Ontology Creation and Semantic Web for Paddy

Marvin P Polson
Dr MGR Educational and Research Institute
Chennai
India

Sumathy Eswaran
Dr MGR Educational and Research Institute
Chennai
India

ABSTRACT

The Semantic Web which is a collaborative movement led by the World Wide Web Consortium (W3C) is gaining popularity among the web users. Ontology forms the backbone of semantic web in representing data. A Paddy ontology created for the agro repository is presented in this paper. Based on Resource Description Framework/Web Ontology Language specifications the created ontology supports the extraction of new knowledge in the areas of high value Paddy, different varieties , possible diseases etc in the southern part of india.This paper makes an attempt to tag the relevant information on Paddy ontology formation which enable user to find, share and combine information more easily.

Keywords

RDFS, RDF, OWL, paddy, Semantic web, Knowledge portal

1. INTRODUCTION

The main idea of the semantic web is to be a web talking to machines [1], i.e. in which machines can provide a better help to people because they can take advantage of the content of the Web. The information on the web should thus be expressed in a meaningful way accessible to computers so that knowledge representation will be more effective. Presently no one cares how the semantic web will work but for the end user, care more about how effectively they can mine knowledge from the semantic web. For the development of a semantically rich semantic web, the first step to be taken is the creation of different ontology for different domains. A domain can be an area of knowledge, like medicine or a specific subject area. Ontology and the Semantic Web (SW) are intrinsically related. Ontology is the backbone of the SW. They are meant to provide the connectivity tissue that will facilitate the realization of an interoperable and intelligent Web. The benefits of ontology appear in the successful integration and interoperability of the whole rather than any specific part. Recognition of this may lead ontologists to keep the use of ontology under the hood, rather than something that is argued for up front. However, it should equally be possible to explain how ontology contributes to the overall success, even if the overall success cannot be claimed for the ontology alone. It is also important to understand how ontologies could be used, what added functionality could be enabled, what frameworks (e.g. Integrated Development Environments) are there that could be used. Moreover, the explicit benefit in using ontologies is not always clear, especially the benefit in using those aspects of ontologies that go beyond controlled vocabularies. The definitions of ontologies are machine readable and they describe basic concepts in the domain and the relations between them. The knowledge, which is encoded in ontologies, is reusable due to the fact that the encoded knowledge can span different domains. More generally, knowledge management, personal or corporate, can take advantage of the semantic web. The semantic web will provide value to any semantically annotated resource by facilitating its retrieving when appropriate. One of the challenges of the current semantic web developments is the design of a framework in which all these understanding can collaborate, because the full benefit of the semantic web can only be attained when computers relate resources from various sources. Ontologies are able to specify the following kinds of concepts, which enable the description of any knowledge [2]:

Classes (things)

Relationships between things

Properties (attributes) of things

The main aim in using ontologies [3] is

To share common understanding of the structure of information among people or software agents
To enable reuse of domain knowledge
To make domain assumptions explicit
To separate domain knowledge from the operational knowledge
To analyze domain knowledge

2. ONTOLOGY CREATION

To create ontology, the W3C has defined a number of description languages. The first published language was the Resource Description Framework (RDF) .RDF uses Extensible Markup Language (XML) as syntax model. A more abstract model is the RDF-Graph. Both syntaxes describe RDF-triples or RDF-statement, and every triple consist of a resource (subject), a property (predicate) and a property value (object). These three elements are the major essentials of a RDF-statement, and also a relationship between things. A RDF-statement is an explanation of a knowledge fact.

2.1 Resources

A resource is a thing, which is described in the RDF-statement. Hence, a resource can be anything, like a website, a real world object (e.g. cat) etc. To identify an object exist the Uniform Resource Identifier (URI), which is a global unique UNICODE based string that bases on a W3C specification.

2.2 Property value

The property value can be a literal or a resource, which is linked over a property to a subject.

The Resource Description Framework (RDF) [6] enables the creation and exchange of resource metadata as normal Web data. To interpret these metadata within or across user communities, RDF allows the definition of appropriate schema vocabularies (RDFS) [6]. However, the fact that several communities, even with similar needs, have developed their own metadata vocabularies independently indicates the need for schema repositories facilitating knowledge sharing. In this way, already defined concepts or properties for a domain can be either reused as such or simply refined to meet the resource description needs of a particular user community, while preserving a well-defined semantic interoperability infrastructure. RDFS extends the RDF-vocabulary so that a user is able to describe a full class structure. Furthermore, RDFS was a solution to the weaknesses in definition of RDF. It does not use the same data-model as RDF, with graph and triples, it is also written in RDF. The new vocabulary of RDFS supplements the RDF-vocabulary with syntax for classes, subclasses, inheritance and more, necessary to develop full hierarchies. The next step of ontology description languages is called Web Ontology Language (OWL). Once again OWL is a step in extending the expressive power and performance to create ontology.

3. PADDY TAXONOMY AND PROPERTY DEFINITION

Rice is the world’s most common staple food. For more than half of mankind, in 118 countries, Rice is the main component of their diet. The world production of Rice is 605 million tons of paddy per year, equal to 403 tons of milled Rice. Unfortunately, the research knowledge repositories for Paddy production and plant production are not well organized and utilized.

3.1 Criteria Construction Process

The concepts of the Paddy production ontology was categorized as classes to provide an initial comprehensive framework that will incorporate every other relevant concept. The Paddy production ontology construction criteria for defining concept categories followed the plant production knowledge model applied from Beverly whole plant model [4]. The process of creating criteria was done according to three stages:

Stage 1: Defining the criteria. Ontology was a data model that provides an organizational framework that allows reasoning about knowledge. The criteria for constructing ontology should be defined as:

a. Criteria for defining concept

b. Criteria for defining term, and

c. Criteria for defining relationship

Stage 2: Formulating preliminary set of criteria and applying to the working process of Paddy production ontology construction. Those criteria, as a result, was modified and adjusted according to the ontology construction process.

Stage 3: Testing and evaluating the criteria. The criteria were presented to domain specific experts and information specialists who were knowledgeable about Paddy production, albeit unskilled about ontologies.Every classes should be properly related to each other with and inverse relationships.Reasoner won’t be able to infer the created ontology in case of ambiguity and isolated classes. After testing and evaluation the created ontology will be ready to parse and display in ontology browser.

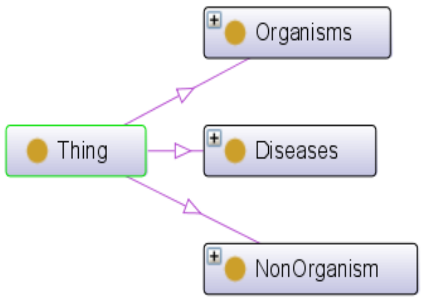

Figure 1 Paddy Ontology

3.2 Criteria for Defining Concept

Concept for defining paddy production ontology should be defined into three concept categories as : conceptual entity concept, object entity concept and functional entity concept. Object entity concepts were divided to organism and non-organism object entity concept as in Figure 1. Functional entity concepts were classified as: plant production process, breeding method, protection process, infecting process and physiological functional entity concepts. The plant production process entity concept was divided into cultivation process, harvesting process, soil preparation process, fertilizing process, irrigation process, propagation process and seed processing functional entity concepts.

3.3 Criteria for Defining Term

All terms which represent concepts should be identified with either one of the following groups.

a) A preferred term is term which is preferably lexicalized the concept. The preferred term was the main term representative of a concept when that concept could be described by various different terms. There was only one term designated as the preferred term as in Figure 2 and 3. The other terms were considered as non-preferred terms or synonyms. Acronyms, abbreviation names, or symbols were not used as preferred terms [2]. Preferred terms were selected from related thesaurus, dictionaries or terms that were accepted or recommended by experts in that domain.

b) Non-preferred terms or synonyms were terms with the same meaning as the preferred term but were not selected as main concept representatives. Some of the non-preferred terms were also called synonyms. Non-preferred terms or synonyms could be in the form of term variants such as spelling variants, acronym, abbreviation names, terms in singular or plural, common names, local names, scientific names, trade names, chemical symbols, chemical formulas, etc. There should be some specific standard to be followed in defining terms. Common terms must be singular and can be noun, word or noun phrase. Terms should be non-capitalized expect for some specific terms such as scientific names, trade names, geographical names etc.For Organisms Scientific names representing plants, animals and related organisms were used in Latin. Common names in English were defined as synonyms for more information. For naming non-organisms the most accepted name entity in the subject domain were used as preferred term and the less defined as synonym (non preferred term). Soil Series names defined by the Land Development Department were used to represent soil. Agricultural chemical substances use Substance common names in English for representing and substance trade names as synonyms. In defining Plant nutrients, element names were used represent plant nutrients and Chemical symbols and chemical formulas were defined as synonyms further to increase clarity. When you find the necessity to define more than one meaning for one term, then you are facing the terminological problem. Each term should correspond to exactly one concept in ontology, since you are not building a dictionary, but a well-organized conceptual structure. Each term is only a label of the concept. You of course can build a dictionary after building ontology. It is usually a good idea to avoid abbreviations in concept names (that is, use Plant Production rather than PP).

.

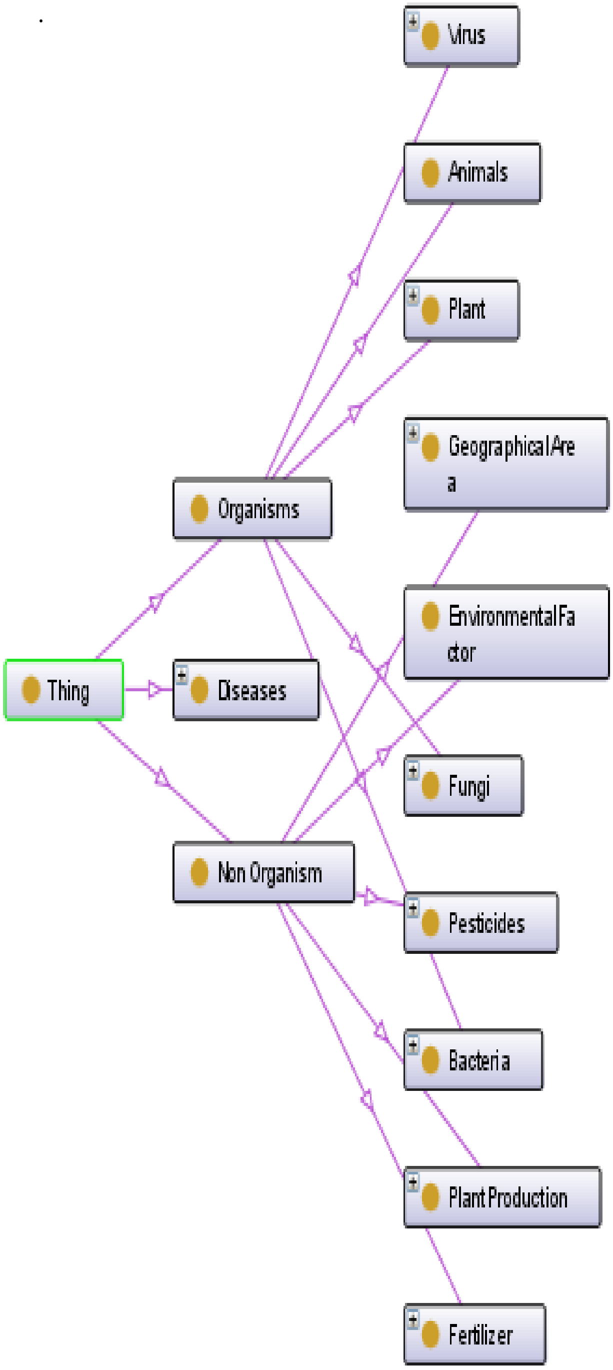

Figure 2. Non Organism Classification

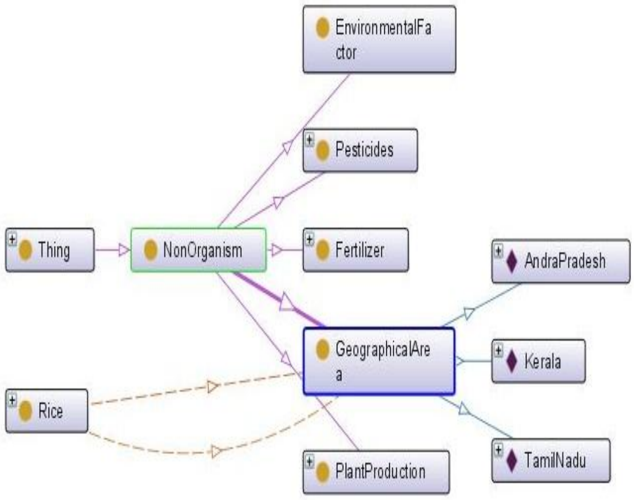

Figure 3. Paddy Ontology with individuals

3.4 Criteria for Defining Relationships

The relationships between classes or entities is of two types as direct relationship and inverse relationship as listed in table1.Based on the connection semantics two criteria are defined:

3.4.1 Criteria for defining hierarchical relationships:

Concepts in the same tree were linked by using hierarchical relation. It was a hierarchical linkage like the subclass and super class or mother and child concept relation. The hierarchical relationship had only one relation, which was the hasSubclassOf and had inverse relation isSubclassOf.

3.4.2 Criteria for defining associative relationships:

Interconnections between concepts in different hierarchies were created. They could be related in different ways and could be divided into functional and conceptual relationships.

Table 1. Collection of Relation and inverse

Relation	Inverse Relation
hasSynonym	isSynonymOf
hasLocalName	isLocalNameOf
hasFertilizingProcess	isFertilizingProcess
hasFertilizingMethod	isFertilizingMethod
hasHarvestingProcess	isHarvestingProcess

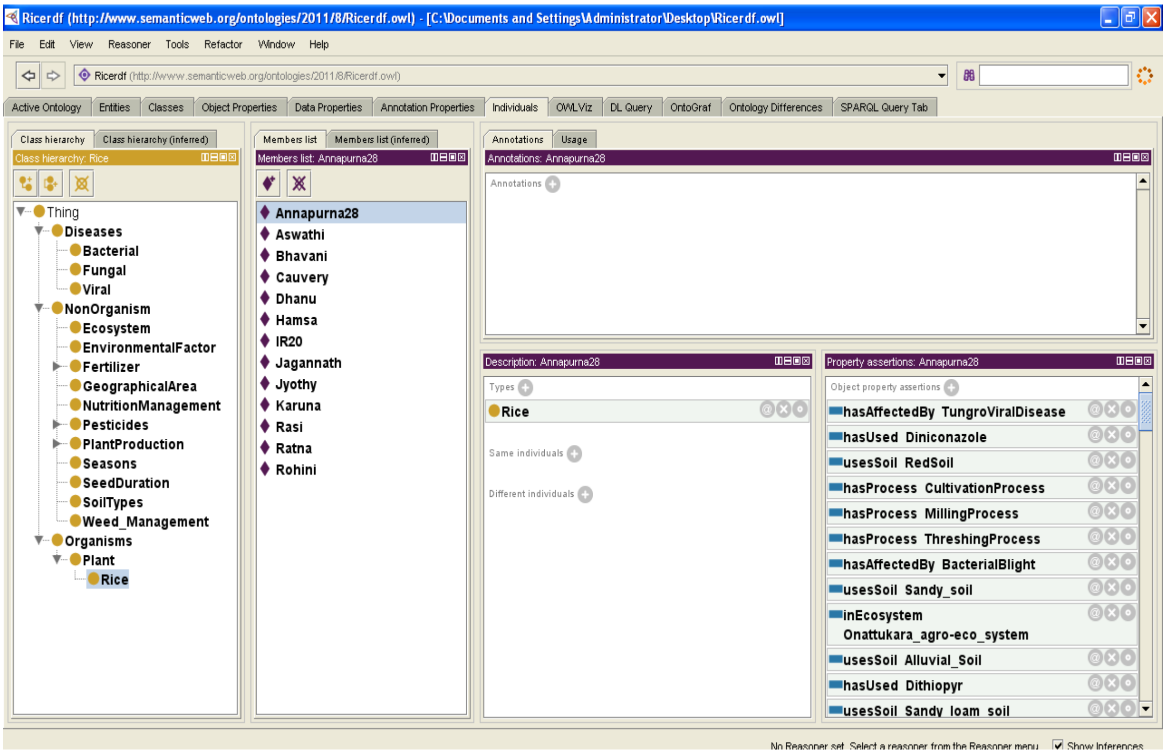

Figure 4. Paddy Ontology creation using protégé

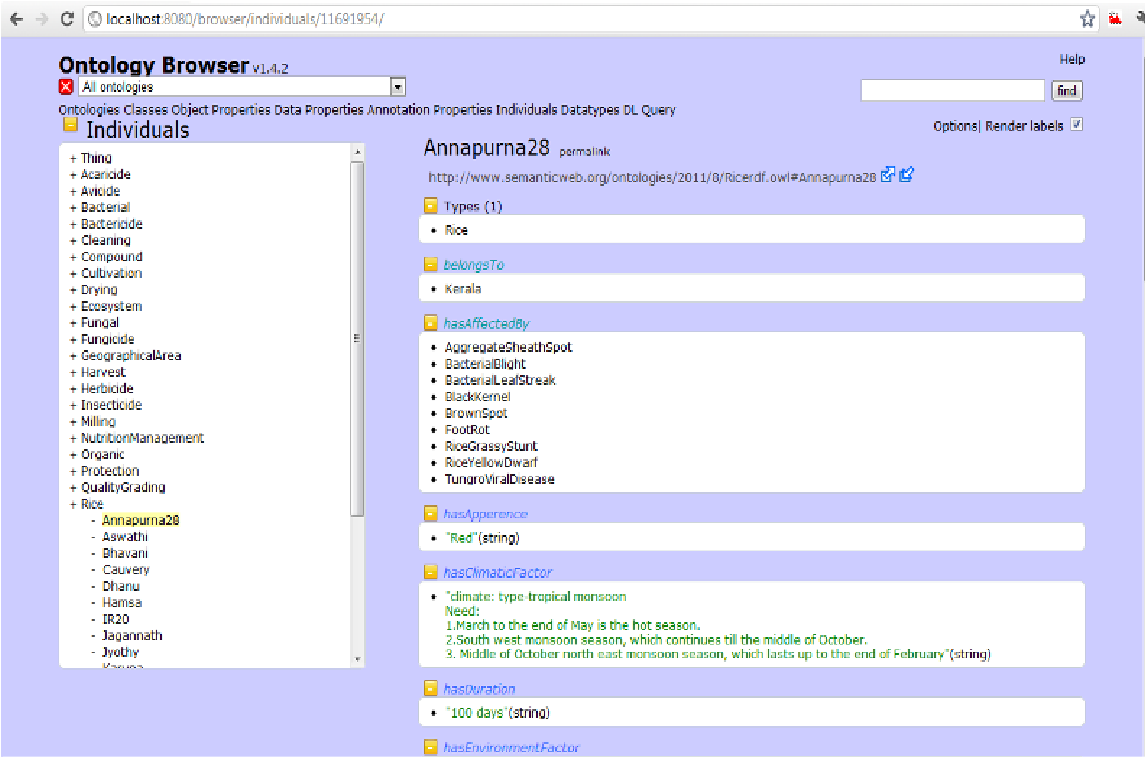

Figure 5. Paddy Ontology in Ontology Browser

Table 2 Collection of Entity Concept

Conceptual Entity Concept	Object Entity Concept	Functional Entity Concept
Taxonomic unit	Organism	Plant production process
..biological taxonomic unit	..plantae	..cultivation process
..soil taxonomic unit	..animalia	..harvesting process
….soil series	..fungi	..soil preparation process
Behavior	..bacteria	..fertilizing process
..animal behavior	..virus	..irrigation process
..plant habit	Non-organism	..propagation process
Composition	..environmental factor	..seed processing
..plant anatomy	….water	Breeding method
..chemical composition	….light	Protection process
Property	….weather	..control method
..biological property	….pollutant	Infecting process
..soil property	..geographical area	Physiological function
Type	..plant nutrient	..growth period
..organism type	..soil	
..non-organism type	..soil amendment	
Appearance	..agricultural substance	
..duration	….fertilizer	
..disorder	….pesticide	
..disease	….plant growth regulator	
..symptom	..product	

4. RESULT

The paddy ontology has given expressive power and syntactic interoperability. The Universal expressive power is high when compared to first and second generation web technologies[4].Since it is not possible to anticipate all potential uses, Ontology have enough expressive power to express any form of data. The ontology is created using protégé as in figure 4. Figure 5 represent the data in a ontology browser in a more expressive way. Another feature is Support for Syntactic Interoperability. By syntactic interoperability we mean how easy it is to read the data and get a representation that can be exploited by applications. . For example, software components like parsers or query APIs should be as reusable as possible among different applications. Syntactic interoperability is high when the parsers and APIs needed to manipulate the data are readily available.

5. CONCLUSION

The semantic web requires as much data as possible to be tagged with identifying markers that each link to other markers that they are related to. All of this information would be visible within one search result. . This all sounds great, but it requires all relevant information to be re-tagged with these identifiers, which is a daunting task for all sources of data. Our farm-agro ontology model based on Paddy gives the necessary guidelines for any researcher to create ontology of his area which will enhance the relevance of semantic web in the world. Future work proposed is to enhance the ontology to more agricultural products.

6. ACKNOWLEDGEMENT

Protege-2000 (http://protege.stanford.edu) was developed by Mark Musenas group at Stanford Medical Informatics. . We generated some of the figures with the OntoGraph plugin to Protege-2000. We thank Stanford University for enabling us to use their Ontology development tool Protégé and become member of Protégé user group. This work is approved and listed at The Protégé Group in Stanford University as Paddy ontology at http://protege.stanford.eu. The tools used in this model are Protégé, OntoGraf and HermiT reasoner. The Paddy ontology which is created based on the above model is available for public at http://ricestudy.servehttp.com/rice.

7. REFERENCES

[1] W3C Candidate Recommendation. March 27, 2000. .

[2] Beverly, R.B., J.G. Latimer and D.A. Smittle. 1993. Natalya F. Noy and Deborah L. McGuinness Stanford University, Stanford, CA, 94305Tavel, P. 2007 Modeling and Simulation Design. AK Peters Ltd.

[3] D. Brickley, R.V. Guha. Resource Description Framework Schema (RDF/S) Specification.

[4] Colin Evans Programming the Semantic Web, Toby Segaran and 2009 Jamie Taylor, OR EILLY Press.

[5] O. Lassila, R. Swick. Resource Description Framework (RDF) Model and Syntax Specification • . W3C Recommendation. February.

[6] Grigoris Antoniou A Semantic Web Primer, 2Ed and 2008 Frank van Harmelen, MIT Press.

[7] James Hendler, Semantic Web for the Working Ontologist Modeling in RDF RDFS, Dean Allemang OWL, and 2008 Morgan Kaufmann Press.

[8] BHAVANI THURAISINGHAM, XML Databases and CRC PRESS 2006 the semantic Web.

[9] Michael C. Daconta,Web Services The Semantic Web: A Guide to the Future of XML and Leo J. Obrst Kevin T. Smith Wiley Publishing Inc 2003 Knowledge Management.

[10] Jefrey T. Pollock,Wiley Publishing inc 2009 Semantic Web For Dummies.

[11] P. Domingos A. Doan and 2003. A. Y. Halevy. Learning to match the schemas of data sources: A multistrategy approach. Machine Learning, 50(3).

[12] Xiangfeng Luo Association Link Network: an In cremental Semantic Data Model on Organizing Web Resources, Zheng Xu and 2009 Wenjun Lu.

[13] Hsin-Chang Yang., Bridging the WWW to the Semantic Web by Automatic Semantic Tagging of Web Pages, 2005

[14] Jos A. Alonso-Jimnez, Foundational Challenges in Automated Seman tic Web Data and Joaqun 2006 Ontology Cleaning,

[15] Chang Yang. Bridging the WWW to the Semantic Web by Automatic Semantic Tagging of Web Pages' 2005