# Semantic Web Development for Tourism Domain from Conventional Web and Improving the Semantic Search by Providing Different Methods of Categorization

Jayaprabha P
Department of Computer Applications,
VidyaaVikas College of Engineering & Technology,
Tiruchengode, Tamilnadu, India – 673 601.

Saradha A
Department of Computer Science Engineering,
Institute of Road& Transport Technology, Erode,
Tamilnadu, India – 673 601.

## ABSTRACT

World Wide Web is the biggest repository that contains all sorts of information. User can search necessary information from the web. All the information in the WWW are unstructured and human readable. Search engine provides result based on the user query is not satisfactory due to unstructured and semi-structured nature of web document. The advent of semantic web overcomes this difficulty, the information in this web are structured, machine readable and it is very easy to integrate. Now-a-days, most of the websites are developed based on the semantics, but there is no technology to convert existing web information into RDF. The proposed approach converts existing web information related to tourism domain in to semantic web using Resource description Framework. To obtain more relevant result the converted RDF web source are categorized and stored in the repository based on different algorithms.

## Keywords

Metadata, Semantic web, Ontology, OWL, RDF, Tourism Information System, Information Retrieval, Search Engine, Web Categorization, Pattern Analysis.

## 1. INTRODUCTION

The Web is becoming a dominant information repository. However, retrieval of relevant information is still a challenging task for most of its users. Ambiguity of words is one of the main hindrances in information retrieval[4][5]. Employment of semantic technologies in search systems is seen as a promising approach to improve the current state of the art [6]. Semantic technologies are applied in different ways: semantic annotations of content [7]; clustering of retrieved documents according to topics[8]; powerful querying languages[9]; or creating structured semantic models of retrieved documents[10].

Information present in the traditional web are unstructured, semi-structured, syntactic nature, difficult to integrate and more human readable. Search engine faces difficulty to extract information from traditional search. Drawbacks of this search engines has been discussed in following section. Data in the Semantic Web is defined and linked in a way that can be used for more effective discovery, automation, integration and reuse across applications. This data can be shared and processed by automated tools as well as people.

## Difference between traditional and semantic search engine

At the core, a semantic search engine has the ability to understand the relationships between keywords, phrases or parts of speech within a search phrase, therefore allowing it understand the underlying meaning of the entire phrase. For example, a semantic search engine would be able to easily distinguish the differences between the following phrases made up of the same 'keywords' but with obvious different implications:

- *Chennai*
- *Madras*

## ***So how is semantic search different from traditional web search?***

In the example above, the phrases are made up of the different keywords, while the subject/action relationships are same. In traditional web search, which are based on ranking algorithms, since the relationships between the sentence parts are unknown, the engines would return identical or nearly identical results, even though it was being asked two completely different questions. Additional problems with web search also arise when the keywords are too specific, producing few or no results, or too general, in which case the results are overwhelming and irrelevant.

Alternatively, since semantic search technology understands the meaning of the above sentences, it would be able to produce highly relevant *answers* to the questions. The goal of semantics is to always provide the direct insights and answers needed to complete research tasks, rather than burying those ideas among scores of irrelevant documents.

The final purpose of this engine is enhancing performance of traditional search engines (especially Precision and Recall). It's possible through understanding the context of documents and queries. One of the most important parts of this type is annotator which responsible for generating metadata for crawled pages. It is needed to generate some metadata for user's query in order to detect its context. Here usually after traditional retrieval, combine matching RDF graphs to obtain better quality of results.

## 2. RELATED WORK:

The next generation of Tourism Information System is expected to be: enable semantics-based information processing, exhibit natural language capabilities, facilitate inter-organization exchange of information in a seamless way, and evolve proactively in tandem with dynamic user requirements [2], [3].

The OntoWebbersystem[11] is an ontology-based approach to website management. It facilitates the design, creation, generation and maintenance of Web sites using a set of software tools. It also enables the personalization of Web site views based on individual users. Another notable approach is the Hera project[12], which is a methodology that supports the design and engineering of Web Information Systems (WIS) using Semantic Web technology. The main focus of the Hera project is to support Web design and implementation particularly hypermedia aspects.

## 3.0 TECHNOLOGIES USED

### 3.1 Semantic web

The Semantic Web is used for representing information in the World Wide Web in a machine-readable fashion: such that it can be used by machines not just for display purposes, but for automation, integration and reuse across applications.

These machine-interpretable descriptions allow more intelligent software systems to be written, automating the analysis and exploitation of web-based information. Software agents will be able to create automatically new services from already published services, with potentially huge implications for models of e-Business.

In the words of Tim Berners-Lee, "The Semantic Web is a web of data, in some ways like a global database," and the Semantic Web effort is developing "languages for expressing information in a machine processable form." The Semantic Web makes use of structured text, rather than natural language, to identify knowledge and its relationship with other knowledge or data. Berners-Lee's original vision involves the action of intelligent software "agents" on computers and handheld devices that would act autonomously to both retrieve data and interact with Web sites through specialized tagging of data and content on these sites.

The Semantic Web functions because of a highly specialized type of data and information tagging that can be implanted within Web pages.

Fundamentally, each Semantic Web tag, known as a triplet, links together a subject, verb, and object, creating a relationship between them. Three simple examples of a semantic tag might be: <Boston><is in the state of><Massachusetts>; <Beacon Hill><is a neighborhood of><Boston>; <I><like><Boston>. These two statements each consist of two nouns separated by a verb and are interlinked with one another through Boston.

Semantic web uses many ontology languages to describe semantic data. Some of the ontology languages are follows

- RDF ( Resource Description Framework)
- OWL ( Web Ontology Language)
- DAML ( DARPA Agent Markup Language)
- SPARQL ( Simple Protocol and RDF Query Language)
- GRDLL ( Gleaning Resource Descriptions from Dialects of Languages)
- OIL ( Ontology Inference Layer)

### 3.2 Ontology

Information integration from different sources needs to be a shared by understanding of the relevant domain. Knowledge representation formalisms provide structures for organizing this knowledge, but provide no mechanisms for sharing it.

Ontologies provide a common vocabulary to support sharing and reuse of knowledge. Ontology is a fundamental component for achieving the Semantic Web. Ontology has the capability to solve a number of problems in tourism. This includes: 1) enabling interoperability of heterogeneous platforms; 2) standardization of business models, business processes, and knowledge architectures; and 3) serving as a model of knowledge representation for the generation of knowledge-based information services [1].

### 3.2 The *Resource Description Framework*

RDF provides a means for adding semantics to a document without making any assumptions about the structure of the document. It is an XML application customized for adding Meta information to Web documents.

The Resource Description Framework attempts to address XML's semantic limitations. It presents a simple model that can be used to represent any kind of data. This data model consists of nodes connected by labeled arcs, where the nodes represent web resources and the arcs represent properties of these resources. It should be noted that this model is essentially a semantic network, although unlike many semantic networks, it does not provide inheritance.

The nodes/arcs model also means that RDF is inherently binary. However, this does not restrict the expressivity of the language because any n-array relation can represented as a sequence of binary relations. RDF can be exchanged using an XML serialization syntax.

The basic syntax consists of a Description element which contains a set of property elements. The about attribute identifies which resource is described. The property rdf:type is used to express that a resource is a member of a given class, and is equivalent to the instance-of link used in many semantic nets and frame systems. There are a number of abbreviated variations of the RDF syntax, which is an advantage for content providers but requires more complex RDF parsers.

It is important to note that all of these syntaxes have a well-defined

```
<?xml version="1.0"?>
<RDF xmlns="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:g="http://schema.org/general#">
<Description about="http://www.state.edu/users/jsmith">
<type resource="http://schema.org/university#Chair" />
<g:name>Jane Smith</g:name>
</Description>
</RDF>
```

An RDF Instance.

```
<?xml version="1.0"?>
<RDF xmlns="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:g="http://schema.org/general#"
xmlns:u="http://schema.org/university#">
<u:Chair about="http://www.state.edu/users/jsmith"
g:name="Jane Smith" />
</RDF>
```

An Abbreviated RDF Instance. Mapping into the RDF data model, and thus avoid some of the problems with representational choices in basic XML. Nevertheless, it is still easy to create different representations for a concept.

To prevent accidental name clashes between different vocabularies, RDF assigns a separate XML namespace to each vocabulary (these vocabularies, called schemas, can be formally defined using RDF Schema as discussed below).

This approach has two disadvantages. First, since namespaces can be used with any element and RDF schemas need not be formally specified, it is possible to write RDF statements such that it is ambiguous as to whether certain tags are RDF or intermeshed tags from another namespace. Second, namespaces are not transitive, which means that each RDF section must explicitly specify the namespace for every schema that is referenced in that section, even for schemas that were extended by a schema whose namespace has already been specified.

A significant weakness of RDF is that it does not specify a schema inclusion feature. Although namespaces allow a document to reference terms defined in other documents, it is unclear as to whether the definitions of these terms should be included. In fact, it is unclear what constitutes the definition of a term.

The data model of RDF provides three object types: resources, property types, and statements.

- **A *resource*** is an entity that can be referred to by a address at the WWW (i.e., by an URI). Resources are the elements that are described by RDF statements.
- **A *property*** defines a binary relation between resources and/or atomic values provided by primitive data type definitions in XML.
- **A *statement*** specifies for a resource a value for a property. That is, statements provide the actual characterizations of the Web documents.

The Semantic Web is a web of data. There is lots of data we all use every day, and it's not part of the web.

## 4. ABOUT TOURISM

Tourism plays major role in entertaining all sorts of people from young age to elder, poor person to rich one's. One reason for high information exchange ratio among different players in tourism industry is the tourism product itself. In comparison to many other products, which are sold online, tourism product is immaterial, heterogeneous and non-persistent. Each of these characteristics has influence on the information exchange within tourism industry.

A trip usually includes many parts such accommodation, transportation, insurance, visa services, guide services, excursions in the destination. Due to the heterogeneity of the travel product, travel agency consultant or a person who is planning the trip itself must have access to different sources of information.

Tourism product is also immaterial, meaning that traveler cannot see or touch the tourism product before the trip. That is why reliable information about destination, accommodation options and other parts of the tourism product is extremely important for both people working in tourism industry and tourists themselves.

Tourism product cannot be stored in storage. If a hotel room or a seat in an airplane remains empty today, this is lost revenue for the tourism company. This is a reason why effective distribution and inventory management are key factors in the tourism business.

## 5. PROPOSED WORK

Domain selected for proposed work is tourism. The popularity of the WWW resulted a flurry of websites covering tourism related information covering almost everything in the universe.

Tourism is one of important domain referring to many factors and has plenty of domain knowledge, which is the essential base of travel information systems.

The proposed work is to convert existing unstructured heterogeneous tourism data into semantic web format such as Resource Description Framework and the same semantic web resources are categorized using two different techniques to enrich the search result of the user query. The work is categorized into six modules namely

1. Preprocess
2. Process1
3. Process2
4. Search

Each module will be elaborated in the following section.

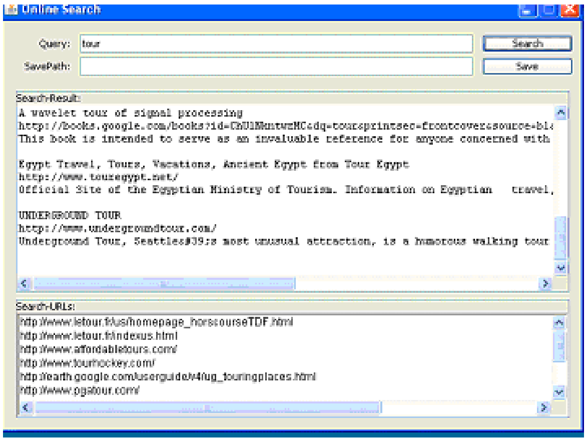
### 5.1 Preprocess

This module extracts heterogeneous tourism related text based on the keyword and converts into semantic web format RDF. This module contains three sub modules namely

1. Downloading unstructured text using online search
2. Converting HTML into XML
3. Mapping Process

### 5.1.1 Downloading unstructured text using online search

This module extracts heterogeneous tourism related text based on the keyword we specify in the online search dialogue. This module links Google search engine and extracts all the related travel information and stores in the specified path and in specified file name. The search result may be either single url with single web page or thousands of url with more web pages. Now these documents are in HTML format. For example, use keyword tour in the online search module and this module downloads all the web pages related to keyword tour and stores in the name of Tourism1 , Tourism2 , Tourism3 etc., This is shown in the fig. 1.



Query: tour
SavePath:

Search-Result:

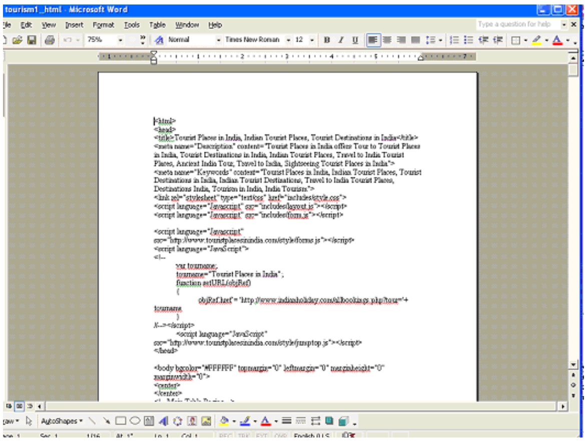
```
A wavelet tour of signal processing
http://books.google.com/books?id=ChU1NuntwzMC&dq=tour&printsec=frontcover&source=bl&
This book is intended to serve as an invaluable reference for anyone concerned with

Egypt Travel, Tours, Vacations, Ancient Egypt from Tour Egypt
http://www.touregypt.net/
Official Site of the Egyptian Ministry of Tourism. Information on Egyptian   travel,

UNDERGROUND TOUR
http://www.undergroundtour.com/
Underground Tour, Seattle&#39;s most unusual attraction, is a humorous walking tour
```

Search-URLs:

```
http://www.letour.fr/us/homepage_horscourseTDF.html
http://www.letour.fr/indexus.html
http://www.affordabletours.com/
http://www.tourhockey.com/
http://earth.google.com/userguide/v4/ug_touringplaces.html
http://www.pgatour.com/
```

**Fig. 1 Downloadingurl's**

From the downloaded url's, consider the first urlie Tourism1.html and it is shown in the fig. 2.


tourism1_.html - Microsoft Word
File Edit View Insert Format Tools Table Window Help
Type a question for help
75%
Normal
Times New Roman
12
1/16 At 1" Ln 1 Col 1 REC TRK EXT OVR English (U.S.

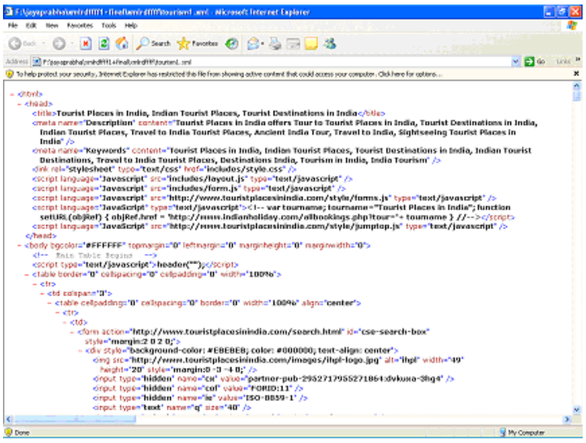
```
<html>
<head>
<title>Tourist Places in India, Indian Tourist Places, Tourist Destinations in India</title>
<meta name="Description" content="Tourist Places in India offers Tour to Tourist Places
in India, Tourist Destinations in India, Indian Tourist Places, Travel to India Tourist
Places, Ancient India Tour, Travel to India, Sightseeing Tourist Places in India">
<meta name="Keywords" content="Tourist Places in India, Indian Tourist Places, Tourist
Destinations in India, Indian Tourist Destinations, Travel to India Tourist Places,
Destinations India, Tourism in India, India Tourism">
<link rel="stylesheet" type="text/css" href="includes/style.css">
<script language="Javascript" src="includes/layout.js"></script>
<script language="Javascript" src="includes/form.js"></script>

<script language="Javascript"
src="http://www.touristplacesinindia.com/style/forms.js"></script>
<script language="JavaScript">
<!--
        var tourname;
        tourname="Tourist Places in India";
        function setURL(objRef)
        {
                objRef.href="http://www.indianholiday.com/allbookings.php?tour="+
tourname
        }
//--></script>
        <script language="JavaScript"
src="http://www.touristplacesinindia.com/style/jumptop.js"></script>
</head>

<body bgcolor="#FFFFFF" topmargin="0" leftmargin="0" marginheight="0"
marginwidth="0">
<center>
</center>
```

**Fig. 2 Tourism1.html**

## 5.1.2 Converting HTML into XML

Tidy is a software which automatically converts HTML into XML. It has some restriction that all the opening tags in HTML should have close tag and it should not contain any comment line etc. So this can't be used for the above implementation. So we used a module called HtmlConvertor which in turn uses SAX parser to read HTML data and converts to XML. It reads each and every tag in HTML and converts into related XML tag. While reading HTML tag, all the opening tags should have closing tag, otherwise it will show error in particular line, then we have to rectify this error manually. In this way HTML content are converted into XML.

Now this Tourism1.html is converted into Tourism1.xml by HtmlConvertor module and shown in the fig. 3


F:\jayaprabha\xmlrdffff1-final\xmlrdffff\tourism1.xml - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address F:\jayaprabha\xmlrdffff1-final\xmlrdffff\tourism1.xml
Go Links
Done
My Computer

To help protect your security, Internet Explorer has restricted this file from showing active content that could access your computer. Click here for options...

```
- <html>
 - <head>
    <title>Tourist Places in India, Indian Tourist Places, Tourist Destinations in India</title>
    <meta name="Description" content="Tourist Places in India offers Tour to Tourist Places in India, Tourist Destinations in India,
      Indian Tourist Places, Travel to India Tourist Places, Ancient India Tour, Travel to India, Sightseeing Tourist Places in
      India" />
    <meta name="Keywords" content="Tourist Places in India, Indian Tourist Places, Tourist Destinations in India, Indian Tourist
      Destinations, Travel to India Tourist Places, Destinations India, Tourism in India, India Tourism" />
    <link rel="stylesheet" type="text/css" href="includes/style.css" />
    <script language="Javascript" src="includes/layout.js" type="text/javascript" />
    <script language="Javascript" src="includes/form.js" type="text/javascript" />
    <script language="Javascript" src="http://www.touristplacesinindia.com/style/forms.js" type="text/javascript" />
    <script language="JavaScript" type="text/javascript"><!-- var tourname; tourname="Tourist Places in India"; function
      setURL(objRef) { objRef.href = "http://www.indianholiday.com/allbookings.php?tour="+ tourname } //--></script>
    <script language="JavaScript" src="http://www.touristplacesinindia.com/style/jumptop.js" type="text/javascript" />
  </head>
 - <body bgcolor="#FFFFFF" topmargin="0" leftmargin="0" marginheight="0" marginwidth="0">
    <!--  Main Table Begins  -->
    <script type="text/javascript">header("");</script>
  - <table border="0" cellspacing="0" cellpadding="0" width="100%">
    - <tr>
      - <td colspan="3">
        - <table cellpadding="0" cellspacing="0" border="0" width="100%" align="center">
          - <tr>
            - <td>
              - <form action="http://www.touristplacesinindia.com/search.html" id="cse-search-box"
                  style="margin:2 0 2 0;">
                - <div style="background-color: #EBEBEB; color: #000000; text-align: center">
                    <img src="http://www.touristplacesinindia.com/images/ihpl-logo.jpg" alt="ihpl" width="49"
                      height="20" style="margin:0 -3 -4 0;" />
                    <input type="hidden" name="cx" value="partner-pub-2952717955271864:dvkuxa-3hg4" />
                    <input type="hidden" name="cof" value="FORID:11" />
                    <input type="hidden" name="ie" value="ISO-8859-1" />
                    <input type="text" name="q" size="40" />
```

**Fig.3 Tourism1.xml**

### 5.1.3 Mapping Process

Travel2.owl is an ontology based RDF file which contains the travel related information in turns of classes, instances, sub instances, object property, relation, data property, general axioms etc., This file has more 500 classes namely travel, Destination, contact, accommodation, hotel, sightseeing etc.,

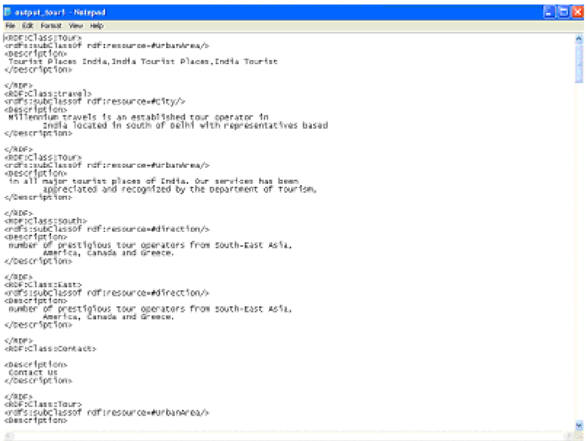
Accommodation has object property hasRating ,isOfferedAt etc., and sub instances are Collection in turn, TwoStarRating, OneStarRating, ThreeStarRating etc.

Destination has data property hasActivity.

Activity is disjoint with Relaxation, Sightseeing, Sports etc.,

Relaxation is sub class of Yoga.

Now the xml tags are mapped with classes in Travel2.owl and converted into RDF and it is shown in the fig. 4.



```
<RDF:Class:Tour>
<rdfs:subClassof rdf:resource=#UrbanArea/>
<Description>
 Tourist Places India,India Tourist Places,India Tourist
</Description>

</RDF>
<RDF:Class:travel>
<rdfs:subClassof rdf:resource=#City/>
<Description>
 Millennium travels is an established tour operator in
        India located in south of Delhi with representatives based
</Description>

</RDF>
<RDF:Class:Tour>
<rdfs:subClassof rdf:resource=#UrbanArea/>
<Description>
 in all major tourist places of India. Our services has been
        appreciated and recognized by the Department of Tourism,
</Description>

</RDF>
<RDF:Class:South>
<rdfs:subClassof rdf:resource=#direction/>
<Description>
 number of prestigious tour operators from South-East Asia,
        America, Canada and Greece.
</Description>

</RDF>
<RDF:Class:East>
<rdfs:subClassof rdf:resource=#direction/>
<Description>
 number of prestigious tour operators from South-East Asia,
        America, Canada and Greece.
</Description>

</RDF>
<RDF:Class:Contact>

<Description>
 Contact Us
</Description>

</RDF>
<RDF:Class:Tour>
<rdfs:subClassof rdf:resource=#UrbanArea/>
<Description>
```

**Fig.4. Tourism1.rdf**

Contact has object property hasEmail, hasStreet, hasZipcode etc.,

XML tag is mapped with classes, sub classes, object property, data property etc., in travel2.owl. If mapping occurs the corresponding is converted into RDF and stored in output file.

Now this RDF file will be available for semantic search.

## 5.2 Process1 – Categorize offspring

After converting the existing web content into semantic web, now these web contents are categorized using four different process. As the World Wide Web changes dynamically, it's impossible to structure a perfect ontology including the relations of all the web resources. In Semantic web, a class is a type of web resources. A class has some properties and it can be a subclass of another class. Thus, it cannot be matched by the ontology-based search engines for a web resource which is not defined as a class in the structured ontologies. To use Web resources efficiently, it becomes an important and emerging issue to detect the potential relations on existing ontologies.

For this purpose, a web resource categorization method is used to extract potential relations among the Web resources. The process1 module extracts the relation by using the knowledge such as the class hierarchy in ontologies, the relations of word in dictionaries, and the description of Web resources.

In this **Proces1**, it is intend to categorize the classes well defined in ontologies. This is a traditional method in Semantic Web research area. If a class has a relation to a category, its subclasses ("rdf:subClassOf") could have a weaker relation to the category. The equivalent classes (“owl:equivalentClass” and “owl:sameAs”) have equivalent relation strength.

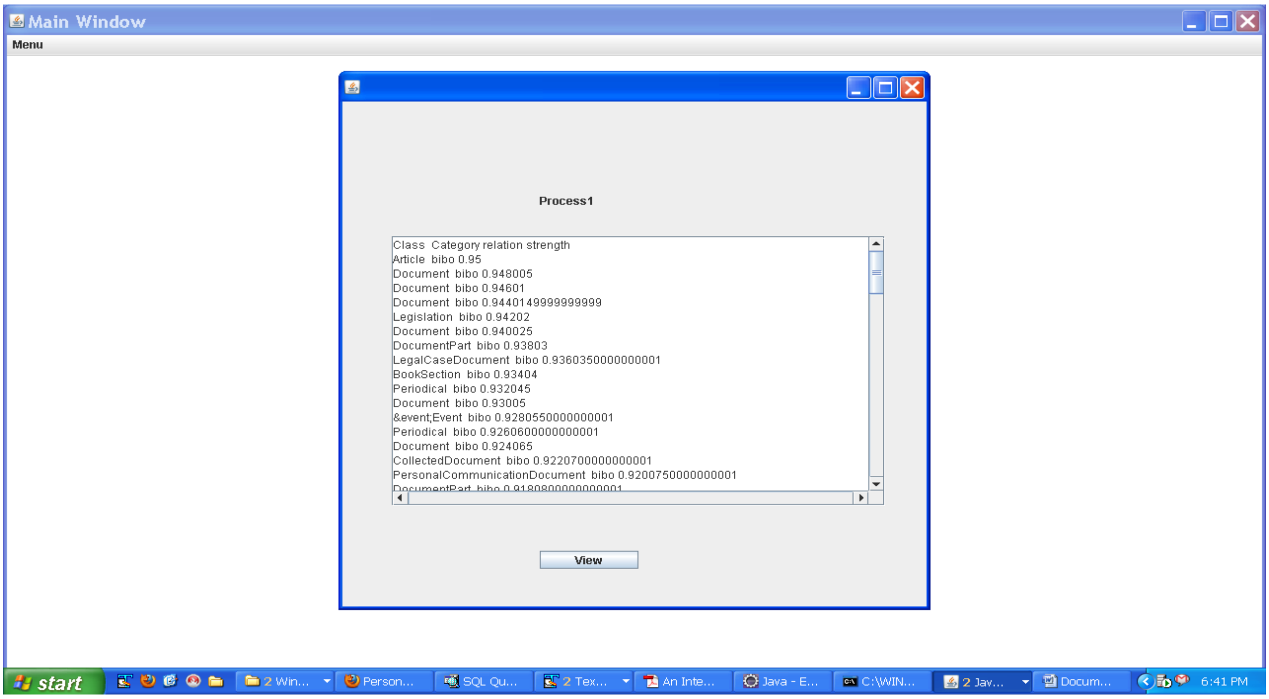
**Process1** is performed by using a recursive function RER(c0, a, rc0,a). Let c0 be a class in category a with the relation strength as rc0,a. If class c, which is an equivalent class or a subclass of class c0 in category a, has not been categorized to category a, RER(c0, a, rc0,a) can categorize class c to category a with relation strength as rc,a. By recursively calling itself, RER(c0, a, rc0,a) categorizes all the offspring classes of c0. Here, given a coefficient k for the degressive strength of relation in the hierarchical structure. While

k = 1 equivalentClass– relation Eq.(2)
k =0.95 subClass− relation.

checking the number of categorized classes for a set of values of k, found that the number increase much while k is not larger than 0.95, and increase very little when k is larger than 0.95. Since too high value of k decreases the precision of categorization result, specify k as 1 for a same level and as 0.95 for a lower level class. The relation strength of class c to category a is calculated by Eq. (1).

$$rc,a = k \times rc0,a \quad \Big\} \quad \text{Eq.(1)}$$

**Fig.5 Process1 – Categorize offspring**



```
Class  Category relation strength
Article  bibo 0.95
Document  bibo 0.948005
Document  bibo 0.94601
Document  bibo 0.9440149999999999
Legislation  bibo 0.94202
Document  bibo 0.940025
DocumentPart  bibo 0.93803
LegalCaseDocument  bibo 0.9360350000000001
BookSection  bibo 0.93404
Periodical  bibo 0.932045
Document  bibo 0.93005
&event;Event  bibo 0.9280550000000001
Periodical  bibo 0.9260600000000001
Document  bibo 0.924065
CollectedDocument  bibo 0.9220700000000001
PersonalCommunicationDocument  bibo 0.9200750000000001
DocumentPart  bibo 0.9180800000000001
```

### 5.3 Process2 – Categorize by important class name

By performing **Process1**, some classes categorized for each category. In Process2, extract the class names that represent the characters of category a, then categorize the classes having the important class names to category a.

In **Process2**, categorize the classes not existing in the ontology based on the important class names. For each category a, extract all the class names by a function getName(a). Then compute a coefficient l to evaluate how much a class name cNamecan represent the character of a by function getW(cName, a) based on tfidf [7]. The weight of class name ci in category ajis denoted by Wi,j, which is calculated by Eq. (3).

$$Wi,j = tfi,j \times \log(\ Ndf) \quad \text{Eq.(3)}$$

tfi,j= number of occurrences of i in j,
dfi= number of categories containing i,
N = total number of categories in A.

Since Wi,jis largely affected by the number of classes in a category, we normalize Wi,jin [0, 1], and W_ i,jdenotes the normalized value which can be calculated by Eq. (4).

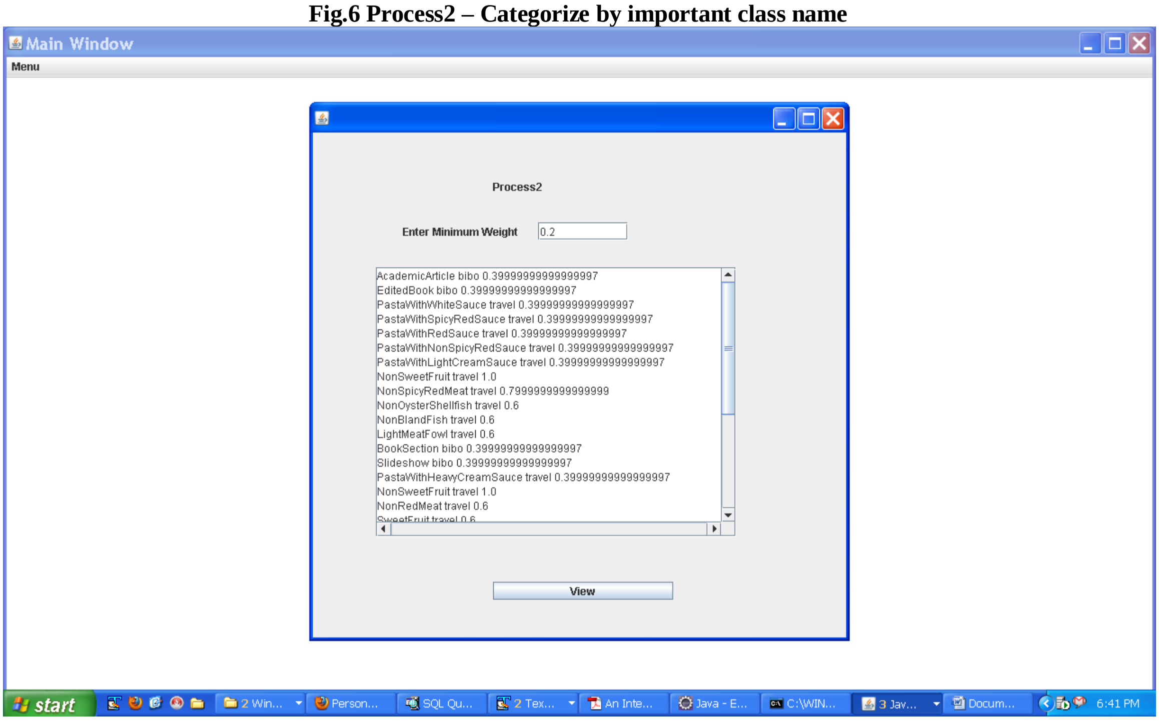
$$W\_\ i,j = Wi,j \ \max(Wi,j) \quad \text{Eq.(4)}$$

Consider the class names having W_ i,jlarger than a threshold minWas the important class names for the category. In the preliminary experiments, we got good result of the categorization while giving a threshold minWranging from 0.02 and 0.04. Here, we specified it as 0.03 by considering both effectiveness and efficiency. The algorithm is described concretely as follows. For each important class name cName, we extract all the classes having class name cNamein category a from all the classes by function getClass(cName). From these classes, extend category a with each class c which denotes a class having a class name of cNamebut not existing in category a. For a set of classes having class name cNameand existing in category a, getR(cName, a) is a function to get the relation strengths of the classes in the set to category a, and avg(getR(cName, a)) denotes the average. Then, the relation strength of c to a can be calculated by Eq. (5).

$$rc,a = \mathrm{avg}(getR(cName, a)) \times W\_\ i,j \quad \text{Eq.(5)}$$

In the **Process2**, extend the categories with all the offspring classes of the newly categorized classes by executing **Process1** again.

**Fig.6 Process2 – Categorize by important class name**



Process2

Enter Minimum Weight: 0.2

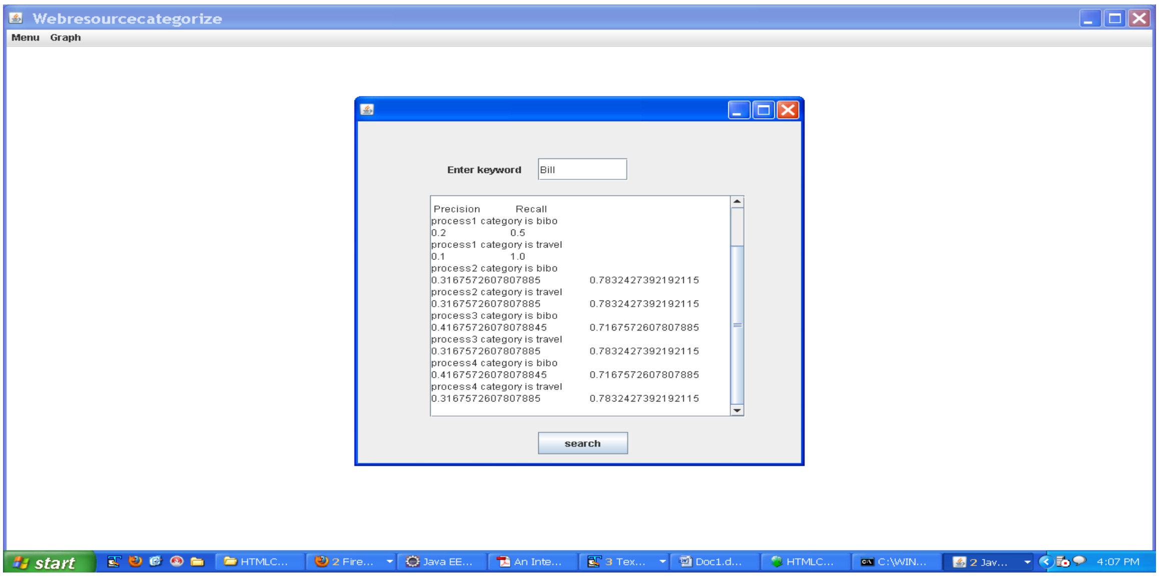
```
AcademicArticle bibo 0.39999999999999997
EditedBook bibo 0.39999999999999997
PastaWithWhiteSauce travel 0.39999999999999997
PastaWithSpicyRedSauce travel 0.39999999999999997
PastaWithRedSauce travel 0.39999999999999997
PastaWithNonSpicyRedSauce travel 0.39999999999999997
PastaWithLightCreamSauce travel 0.39999999999999997
NonSweetFruit travel 1.0
NonSpicyRedMeat travel 0.7999999999999999
NonOysterShellfish travel 0.6
NonBlandFish travel 0.6
LightMeatFowl travel 0.6
BookSection bibo 0.39999999999999997
Slideshow bibo 0.39999999999999997
PastaWithHeavyCreamSauce travel 0.39999999999999997
NonSweetFruit travel 1.0
NonRedMeat travel 0.6
SweetFruit travel 0.6
```

In the above process, the semantic web resources are categorized for efficient search of data for user query.

## 5.3 Search:

This module is used to find precision and recall factor for particular keyword that occurred in all the above four process. That is relation strength of a category and the classes.

**Fig.7 Search module to find relation strength among four process**



Enter keyword: Bill

```
 Precision          Recall
process1 category is bibo
0.2                0.5
process1 category is travel
0.1                1.0
process2 category is bibo
0.3167572607807885          0.7832427392192115
process2 category is travel
0.3167572607807885          0.7832427392192115
process3 category is bibo
0.41675726078078845         0.7167572607807885
process3 category is travel
0.3167572607807885          0.7832427392192115
process4 category is bibo
0.41675726078078845         0.7167572607807885
process4 category is travel
0.3167572607807885          0.7832427392192115
```

## 6. CONCLUSION AND FUTURE ENHANCEMENTS

Tourism related heterogeneous unstructured data from various web sites is extracted and using tools, it is converted into semantic web format called RDF. The proposed work is to convert existing unstructured heterogeneous tourism data into semantic web format such as Resource Description Framework and the same semantic web resources are categorized using four different techniques to enrich the search result of the user query. Implementation has been done in java. In future this work will be extended to several domains.

## 7. REFERENCES

[1] Daramola, O., Adigun, M., Ayo, C., Building an Ontology-based Framework for tourism Recommendation Services, ENTER 2009, pp. 135-147, Amsterdam, Netherlands (2009).

[2] Werthner, H. and Klein, S.: Information Technology and Tourism—A Challenging Relationship, Springer-Verlag, New York, 23. (1999).

[3] Staab, S., Werthner, H., Ricci, F., Zipf, A., Gretzel, U., Fesenmaier, D.R., Paris, C., and Knoblock, C.: Intelligent systems for tourism, IEEE Intelligent Systems, Volume 17, Issue 6, Nov/Dec, 53-66. (2002).

[4] Bhogal, J., Macfarlane, A. & Smith, P. (2007) 'A review of ontology based query expansion', Inf. Process. Manage., Vol. 43, No. 4, pp. 866-886.

[5] Carmel, D., Yom-Tov, E., Darlow, A. &Pelleg, D. (2006) 'What makes a query difficult?', paper presented to the Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, Seattle, Washington, USA.

[6] Horrocks, I. (2007) 'Semantic web: the story so far', paper presented to the Proceedings of the 2007 international cross-disciplinary conference on Web accessibility (W4A), Banff, Canada.

[7] Moscato, F., Martino, B.D., Venticinque, S. &Martone, A. (2009) 'OVerFA: a collaborative framework for the semantic annotation of documents and websites', Int. J. Web Grid Services, Vol. 5, No. 1, pp. 30-45.

[8] Panagis, Y., Sakkopoulos, E., Garofalakis, J. &Tsakalidis, A. (2006) 'Optimisation mechanism for web search results using topic knowledge', International Journal of Knowledge and Learning, Vol. 2, pp. 140-153.

[9] Bry, F., Koch, C., Furche, T., Schaffert, S., Badea, L. & Berger, S. (2005) 'Querying the Web Reconsidered: Design Principles for Versatile Web Query Languages', Int. J. Semantic Web Inf. Syst., Vol. 1, No. 2, pp. 1-21.

[10] Noah, S.A., Alhadi, A.C. &Zakaria, L.Q. (2005) 'A semantic retrieval of web documents using domain ontology', Int. J. Web Grid Services, Vol. 1, No. 2, pp. 151-164.

[11] Jin, Y., Xu, S., Decker, S., and Wiederhold, G.: Managing Web Sites with OntoWebber. In Proceedings of the 8th International Conference on Extending Database Technology, Prague, Czech Republic, Springer, 766-768, (2002).

[12] Vdovjak, R., Frasincar, F., Houben, G.J., Barna, P.: Engineering Semantic Web Information Systems in Hera. Journal of Web Engineering (JWE), Vol. 2, Nr. 1-2, Rinton Press, 3-26. (2003).