



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 5, October 2014

An Ontology-Based Information Retrieval Model for Domesticated Plants

Ruban S¹, Kedar Tendolkar², Austin Peter Rodrigues², Niriksha Shetty²

Assistant Professor, Department of IT, AIMIT, St Aloysius College, Beeri, Mangalore, India¹

M.Sc. [ST] III Semester, AIMIT, St Aloysius College, Beeri, Mangalore, India²

ABSTRACT: Search engines have become part and parcel of human life. Though search engines have experienced many changes from past few years. It still relies on keyword based techniques. With the rapid increment of the information on the web, traditional information retrieval based on the keywords is far from user's satisfaction in recall and precision. In order to improve the accuracy of retrieval in the domain of domesticated plants, an information retrieval model based on the domesticated plants ontology is constructed using protégé. The key purpose of semantic web and ontology is to incorporate heterogeneous data and enable interoperability among distinct systems. Ontologies provide lexical items, allow conceptual normalization and provide different types of relations. The domesticated plants ontology is constructed by gathering and analysing domesticated plants domain information from the web. This ontology is used to create an information retrieval system which facilitates context based search thus replacing the keyword based search. In this paper we have presented the methodology by which the ontology is created and also have shown the experiments that was conducted to prove its performance is better than the traditional keyword based that is being used right now.

KEYWORDS: Information Retrieval, Ontology, Semantic Web.

I. INTRODUCTION

Information retrieval is a process of attaining resources relevant to an information desired from a collection of information resources. Searches are based on metadata or on full-text indexing. An Information Retrieval process initiates when a user enters a query into the system. Although search engine technology has evolved a lot in last decade, the Information Retrieval (IR) system is currently built on mostly keywords. The keyword-based IR uses word form matching. There are numerous problems associated with this model and therefore most of the time does not meet user's needs. The key to solve this problem is to move from keyword matching to semantic matching, understood as searching by meaning rather than literal strings. (M Fernandez, 2011)

Retrieval of information efficiently and accurately has become more and more important. A domain ontology (or domain-specific ontology) characterises concepts which are a part of the world. At present, domain ontology serves as a backbone of the Semantic Web by providing vocabularies and formal conceptualization of the given domain to facilitate information sharing and exchange (Guber, 1993). In this study, ontology is used to achieve semantic extension and provide users with better information services to improve customer needs.

In this paper, a model of IR is proposed for domesticated plants based on domain ontology. The experimental data shows that the precision ratio of the ontology based IR model is higher than the general key word based web information retrieval model at a certain extent. The structure of the paper is as follows. First, a brief construction process and implantation on the domesticated plants ontology is given. Second, the IR model is discussed in detail, which includes the key modules and key functions. Next, the test and evaluation of the model are mentioned and finally the conclusion is given.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 5, October 2014

II. THE CONSTRUCTION OF THE ONTOLOGY

Construction method of domain ontology

An ontology is defined as a formal, explicit specification of a shared conceptualization.(Guber, 1993) It provides a common vocabulary to denote the types, properties and interrelationships of concepts in a domain. An ontology may take a variety of forms, but necessarily it will include a vocabulary of terms, and some specification of their meaning. This includes definitions and an indication of how concepts are inter-related which collectively impose a structure on the domain and constrain the possible interpretations of terms. (What is an Ontology?, 2014)Classes are the focus of most ontologies. Classes define concepts in the domain.(McGuinness, 2000)The following steps are needed for building domesticated plants ontology.

Determine the scope and purpose of the domain ontology. This stage clarifies the aim, the scope, and the function which the domain ontology are constructed for. Before the construction, the purpose of the domain ontology should be clear. The domesticated plants ontology provides certain semantic help to improve the efficiency of information retrieval for the webpage information. Therefore the semantic relationship of the concept should be provided as much as possible to improve the information service based on ontology.

Domain information collection and analysis.This stage is a precondition for building domesticated plants ontology. Only when the information and understanding the domain knowledge were fully collected, it is able to build an available and correct ontology with sufficient amount of information. The sources information of domesticated ontology came from Wikipedia, books on domesticated plants, domain experts and other relevant ontologies already existed.

Define the classes and the class hierarchy. Currently there are three class design methods. A top-down development process which starts with the classification of the most general concepts in the domain and consequent specialization of the concepts, a bottom-up development process starts with the classification of the most specific classes, and grouping of these classes into more general concepts and a combination development process is a blend of the top-down and bottom-up approaches.(McGuinness, 2000) In this paper, a top-down approach was applied to construct domesticated plants ontology.

Define the properties of classes. The classes alone will not provide enough information to answer the competency questions. Once we have defined some of the classes, we must describe the internal structure of concepts.(McGuinness, 2000). Initially, it is important to get a comprehensive list of terms between concepts they represent, relations among the terms, or any property that the concepts may have. These terms contain object properties and data type properties. All the sub-classes inherit the properties of the classes.

Create instances. The last step is to create individual instances of the class hierarchy. Defining an individual instance of a class requires 1) choosing a class, 2) creating an individual instance of that class, and 3) filling in the slot values. (McGuinness, 2000)

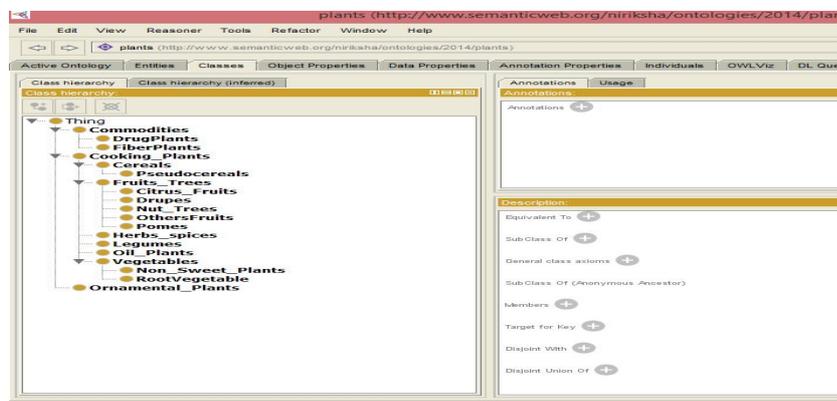
III. THE BUILDING OF THE DOMESTICATED PLANTS ONTOLOGY

The domesticated plants domain ontology describes the concepts and the relationships of the concepts of domesticated plants. In this model, OWL DL was used to describe the ontology concepts, and achieve the domain ontology with the tool of Protégé 5.0 Beta. The plant ontology has many categories to compose, for example, the commodities, cooking plants, ornamental plants, etc., each level comprises of many concept sets (classes, classes' relations and properties). The model determines the category, and then defines the class, the property and the relations to correspond to that category. Finally the classes will be filled with instances. Domain ontology defines the classes and subdivides the corresponding class information. The domesticated plants ontology describes the property of each class, the relationship and the expansion relation.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 5, October 2014



IV. CONSTRUCTION OF THE INFORMATION RETRIEVAL MODEL BASED ON DOMAIN ONTOLOGY

Conventional Web consists of human operator and uses computer systems for tasks like finding, searching and aggregating whereas Semantic Web is the one understood by computers, does the searching, aggregating and combining information without a human operator. It is easily processable by machines, on a global scale. It is the efficient way of representing data on the World Wide Web. A semantic retrieval system was constructed based on the domain ontology, which aims to realise higher efficiency than the keyword-based search engines. This model uses domain ontology to achieve semantic annotation for the domesticated plants websites, and construct semantic information retrieval system by the domesticated plants ontology.

Any Information Retrieval system is supported by the Retrieval process which involves three basic processes, which are as follows:

- i) The representation of the content of the documents,
- ii) The representation of the user's information need, and
- iii) The comparison of the two representations.

The processes are visualized in Figure 1.

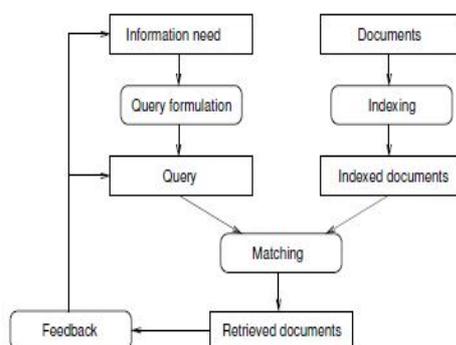


Figure 1: The Retrieval Process

Information Retrieval can be useful for the development, implementation and evaluation of a search engine. Representing the documents is usually called the indexing process. The process of representing the information need of a user is often referred to as the query formulation process. The resulting representation is the query. The comparison of the query against the document representations is called the matching process. Retrieval strategy refers to



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 5, October 2014

information retrieval model. Retrieval strategies assign a measure of similarity between a query and a document. Any Information Retrieval system is based on Information Retrieval process.

Query and Query Expansion

A query consists of a user information need. In its simplest form, a query is composed of keywords and the documents comprising such keywords are searched for. A query can be simply a word or a more complex combination of operations involving several words. The most elementary query that can be formulated in a text retrieval system is a word. The result of word queries is the set of documents containing at least one of the words of the query.

A simple way to formulate user queries is using keywords in place of natural languages to match user's information needs. In Information Retrieval, the user's input queries usually are not detailed enough to allow fully satisfactory results to be returned. Query expansion can solve this problem.

Query Expansion is generally aimed to formulate a user query into one that is more responsive for Information retrieval. Earlier findings have showed that though query expansion had limited retrieval improvement on detailed queries, it demonstrated great potential for significantly improving results given short queries. Since then there have been lot of work that is carried on query expansion by the researchers in the IR community. In literature many query expansion approaches are proposed, and each of them have their own benefits and limitations.

The query given by the user is parsed to identify the parts of speech of the words contained in the query. Then the related synsets for keywords contained in the query are retrieved. The domain keywords that are semantically related to the query are extracted from the ontology built. This step results in the retrieval of more number of semantically related words. These domain keywords are then used for the formation of refined queries. These refined queries are queries with expanded keywords and that has more semantic relevance involved. (Swathi Rajasurya, 2012)

V. IMPLEMENTATION OF THE PROTOTYPE SYSTEM AND ANALYSIS OF THE TEST DATA

1. **ONTOLOGY CONSTRUCTION:** ONTOLOGY, which is a formal representation of knowledge, is a set of concepts within a domain and it forms the knowledge base for our project that is constructed based on the concepts related to the domesticated plants domain. By referring through various plants websites a handful of information is gathered and based on that, an ontology is constructed taking into consideration, the various important areas under the domesticated plants domain.

TOOL USED: Protégé 5.0 Beta

2. **USER INPUT:** The user of the system enters a query related to domesticated plants domain in natural language. The expected output of this query is the semantically relevant web links. The irrelevant links are filtered out.

3.

4. **PARSING OF INPUT QUERY:** The input query given by the user is initially parsed by means of the parser. The parsing is done to analyze the query syntactically which determines the part of speech of each and every word in the query. In this way the given query is analyzed grammatically.

TOOL USED: Stanford Parser

5. **WORDNET:** The output obtained from the parser is sent to the wordnet to get the related synsets of various words contained in the query. So here semantically related words are obtained from the output of the wordnet.

TOOL USED: Wordnet API

6. **EXTRACTION FROM ONTOLOGY:** This process is of more importance, where the information related to the given user query is extracted from the built ontology. The initially given query after passing through the Stanford parser and wordnet, a set of classified and semantically analyzed words are obtained. These words are matched with the concepts contained in the ontology to get a set of more related key words. At the end of this process we get a collection of words which are semantically related and domain specific key words.

TOOL USED: Jena API



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 5, October 2014

7. **FORMATION OF REFINED QUERY:** Next process is the query creation which is done using the collection of these words. The queries formed will be more refined and will fetch more semantically related web links, when we pass these queries as input to the search engines. The refined queries are sent to search API which fetches the web links related to the user query.

TOOL USED: Google Search API

VI. RESULTS AND DISCUSSION

Our experiments reveal the refined query gives more semantically relevant web links and is significantly performing well than giving query directly to Google Search. We have done these experiments during the month of August and September 2014. The first 100 links were manually checked for relevance. The sample queries that were processed in given in the table below.

	Sample Queries	Google	Refined Queries	With Enhanced Query
		Precision		Precision
1	Kinds of cereals	0.83	Variety or Kinds of cereals or Rice	0.93
2	Most popular crop grown by farmers	0.85	Crow or Most popular crop grown by farmers or herbs	0.90
3	Examples of ornamental plants	0.87	Good example or Examples of ornamental plants or Dahlia	0.89
4	Types of fibre plant	0.88	Type or Types of fibre plant or Silk	0.86
5	Pesticides used for vegetables	0.86	Veggie or Pesticides used for vegetables or Potatoes	0.87
6	Plants used for vines	0.88	Plant or Plants used for vines or Hedera	0.92
7	List of nut trees	0.93	Name or List of nut trees or Almonds	0.92
8	Fast growing vegetables	0.85	Rise or Fast growing vegetables or Vegetable oil	0.90
9	Grasses of India	0.85	Grass or Grasses of India or Cypreus	0.87
10	Palm Plants	0.86	Palm tree or Palm Plants or Edible plants	0.89
11	Sweet potatoes good for health	0.87	White potatoes or Sweet potatoes good for health or potatoes	0.87
12	Uses of herbs	0.82	Usage or Uses of herbs or herbal tea	0.81
13	Description about grassland plants	0.82	Verbal description or Description about grassland plants or blue grama grass	0.85
14	Types of garden plants	0.83	Type or types of garden plants or cherry blossom	0.90
15	Description of passiflora	0.85	Genus passiflora or Description of passiflora or passiflora	0.91

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 5, October 2014

Google vs Ontology Based Search



Google vs Ontology Based Search

Independent sample t-test is performed using SPSS and the hypothesis can be set as follows:

H_0 – there is no significant difference between Google search and ontology search.

H_1 – there is significant difference between Google search and ontology search.

T-TEST GROUPS=Algorithm ('Google' 'Ontology')

/MISSING=ANALYSIS

/VARIABLES=Precision

/CRITERIA=CI (.95).

Table 1

Group Statistics

Algorithm		N	Mean	Std. Deviation	Std. Error Mean
Precision	Google	15	.8567	.02820	.00728
	Ontology	15	.8860	.03158	.00815

Table 2

Independent Samples Test											
		Levene's Test for Equality of Variances		t-test for Equality of Means						95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper	
Precision	Equal variances assumed	.415	.525	-2.683	28	.012	-.02933	.01093	-.05172	-.00694	
	Equal variances not assumed			-2.683	27.649	.012	-.02933	.01093	-.05174	-.00693	



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 5, October 2014

Since Levene's test significance value = 0.525 which is > 0.05 two groups are having equal variance. The corresponding p value is $0.012 < 0.05$. So, we reject the null hypothesis.

It has been concluded that "there is significant difference between Google search and ontology search." From Table 1, it is observed that mean for precision using ontology is $0.886 >$ mean for precision used for Google 0.8567 . So, it is statistically confirmed that precision is improved using the proposed query refinement.

VII. CONCLUSION

In this paper we have presented results of some experiments performed in order to evaluate retrieval efficiency of an ontology based approach. It is of no doubt that the ontology we have used has increased the accuracy of the search results. But the accuracy heavily depends on the ontology that is being used for the query refinement process. Though lot of research is being done to develop a unified global ontology, it is yet to become a reality.

REFERENCES

1. *Applied ontology - Clinfowiki*. (2014, September). Retrieved from Applied ontology - Clinfowiki: http://www.informatics-review.com/wiki/index.php/Applied_ontology
2. Awes Siddiqui, P. A. (2013). ONTOLOGICAL APPROACH FOR IMPROVING SEMANTIC WEB SEARCH RESULTS . *IJRET: International Journal of Research in Engineering and Technology* , 30-33. Retrieved from <http://www.slideshare.net/ijretditor/ontological-approach-for-improving-semantic-web-search-results>
3. Guber, T. (1993). A translation approach to portable ontology specification. *Knowledge Acquisition*, 199-220.
4. *Information retrieval - Wikipedia, the free encyclopedia*. (2014, September). Retrieved from Wikipedia, the free encyclopedia: http://en.wikipedia.org/wiki/Information_retrieval
5. M Fernandez, I. C. (2011). Semantically enhanced information retrieval: An ontology-based approach. *Web Semantics: Science, Services and Agents on the World Wide Web*, 432-452.
6. McGuinness, N. F. (2000). *What is an ontology and why we need it*. Retrieved from http://protege.stanford.edu/http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html
7. Ming, T. T.-y. (2012). An Ontology-Based Information Retrieval Model for Vegetables E-Commerce. *Journal of Interactive Agriculture*, 800-807.
8. *Ontology (information science) - Wikipedia, the free encyclopedia*. (2014, August). Retrieved from Wikipedia, the free encyclopedia: [http://en.wikipedia.org/wiki/Ontology_\(information_science\)](http://en.wikipedia.org/wiki/Ontology_(information_science))
9. *Social Research Glossary*. (2007). Retrieved from Quality Research International: <http://www.qualityresearchinternational.com/socialresearch/ontology.htm>
10. Swathi Rajasurya, T. M. (2012). SEMANTIC INFORMATION RETRIEVAL USING ONTOLOGY IN UNIVERSITY DOMAIN.
11. *What is an Ontology?* (2014, September 24). Retrieved from School of Computer Science, The University of Manchester: <http://www.cs.man.ac.uk/~stevensr/onto/node3.html>