding classification term classes defined in step 2.
7. Repeat Step 4 and Step 5 for all the concept classes.
8. Create individuals for all the sub classes of the ontology and assign values to the different properties for the individuals of different classes of the ontology.
9. Refine and enhance the ontology by adding appropriate restrictions and other annotation properties that are out of the scope of taxonomy.

2.1 Building Soil Ontology from Soil Taxonomy in Protégé

By using the above approach, the Soil Ontology in Protégé with OWL plug-in is implemented. However, the approach is generic and can be implemented with any tool that provides support to implement ontology in OWL. First step is to identify the hierarchy in which the soil taxonomy is classified. In Soil Ontology the hierarchy is:

Order (*e.g.* Alfisols) → Sub order (*e.g.* Aqualfs) → Great group (*e.g.* Albaqualfs) → Sub group (*e.g.* Typic Albaqualfs).

2.1.1 Creating classes

- Being Alfisols is at the top of the hierarchy, the Alfisols class is created as the base class from which other classes are derived. Aqualfs class is the first sub class, Albaqualfs class is the next sub class and finally Typic Albaqualfs is last class of the hierarchy.
- Class Alfisols is created as a top level class under *owl:thing* class. Also, the *rdfs:subClassOf* relationship is established as a necessary condition between the concept classes and corresponding classification term classes defined in previous step. For example: class Albaqualfs is a subclass

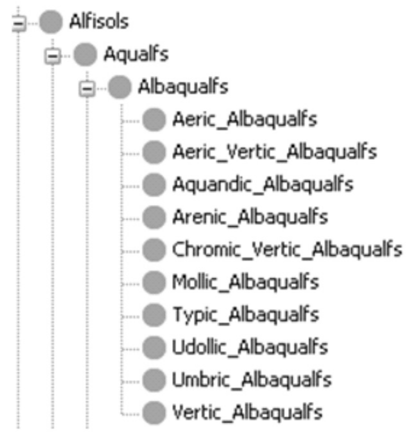

Fig. 1. Class hierarchy of soil ontology in Protégé

of Aqualfs and class Aqualfs is a subclass of Alfisols [Fig. 1].

- After creating all classes present in the Soil Taxonomy hierarchy, classes like *Basic_Property*, *Horizons_and_Characteristics_Diagnostic*, and *Other_Property* were created to keep the properties of the classes created before [Fig. 2].
- Basic_Property has several sub classes for basic properties of all orders present in Soil Taxonomy. Similarly other two classes created in previous step have their own sub classes.

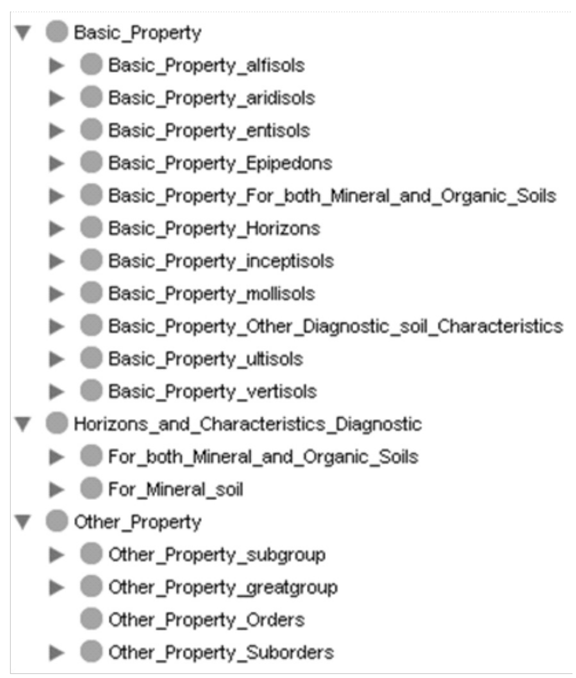

Fig. 2. Property classes and their subclasses in soil ontology

2.1.2 Creating individuals

After creating all classes it's time to create their individuals. Individuals of all classes were created in the individual editor of Protégé OWL plug-in. Individuals of all classes of Soil Taxonomy hierarchy were created with same name of their classes *e.g. Alfisols* class has individual *alfisols* and *Aqualfs* has individual *aqualfs*, similarly other classes of Soil Taxonomy have their individuals, with all letters in small alphabets [Fig. 3].

Fig. 3. alfisols as individual of Alfisols class in soil ontology

2.1.3 Creating properties

After creating classes and individuals it's time to create properties for classes and their individuals in Soil Ontology. Properties were created in Property editor of Protégé OWL plug-in, where *domain* and *range* for each property were declared *e.g. hasBasicProperty* is a object property, it's domains are *Order* and *Horizons_and_Characteristics_Diagnostic* classes and range is *Basic_Property*. Similarly other properties were created and their domain and range were also specified [Fig. 4].

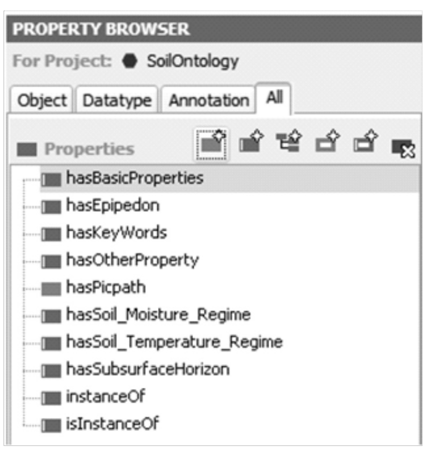

Fig. 4. List of all properties in soil ontology

2.1.4 Creating property restrictions

In addition to designating property characteristics, it is possible to further constrain the range of a property in specific contexts in a variety of ways, called *property restrictions*. All restrictions are made under a tag *owl:Restriction*. The *owl:onProperty* element indicates the restricted property. It is easy to build Ontologies by creating restrictions in ontology editors like Protégé, *e.g. hasBasicProperty* property is applied to class *Alfisols* as domain and a restriction is created for its range and the restriction is "*Alfisols* should contain some of individuals of class *Basic_Property_Alfisols* as range for the property *hasBasicProperty*" [Fig. 5].

Fig. 5. Restriction applied to *hasBasicProperty* property

Similarly other restrictions for *Alfisols* class are done. As orders are categorized on the basis of presence or absence of major diagnostic horizons (surface & subsurface horizons) restriction can be done for *hasEpipedon* and *hasSubsurfaceHorizon* properties. The restriction for these properties when applied to class "*Alfisols* is *minimum cardinality must be one*" [Fig. 6 and Fig. 7].

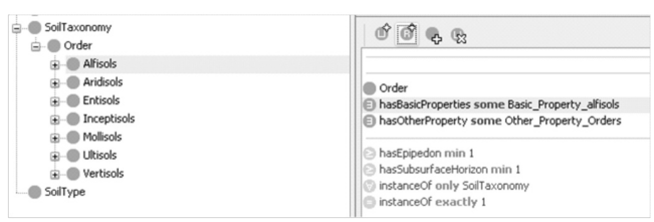

Fig. 6. Alfisols class with its restrictions

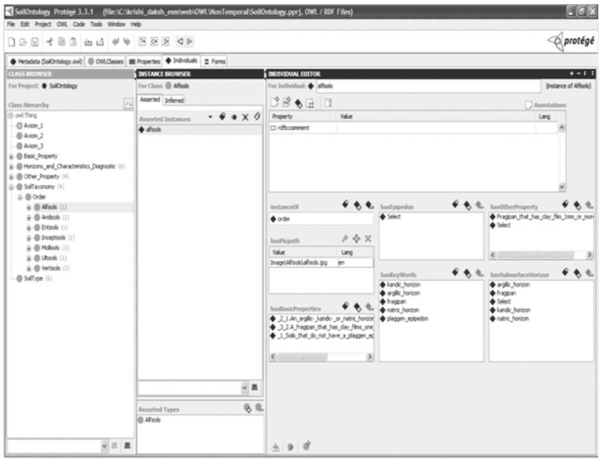

Fig. 7. Alfisols class with its individual alfisols with its properties and their corresponding values in Protégé OWL Plug-in

3. QUERYING SOIL ONTOLOGY

Since, the USDA Soil Taxonomy is in the form of OWL Ontology, the knowledge can be retrieved from the Soil Ontology. Currently soil ontology for Indian soils (7 orders) is available. Protégé [Gennari *et al.* (2003); Golbeck *et al.* (2003)] provides a query interface known as SPARQL query panel where one can write queries to find out particular information from the Ontology. SPARQL is a syntactically-SQL-like language for querying RDF graphs [Clark (2008)].

3.1 Simple Queries and Results

- Find out all orders in the Soil Ontology?

  SPARQL syntax:

  SELECT ? order

  WHERE {?order rdfs:subClassOf :Order}

  ORDER BY (?order)

SELECT clause identifies the variables to appear in the query results, and the WHERE clause provides the basic graph pattern to match against the data graph. ORDER BY clause gives the result in alphabetical order [Fig. 8].

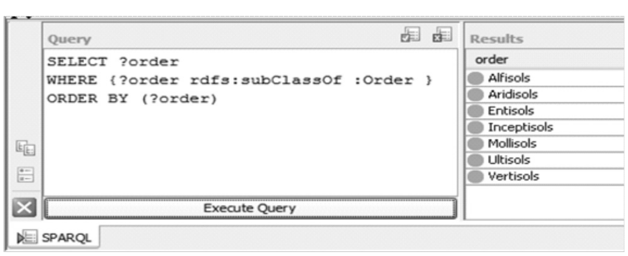

Fig. 8. SPARQL query interface with query to find out all orders in soil ontology

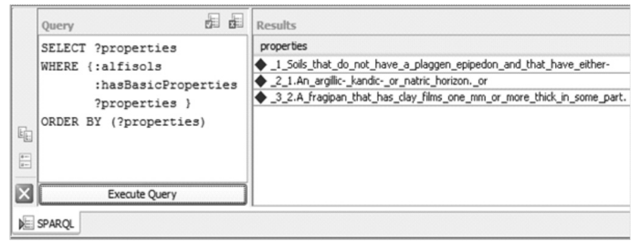

Fig. 9. SPARQL query interface with query to find out properties of Alfisols

- Find out the properties of order alfisols in soil ontology? [Fig. 9]

  SPARQL syntax:

  SELECT ?properties

  WHERE {:alfisols :hasBasicProperties ?properties }

  ORDER BY (?properties)

3.2 Other Queries

Other queries and their results are given in Table 1.

Table 1. Other SPARQL queries and their results

S. No.	Query	SPARQL Syntax	Result
1.	Find the keywords in order Alfisols?	SELECT ?Keywords WHERE {:alfisols :has KeyWords ?Keywords} ORDER BY (?Keywords}	Argillic horizon, fragipan, kandic horizon, etc.
2.	Find the path of image for Alfisols?	SELECT ?Picpath WHERE {:alfisols :has Picpath ?Picpath}	Image/Alfis ols/alfisols.jpg
3.	Find the order(s) containing Argillic and Kandic horizons?	SELECT ?Order WHERE {?Order :instance Of :order. ?Order :has Subsurface :has Subsurface ?Order :has Subsurface Horizon :kandic } ORDER BY (?Order)	Alfisols Mollisols Ultisols
4.	Find the order(s) containing Argillic, Kandic horizons and Mollic epipedon?	SELECT ?Order WHERE {?Order :instanceOf :order. ?Order :has Subsurface Horizon :argillic. ?Order :has Subsurface Horizon :kandic. ?Order : hasEpipedon :mollic } ORDER BY (?Order)	Mollisols

4. CONCLUSION

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) [Smith *et al.* (2004)] 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 human beings and computing machines. Designing and creation of ontology is not very tedious as there are many software tools (like Protégé) with OWL plug-in are available freely on the web. Using SPARQL [Clark (2008)] one can easily retrieve the knowledge from ontologies. Building ontologies in different domains of agriculture will be help to convert unstructured knowledge into structured one that can be shared across applications. This has been illustrated by developing soil ontology keeping in view USDA soil taxonomy through Protégé OWL editor from Order to Sub group level for Indian soils. Using this soil ontology, a query interface can be developed that will help detailed study of soil taxonomy, classification of new soil as well as exchange knowledge between software agents and systems.

REFERENCES

Bedi, P. and Marwaha, S. (2004). Designing Ontologies from Traditional Taxonomies. In: *Proceedings of International Conference on Cognitive Science*, Allahabad, India.

Berners-Lee, T., Hendler, J. and Lassila, O. (2001). The Semantic Web. *Scientific American*, **284(5),** 34-43.

Clark, K. (2008). *SPARQL Protocol for RDF*, *W3C Recommendation*. http://www.w3.org/TR/rdf-sparql-protocol/.

Dean, M., Connolly, D., Harmelen, F.V., Hendler, J., Horrocks, I., McGuinness, D.L., Patel-Schneider, P.F. and Stein, L.A. (2003). *OWL Web Ontology Language Reference, W3C Working Draft.* http://www.w3.org/TR/2003/WD-owl-ref-20030331.

Gene Ontology Consortium. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, **25,** 25-29.

Gennari, J.H., Musen, M.A., Fergerson, R.W., Grosso, W.E., Crubezy, M., Eriksson, H., Noy, N.F. and Tu, S.W. (2003). The evolution of Protege: An environment for knowledge-based systems development. *Intern. J. Human Computer Studies*, **58(1),** 89-123.

Golbeck, J., Fragoso, G., Hartel, F., Hendler, J., Parsia, B. and Oberthaler, J. (2003). The National Cancer Institute's thesaurus and ontology. *J. Web Semantics*, **1(1),** 7-26.

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

Plant Ontology. (2002). Plant Ontology Consortium. http://www.plantontology.org/.

Smith, M.K., Welty, C. and McGuinness, D.L. (2004). *OWL Web Ontology Language Guide (W3C Recommendation)*. http://www.w3.org/TR/owl-guide/.

USDA, NRCS. (2010). *Keys to Soil Taxonomy*, Eleventh Edition. ftp://ftp-fc.sc.egov.usda.gov/NSSC/Soil_Taxonomy/keys/keys.pdf.