Volume 2, Issue 11, November 2012 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Semantic Information Retrieval Using WAMP Server

Meena Unni
Department of Computer Science
Karpagam University
Coiimbatore, 641 021, India

Dr. K. Baskaran
Associate Professor, Department of Computer Science
Government College of Technology
Coimbatore 641 013, India

Abstract -With the enormous growth of internet and its global use, users find it difficult to extract meaningful information from it. To overcome this problem, semantics methods for retrieval of required data are being studied. This research paper explores the possibility of extracting semantic based retrieval of information. First ontology was built which was published on the web. Then the ontology was loaded to MySQL data store. Using PHP language information from this ontology is retrieved using SPARQL. WAMP server which is a package comprising of Apache, MySQL and PHP is used for the process. This paper demonstrates the dominant nature of information retrieval from the web using semantic technologies.

Keywords: semantic web, WAMP server, ARC, Knowledge base, SPARQL

I. INTRODUCTION

World Wide Web is growing at an alarming rate. It is composed of documents written in Hypertext Markup Language (HTML) which is a huge repository of unstructured data. Everyday thousands of pages are added. Managing this large amount of data is a formidable task. Information in the current web is only for human consumption. This has made extracting information difficult. Tim Berners Lee, founder of World Wide Web, recognized its big potential and coined semantic web, his vision for the next generation of the web. The Semantic Web [1] allows the representation and exchange of information in a meaningful way, facilitating automated processing of descriptions on the Web.

Semantic web enriches human readable data with machine readable annotations. It will be easy for machines to understand such a web. The biggest obstacle to annotations is that they adhere to different schemes and structures. Annotations explicitly express links between different web resources and connect these resources to formal terminologies. Such structures are called ontologies. Ontologies provide common vocabularies to be used on the Semantic Web. To further simplify the data integration and automation work, W3C has developed meta data standards such as Resource description framework (RDF) [2], and the web ontology language (OWL).

Part 2 of this paper is dedicated to introducing the components that make up the semantic web: RDF and Ontology. Part 3 proposes design methodology. Part 4 gives detail explanation of implementation using WAMP server. This paper presents the possibility of using the Semantic Web technology to represent, integrate and analyze the information on the web.

II. RDF, ONTOLOGY AND SPARQL

RDF, Ontology and SPARQL form the three core component of semantic web. Semantic web is a web of databases and not of documents, queried by SPARQL [7]. Ontology, RDF and SPARQL collectively play important role in transforming current web to semantic web.

A. RDF

RDF [4] is the first language developed especially for the Semantic Web. It is recommended by W3C for writing machine processable annotations. RDF defines resources using XML. RDF is also called triple because it has three parts subject, object and predicate. Subject and object are names for resources and predicate is the relationship that connects these two things.

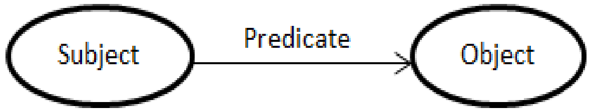

Fig1: Structure for RDF

All Information can be represented in the form of triples. RDF represents relationship between any two data elements, allowing for a very simple model. Below shows information about Carrot expressed as subject predicate object.

Subject	predicate	object
Carrot	is_a	Vegetable
Carrot	HasAlphaCarotene	3477
Carrot	BetaCarotene	8509
Carrot	LuteinZeaxanthin	256
Carrot	HasLycopene	1
Carrot	HasPlantType	Biennial
Carrot	HasVegType	Root

Table 1: Part of the RDF triple relationship for Carrot

B. Ontology

Ontologies link computer and human understanding of symbols. These symbols are also called as relations. Ontology is a specification of a shared conceptualization [3]. Ontology is specific to a domain, and it represent *an area of* knowledge Hence users and domain experts should agree on the knowledge being represented by ontology so that it can be shared and reused.

C. *SPARQL*

SPARQL was standardized by W3C. SPARQL is a query language that is used to query RDF data. It can also be used to query remote RDF server. Like RDF, basic building block of SPARQL query is the triple pattern. A triple pattern is like a triple, but it can have variables in place any of the three positions: subject, predicate or object.

III. DESIGN METHODOLOGY

The key steps for semantic information retrieval using WAMP server is as follows

Workflow steps:

1. Install WAMP server
2. Install ARC
3. Configure MySQL
4. Configure SPARQL Endpoint
5. Create Ontology
6. Load ontology into Mysql
7. Query using SPARQL in PHP

IV. IMPLEMENTATION:

This section describes the implementation of the proposed approach for semantic information retrieval.

A. Install WAMP Server:

For this purpose, open source WAMP server was downloaded and installed. It groups programs that are run on Microsoft windows operating system.

WAMP is an acronym formed from the initials of the operating system Microsoft Windows and the principal components of the package: Apache, MySQL and one of PHP, Perl or Python. Apache is a web server. MySQL is an open-source database. PHP is a scripting language that can manipulate information held in a database and generate web pages dynamically each time content is requested by a browser.

B. Install ARC

ARC provides a RDF system for PHP developers of semantic web. It is an open source, easy to use system. It runs in almost all of the web server environments.

C. Configure MYSQL server

In local MYSQL server, create the database. Create a user and give all permissions on this database to the user.

To create RDF store in MYSQL, augment config.php file with the following code.

```
$store = ARC2::getStore($config);
if (!$store->isSetUp()) {
$store->setUp();}
```

D. Configure SPARQL Endpoint:

A *SPARQL endpoint is* an interface that users use to query an RDF data store by using SPARQL query language. It is a machine friendly interface towards knowledge base. It accepts queries and return result accordingly. This endpoint could be a stand-alone or a web based application. Endpoint returns results in a number of different formats like HTML table, RDF/XML, Turtle, JSON.
Most of the time, results are returned in the form of a HTML table, which is constructed by applying XSL transformations to XML result.
SPARQL endpoints can be classified as generic endpoints and *specific* endpoints. If the endpoint is tied to a specific dataset, it is called as specific endpoint. If the endpoint can query any RDF dataset that is stored locally or accessible from the web, it is called as generic.
SPARQL endpoint can be configured by giving appropriate host name, database name, database user name, database password and store name.
This paper uses ARC SPARQL endpoint to query the RDF datastore.

E. Create Ontology:

Example ontology is created for nutrition domain for Vitamin A. It has DietarySource, Disease, EffectsOnHuman, FormsofVitaminA, Interaction, Person, Supplement as the main classes. Ontology is created as class subclass relationship.

Below shows class subclass hierarchy for Carrot.

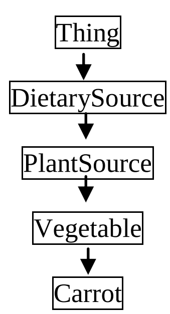

Fig2 : Ontology Classes

Ontology also defines properties and restrictions on the domain. Individuals also called instances are created. Ontology is then published on the web. For each instances data properties are defined like AlphaCarotene, BetaCarotene, LuteinZeaxanthin, Lycopene, HasName,HasPlantType, HasQuantity, HasUnitOfMeasurement, HasVegType. RDF/XML representation for Carrot from the ontology is given below.

```
<Vegetable rdf:ID="Carrot">
    <HasAlphaCarotene rdf:datatype="&xsd;int">3477</HasAlphaCarotene>
    <HasBetaCarotene rdf:datatype="&xsd;int">8509</HasBetaCarotene>
    <HasLuteinZeaxanthin rdf:datatype="&xsd;int">256</HasLuteinZeaxanthin>
    <HasLycopene rdf:datatype="&xsd;int">1</HasLycopene>
    <HasName rdf:datatype="&xsd;string">Carrot</HasName>
    <HasPlantType rdf:datatype="&xsd;string">Biennial</HasPlantType>
    <HasQuantity rdf:datatype="&xsd;int">100</HasQuantity>
    <HasUnitOfMeasurement rdf:datatype="&xsd;string">gram</HasUnitOfMeasurement>
    <HasVegType rdf:datatype="&xsd;string">Root</HasVegType>
  </Vegetable>
```

Once ontology is finalised, it is published on the web.

F. Load data into MySQL:

After finalising the ontology, it has to be loaded into RDF data store, MySQL. There are two ways to load OW/RDF data into MYSQL OWL/RDF data can be loaded into MYSQL either through command line or through PHP application code. Whichever is the method used to load OWL/RDF data into MYSQL, the basic load statement is given below. Either this can be executed at the command prompt or inserted in a PHP application code.

```
$store->query('LOAD <http://mu123.site90.net/vit2.owl>');
```

G. Retrieval of Information

Once data is loaded into RDF data store, SPARQL query can be embedded into PHP code to get the desired result.

Below shows PHP code using SPARQL to retrieve information regarding the resource name and their vitamin A content, from the RDF store.

```
<?php
include_once("arc/ARC2.php");
include_once('config.php');
$store = ARC2::getStore($arc_config);
if (!$store->isSetUp())
  {  $store->setUp(); }
$q ='PREFIX ta:
<http://www.owl-ontologies.com/vit2.owl#>
SELECT *
from <http://mu123.site90.net/vit2.owl>
WHERE
{
            ?s1 ta:HasName ?s.
            ?s1 ta:HasVitaminAinIU ?p.
}';

$rows = $store->query($q, 'rows');
$r = '';
if ($rows = $store->query($q, 'rows'))
```



```
{
        $r = '<table border=1>
        <th>Name</th><th>VitaminA</th>'."\n";
        foreach ($rows as $row)
        {       $r .= '<tr><td>'.$row['s'] .
                '</td><td>'.$row['p'] . '</td></tr>'."\n";
        }
        $r .='</table>'."\n";
}
Else
{       $r = '<em>No data returned</em>'; }
echo $r;
?>
```

The output of the given code is shown below

Name	VitaminA
Apricot	1926
Papaya	1094
Spinach	9376
Beef	0
Mango	765
Parsley	8425
ChestNut	28
Melon	569

Fig 3: Output Screen

V. CONCLUSION AND FUTURE WORK:

As of today Web consists of millions of web pages. Data is represented in HTML pages. Hence it is inefficient for meaningful information extraction. This paper focusses on ways to enhance the search results by using ontology and RDF. WAMP server is used to extract information from RDF store where in SPARQL query is embedded in PHP program. This work can be further enriched to achieve intelligent fuzzy retrieval.The semantic web is better suited for data integration and for knowledge representation. RDF and OWL along with SPARQL play an important role in information retrieval from the semantic web.

REFERENCES

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

[2] Davis, Ian, “An Introduction to RDF”, http://research.talis.com/2005/rdf-intro/

[3] Dieter Fensel · Holger Lausen · Axel Polleres, Jos de Bruijn · Michael Stollberg · Dumitru, Roman John Domingue Enabling SemanticWeb Services, SPRINGER, 2007

[4] G. Klyne and J. J. Carroll, editors. *Resource Description Framework (RDF): Concepts and Abstract Syntax*. W3C Recommendation, 10 February 2004. Available from http://www.w3.org/TR/rdf-concepts/.

[5] www.wampserver.com/en

[6] Noy, N., McGuinness, D.L, Ontology Development 101: A guide to creating your first ontology, Stanford Medical Informatics Technical Report No SMI-2001-0880, 2000

[7] Siddharth Gupta , Narina Thakur, “Semantic Query Optimization with Ontology Simulation” , International journal of web & semantic Technology (IJWesT) Vol.1 , Num 4 , October 2010.

