ANALYSIS OF RDBMS AND SEMANTIC WEB SEARCH IN UNIVERSITY SYSTEM

Vishal Jain¹ and S. V. A. V. Prasad²
¹Research Scholar, Computer Science and Engineering Department, Lingaya’s University, Faridabad, India
²Professor, Electronics and Communication Engineering Department, Lingaya’s University, Faridabad, India

ABSTRACT

To retrieve information from documents, there are many Information Retrieval (IR) techniques. Current IR techniques are not so advanced that they can be able to exploit semantic knowledge within documents and give precise results. IR technology is major factor responsible for handling annotations in Semantic Web (SW) languages. With the rate of growth of web and huge amount of information available on the web which may be in unstructured, semi structured or structured form, it has become increasingly difficult to identify the relevant pieces of information on the internet. In this paper, implementation of new proposed model, “Mining in Ontology with Multi Agent Systems” has been discussed and analyzed the model for comparative study in the search of RDBMS system and Ontology based system. In this model, the Semantic Web addresses the first part of this challenge by trying to make the data also machine understandable in the form of Ontology, while Multi-Agent addresses the second part by semi-automatically extracting the useful knowledge hidden in these data, and making it available.

KEYWORDS: Information Retrieval, Semantic Web, Ontology, Multi Agent Systems.

I. INTRODUCTION

The Semantic Web is an evolution of the current Web that represents information in a machine-readable format, while maintaining the human-friendly HTML representation and it avoid key word searching. Data handling and retrieving has an essential importance where the data size is larger than a certain amount. Storing transactional data in relational form and querying with Structured Query Language (SQL) is very preferable because of its tabular structure. Data may be also stored in an ontological form if it includes numerous semantic relations. This method is more suitable in order to infer information from relations. When transactional data contain many semantic relations inside as in our problem, it is not easy to decide the method of storing and querying. Improving Information retrieval by employing the use of Ontologies to overcome the limitations of syntactic search has been one of the inspirations since its emergence. Ontologies are important components of web-based applications. While the Web makes an increasing number of ontologies widely available for applications, how to discover ontologies in the Web becomes a more challenging issue. Existing approaches are mainly based on keywords and metadata information of Ontologies, rather than semantic entailments of ontologies. The current search engines performs search based on the syntax not on semantics. The keyword based search engines fails to understand and analyze the con-text in which keywords are used. The situation worsens when the search phrase is a combination of keywords. The quality of the results degrades with irrelevant results of documents which contains only the part of search phrase leaving the meaning aside. A semantic search –Tim Berner Lee’s unrealized dream - resolves this issue by analyzing the contexts and the relationships between the key words and thus producing a “quality high” and “quantity less” results. The basic idea of semantic web is to enrich the current Web with machine cognitive information about the semantics of information
The remaining sections of paper are as follows. Section 2 makes readers aware of related work. In this section, we have discussed about importance of Semantic Web over Current Web. Section 3 defines proposed solution for making information retrieval more relevant and fast. In section 4, proposed model has been discussed with the help of a case study on University System. In further part of this paper, discusses the development part of the model and shows the query result.

II. LITERATURE SURVEY

Keyword based Search engines are not able to provide relevant search result because they suffer from the fact that they do not know the meaning of the terms and expression used in the web pages and the relationship between them. This paper compares the semantic search performance of both keyword-based and semantic web based search engines [3, 4]. Initially, two keyword based search engines (Google and Yahoo) and three semantic search engines (Hakia, DuckDuckGo and Bing) are selected to compare their search performance on the basis of precision ratio and how they handle natural language queries. Ten queries, from various topics was run on each search engine, the first twenty documents on each retrieval output was classified as being “relevant” or “non-relevant”. Afterwards, precision ratios were calculated for the first 20 document retrieved to evaluate performance of these search engines. In our study it was found that relevant document retrieved by Bing is more (145 out of 200) than any other search engine. Hakia overall performance in terms of average percentage is lowest (56%). Figure 3 present overall graphical representation of mean precision ratio of search engine for first 20 documents. While figure 4 represents graphical representation for queries (number 1 to number 10).

![Figure 1: Precision ratio of search engines for first 20 documents [3, 4]](image-url)

A number of ontology reasoning systems have been developed for reasoning and querying the semantic web. Since they implement different reasoning algorithms and optimization techniques, they differ in a number of ways. Previous attempts at comparing performance of ontology reasoning systems have mainly considered performances of individual query requests. In this paper, authors, Chuik Lee, Sungchan Park, Dongjoo Lee, Jae-won Lee, Ok-Ran Jeong, Sang-goo Lee, presented the results of testing four of the most popular ontology reasoning systems on query sequences that reflect real world use cases. We believe that using query sequences is a more effective way to evaluate ontology reasoning systems [6]. Vipin Kumar.N, Archana P. Kumar, Kumar Abhishek presented a comparative study on SPARQL and SQL [7]. In this paper authors presented the benefits of SPARQL over SQL. Abrar Ahmad H, Muhammad Ruknuddin Ghalib [8] presented a system framework for building the semantic web support for intelligent search using RDF, ontology and SPARQL queries. J. Uma Maheswari, G. R. Karpugam, proposed a framework for Ontology based information retrieval model [9]. Venkata Sudhakara Reddy .Ch, Hemavathi .D proposed an new algorithm, Stemming Algorithm for semantic web search. In this paper, authors have presented a comparative study of RDBMS and Semantic Web search. Their experiments show that in the outcome of deployment of a new module, our incremental extraction approach minimizes the processing time by 92 percent as compared to a traditional pipeline approach. By using our methods to a corpus of 20 million biomedical abstracts, experiments indicate that the query performance is efficient for real-time applications. Experiments also uncovered that our approach achieves high quality extraction results [10]. Chintan Patel, Kaustubh Supekar, Yugyung Lee, E.K. Park, worked on OntoKhoj: A Semantic Web Portal for Ontology Searching, Ranking and Classification [11]. The methodology in developing OntoKhoj is based on algorithms used for searching, aggregating, ranking and classifying ontologies.
in Semantic Web. Jeff Z. Pan, Edward Thomas and Derek Sleeman, worked on Ontosearch2: Searching and Querying Web Ontologies [12]. ONTOSEARCH2 is able to reliably query large data sets faster than comparable database driven knowledge management systems. The recall and precision figures from the tests performed are encouraging but there are situations in which incomplete results can be returned. They have evaluated the ONTOSEARCH2 system using the Lehigh University Benchmark (LUBM) [13] to measure its performance on large data sets. They have run benchmarks using generated data sets representing 1, 5, 10, 20 and 50 universities, these are generated using the same seed and index values as used in [14]. Mariano Rodríguez-Muro, Roman Kontchakov and Michael Zakharyaschev [15], worked on OBDA with Ontop. Ontop (ontop.inf.unibz.it) is an ontology-based data access (OBDA) system implemented at the Free University of Bozen-Bolzano. Hyun Hee Kim, Soo Young Rieh, Tae Kyung Ahn, Woo Kwon Chang implemented an ontology-based knowledge management system which makes knowledge assets intelligently accessible to Korean financial firms [16]. This paper introduced the ontology model by illustrating the four components and reports on the implementation and evaluation of the information ontology for the searching of web resources. Based on the content analysis of eight international bank web sites, a pilot system of information ontology for web resources consisting of Publication, Project, Member, Person, and Organization was constructed. A comparative experiment was conducted to evaluate the information ontology for web resources. The performance of the ontology-based system was compared with that of web search engines in terms of relevance and search time. Ten researchers from an economic research institution were recruited in October, 2002, to conduct experiments in the researchers’ offices except on two occasions when both of which were conducted in the library. Before the participants conducted their searches, they were provided by the experimenter with a half-hour presentation about the system. The participants were given a list of twenty tasks and asked to perform both on a search engine of their choice and the ontology-based pilot system. The tasks included answering questions about the locating of scholarly literature, statistical data, conference information, news, people searches, and project searches.

Sareena Rose [17] presented a conceptual study on new searching techniques for Ontology based search engines. This paper focused on a study and improvisation of searching techniques used in semantic search engines keeping time complexity as the major factor. For analysis purpose, the keyword “Blood Cancer” was selected. After the processing of first and second phase the out-put was semantically expanded query combined with OR and AND operators. Li Ma, Yang Yang, Zhaoming Qiu, Guotong Xie, Yue Pan, Shengping Liu [18], build a complete benchmark for better evaluation of existing ontology systems, we extend the well-known Lehigh University Benchmark in terms of inference and scalability testing. The extended benchmark, named University Ontology Benchmark (UOBM), includes both OWL Lite and OWL DL ontologies covering a complete set of OWL Lite and DL constructs, respectively. Oracle 11g OWL is a scalable, efficient, forward chaining based reasoner that supports an expressive subset of OWL-DL. Oracle evaluated the inference performance of RDF and OWL Prime on the LUBM dataset [19]. Oracle evaluated this performance on database installed on machine “semperf3”. M. Rodriguez-Muro, R. Kontchakov, M. Zakharyaschev [20], here authors has provided the detailed information for the LUMMex20 experimentation. In this case there is no given SQL schema; and hence, no mappings. Ontop creates a relational schema together with mappings itself using the "Semantic Index" technique described in the paper. Here, a detailed comparison has been described on ontop-DB2, ontop-MySQL, OWLIM and Stardog with the use of LUMMex20 (Lehigh University Benchmark, Extended) dataset scenario.

Anarosa Alves, Franco Brandão, Viviane Torres da Silva, Carlos José Pereira de Lucena [21], presented a work on a model driven approach to develop multi-Agent Systems. In this paper, authors described a model driven approach to develop multi-agent systems that begins with an ontology based on the TAO conceptual framework. Quynh-Nhu Numi Tran, Graham Low [22], worked on MOBMAS: A methodology for ontology-based multi-agent systems development. In this authors, authors proposed a new framework and compared MOBMAS against sixteen well known methodologies: MaSE, MASSIVE, SODA, GAIA, MESSAGE, Methodology for BDI Agent, INGENIAS, Methodology with High-Level and Intermediate Levels, Methodology for Enterprise Integration, PROMETHEUS, PASSI, ADELFE, COMOMAS, MAS-Common, KADS, CASSIOPEIA and TROPOS. Pakornpong Pothipruk and Pattarachai Lalitrojwong [23], worked on an Ontology-based Multi-agent System for Matchmaking. Csongor Nyulas, Martin J. O’Connor,
Samson Tu1, David L. Buckeridge, Anna Akhmatovskiaia, Mark A. Musen [24], presented their work on An Ontology-Driven Framework for Deploying JADE Agent Systems. Authors described a methodology and suite of tools to support the modeling and deployment of agents on the JADE platform. These models are encoded using the Semantic Web ontology language OWL and provide detailed computer-interpretable specifications of agent behavior in a JADE system. Gajun Ganendran, Quynh-Nhu Tran, Pronab Ganguly, Pradeep Ray and Graham Low [25], proposed a methodology on An Ontology-driven Multi-agent approach for Healthcare. In this paper, authors described an ontology-driven multi-agent approach to the development of healthcare systems, with a case study in diabetes management. Wongthongtham, P., Chang, E., Dillon, T.S. [26], worked on Ontology-based Multi-agent system to Multi-site Software Development. Authors described software agent utilized ontology as its intelligence in MSSD and found that it has benefits. Ontology gives computers more knowledge that the agent can utilize. Maja Hadzic, Elizabeth Chang [27], worked on use of ontology-based multi-agent systems in the biomedical domain. Authors have shown how the ontologies can be used by multi-agent systems in intelligent information retrieval processes. The ontologies can be used to support some important processes involved in the information retrieval such as posing queries by the user, problem decomposition and task sharing among different agents, result sharing and analysis, information selection and integration, and structured presentation of the assembled information to the user.

III. PROPOSED SOLUTION

In order to implement semantic search, new Ontology based information retrieval model; “Mining in Ontology based Multi-Agent System for Information Retrieval” has been proposed in this paper. Proposed model has been implemented in University System.

IV. CASE STUDY

Details of the Proposed System Responsibilities:

1. Functionalities:
   - To search Universities/Colleges for taking admission in Ph.D, M.Tech, B.Tech, MCA, MBA, BCA, BBA
   - To Search particular Book, Journal, Magazine, Periodicals in College/University Library
   - To search person either Student, Teaching Staff, Non Teaching Staff in any mentioned universities
   - To search Learning Material/Study Material of any particular subject.

2. Details of Universities:
   - University Name
   - Type of Universities (Central Govt./State Govt./Deemed/Private)
   - Address
   - Affiliated Colleges
   - Approved Courses
     - Course Duration
     - Mode of Admission (Process)

3. Details of Library
   - Books
     - Title
     - Author
     - Publisher
     - Pages
     - Year of Publication
   - Journals
     - Title
     - Type (National/International)
     - Topic (Engineering, Humanities, Arts)
     - Issue No.
- Magazine
- Periodicals

4. Details of Person
- Person Name
- Category (Student/Staff)
- College/University Name
- Contact Details
  - Address
  - Contact Number
  - E-Mail Id

5. Details of Course Material
- Branch Name (CSE, ECE, EEE, IC, MAE, Civil)
- Subject Name
- Format (PPT/PDF)

6. Working Scenario of the Project
- Ontology can be used as database in the backend
- JSP/Java based web site to design the GUI, where user will give query
- JADE will retrieve data of query, fired by the user
- Mining can be used to extract pattern from the database

5. Technology Stack

1. **OWL** - Web Ontology Language is used here to represent the university domain including, Universities, Courses, People, Library and Materials.
2. **Protégé** - A free, open-source ontology editor and framework for building intelligent systems.
3. **OWLIM** - OWLIM is a family of semantic repositories, or RDF database management systems, with the following characteristics:
   - native RDF engines, implemented in Java
   - delivering full performance through both Sesame and Jena
   - robust support for the semantics of RDFS, OWL 2 RL and OWL 2 QL
   - best scalability, loading and query evaluation performance
4. **WEKA** - Apriori Algorithm - Weka is a collection of machine learning algorithms for solving real-world data mining problems. It is written in Java and runs on almost any platform. The Apriori Algorithm is an influential algorithm for mining frequent itemsets for boolean association rules.
5. **Empire** - Empire provides Java developers an easy way to integrate and start using SPARQL & RDF via JPA persistence annotations.
6. **JADE** - Java Agent Development Framework, (JADE) is used for the development of agents, implemented in Java. JADE system supports coordination between several agents FIPA and provides
a standard implementation of the communication language FIPA-ACL, which facilitates the communication between agents and allows the services detection of the system

V. DEVELOPMENT

5.1 Ontology Development

Figure 3: Ontology Entities

Figure 4: Ontology Classes

Figure 5: Data Properties
5.2 Agents Development
Figure 9: Add Agents

Figure 10: Load Agents

Figure 11: University Agent Details
5.3. Screenshots

1. Login

![Login](image)

**Figure 12:** Proposed University System

2. Search in University

![Search](image)

**Figure 13:** Search in University

3. Search Person in University

![Search Person](image)

**Figure 14:** Search Person in University

VI. ANALYSIS

Analysis of Proposed System by giving following queries:

- **Query 1: Universities within Chennai**
  Solution: University(x) ^ state(x, y) ^ address(y, “chennai”)

- **Query 2: Universities offering MCA course**
Solution: University(x) ^ offeredCourses(x,y) ^ courseDec(y, “MCA”)

**Query 3: Professor handling MBA course**
Solution: Person(x) ^ type(x, Staff) ^ associatedCourse(x, MBA)

**Query 4: Students taking BE course**
Solution: Person(x) ^ type(x, Student) ^ takingCourse(x, BE)

**Query 5: Books published in the year of 2014**
Solution: LibraryItem(x) ^ type(x, Book) ^ publishedYear(x, 2014)

**Query time in seconds**

| Table 1: Comparison between RDBMS and Semantic Web Search |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| **Data**       | **System**      | **Q1**          | **Q2**          | **Q3**          | **Q4**          | **Q5**          |
| 103K            | SW              | 22.820          | 2.604           | 7.105           | 0.287           | 3.879           |
|                 | RDBMS           | 28.410          | 3.014           | 9.215           | 0.216           | 4.109           |
| 2.8M            | SW              | 176.972         | 8.312           | 27.789          | 1.531           | 16.889          |
|                 | RDBMS           | 260.350         | 9.300           | 34.021          | 0.816           | 18.313          |
| 13.9M           | SW              | 308.198         | 65.713          | 5.105           | 3.968           | 111.368         |
|                 | RDBMS           | 1382.824        | 91.452          | 176.516         | 5.368           | 133.887         |

**Figure 15: Comparison between RDBMS and Semantic Web Search (103K Data)**

**Figure 16: Comparison between RDBMS and Semantic Web Search (2.8M Data)**

**Figure 17: Comparison between RDBMS and Semantic Web Search (13.9M Data)**
ACKNOWLEDGEMENTS

I, Vishal Jain, would like to give my thanks to Mr. Anandam Raj for help me in programming for the model and sincere thanks to Dr. M. N. Hoda, Director, Bharati Vidyapeeth’s Institute of Computer Applications and Management (BVICAM), New Delhi, for giving me opportunity to do Ph.D from Linagay’s University, Faridabad.

VII. CONCLUSIONS

A new model has been defined with the use of Mining in Ontology with Multi Agent system for information retrieval, whereas ontology can be used as a repository, mining for data extraction and multi agent system can be used for data representation. This paper will also helpful to other researchers, who would like to do work in this area. In this paper, Analysis of proposed model has been shown that Semantic Web search may be better than RDBMS search.

REFERENCES

[3]. Yogender Singh Negi, Suresh Kumar, “A Comparative Analysis of Keyword Based and Semantic Based Search Engines”, Ambedkar Institute of Advanced communication Technologies & Research, Delhi, India.
[15]. Mariano Rodr guez-Muro, Roman Kontchakov and Michael Zakharyaschev, “OBDAwith Ontop”, Faculty of Computer Science, Free University of Bozen-Bolzano, Italy.
[16]. Hyun Hee Kim, Soo Young Rieh, Tae Kyoung Ahn, Woo Kwon Chang, “Implementing an Ontology-Based Knowledge Management System in the Korean Financial Firm Environment”, Korea Research Foundation Grant (KRF-2002-005-B20006)
[18]. Li Ma, Yang Yang, Zhaoming Qiu, Guotong Xie, Yue Pan, Shengping Liu, “Towards A Complete OWL Ontology Benchmark”, IBM China Research Laboratory, Building 19, Zhongguancun Software Park, ShangDi, Beijing, P.R. China.
[28]. Maja Hadzic, Elizabeth Chang, “Use Of Ontology-Based Multi-Agent Systems in the Biomedical Domain”, Curtin University of Technology, School of Information Systems Perth, Western Australia, 6845, Australia.

AUTHORS

Vishal Jain has completed his M.Tech (CSE) from USIT, Guru Gobind Singh Indraprastha University, Delhi and doing Ph.D from Computer Science and Engineering Department, Lingaya’s University, Faridabad. Presently he is working as Assistant Professor in Bharati Vidyapeeth’s Institute of Computer Applications and Management, (BVICAM), New Delhi. His research area includes Web Technology, Semantic Web and Information Retrieval. He is also associated with CSI, ISTE.

S. V. A. V. Prasad has completed M.Tech., Ph.D. Presently working as professor and Dean (Academic Affairs), Dean (R&D; IC), Dean School of Electrical Sciences. Sir has actively participated and organized many refresher courses, seminars, workshops on ISO, ROHS, Component Technology, WEEE, Organizational Methods, Time Study, Productivity enhancement and product feasibility etc. Sir has developed various products like 15 MHz dual Oscilloscope, High Voltage Tester, VHF Wattmeter, Standard Signal, Generator with AM/FM Modulator, Wireless Becom, High power audio Amplifier, Wireless microphone and many more in the span of 25 years (1981 – 2007). Sir has awarded for excellence in R&D in year 1999, 2004 and National Quality Award during the Year 1999, 2000, 2004 and 2006. Sir is Fellow member of IEEE and life member of ISTE, IETE, and Society of Audio & Video System. Sir has published more than 90 research papers in various National & International conferences and journals. Sir research area includes Wireless Communication, Satellite Communication & Acoustic, Antenna, Neural Networks, and Artificial Intelligence.

APPENDIX - A
Code Snippet
1. University Ontology
package Vishal;
import java.util.Collection;
import org.protege.owl.codegeneration.WrappedIndividual;
import org.semanticweb.owlapi.model.OWLNamedIndividual;
import org.semanticweb.owlapi.model.OWLOntology;

/**
 * <p>
 * Generated by Protege (http://protege.stanford.edu). <br>
 * Source Class: University <br>
 * @version generated on Thu Aug 28 20:39:13 IST 2014 by Charvi <br>
 */

public interface University extends WrappedIndividual {

    /* ***************************************************
     * Property http://localhost/Institute#centretHead
     */
    /**
     * Gets all property values for the centretHead property.<p>
     * @returns a collection of values for the centretHead property.
     */
    Collection<? extends Teacher> getCentretHead();

    /**
     * Checks if the class has a centretHead property value.<p>
     * @return true if there is a centretHead property value.
     */
    boolean hasCentretHead();

    /**
     * Adds a centretHead property value.<p>
     *
     * @param newCentretHead the centretHead property value to be added
     */
    void addCentretHead(Teacher newCentretHead);

    /**
     * Removes a centretHead property value.<p>
     *
     * @param oldCentretHead the centretHead property value to be removed.
     */
    void removeCentretHead(Teacher oldCentretHead);

    /* ***************************************************
     * Property http://localhost/Institute#listedCourse
     */
    /**
     * Gets all property values for the listedCourse property.<p>
     * @returns a collection of values for the listedCourse property.
     */
    Collection<? extends Course_Name> getCourse();

    /**
     * Checks if the class has a listedCourse property value.<p>
     * @return true if there is a listedCourse property value.
     */
    boolean hasCourse();

    /**
     * Adds a listedCourse property value.<p>
     *
* @param newListedCourse the listedCourse property value to be added
*/
void addListedCourse(Course_Name newListedCourse);
/**
 * Removes a listedCourse property value.<p>
 *
 * @param oldListedCourse the listedCourse property value to be removed.
 */
void removeListedCourse(Course_Name oldListedCourse);
/* *************************************************
**
* Common interfaces
*/
OWLNamedIndividual getOwlIndividual();
OWLOntology getOwlOntology();
void delete();
}

2. N-Triple Code

```
<http://localhost/Institute#Vishal_Kumar> <http://localhost/Institute#city> "Ghaziapur city"^^<http://www.w3.org/2001/XMLSchema#string>.<br />
<http://localhost/Institute#Vishal_Jain > <http://localhost/Institute#hasQualification> <http://localhost/Institute#BSc>.<br />
<http://localhost/Institute#Vishal_Jain > <http://localhost/Institute#hasResearchInterest> <http://localhost/Institute#Digital_library>.<br />
<http://localhost/Institute#Vishal_Jain > <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2002/07/owl#NamedIndividual>.<br />
<http://localhost/Institute#Vishal_Jain > <http://localhost/Institute#ZIP> "233305"^^<http://www.w3.org/2001/XMLSchema#string>.<br />
<http://localhost/Institute#Vishal_Jain > <http://localhost/Institute#hasDiscipline> <http://localhost/Institute#Library_and_Information_Science>.<br />
<http://localhost/Institute#Vishal_Jain > <http://localhost/Institute#affiliatedUnit> <http://localhost/Institute#DRTC>.<br />
<http://localhost/Institute#Vishal_Jain > <http://localhost/Institute#dateOfBirth> "1984-08-02"^^<http://www.w3.org/2001/XMLSchema#date>.<br />
<http://localhost/Institute#Vishal_Jain > <http://localhost/Institute#belongsToCentre> <http://localhost/Institute#ISIBC>.<br />
<http://localhost/Institute#Vishal_Jain > <http://localhost/Institute#email> "Vishal@drtc.isibang.ac.in"^^<http://www.w3.org/2001/XMLSchema#string>.<br />
<http://localhost/Institute#Vishal_Jain > <http://localhost/Institute#country> "India"^^<http://www.w3.org/2001/XMLSchema#string>.<br />
<http://localhost/Institute#Vishal_Jain > <http://localhost/Institute#phoneNo> "9986180163"^^<http://www.w3.org/2001/XMLSchema#string>.<br />
```
3. OWL Code

<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:p1="http://localhost/Institute#"
  xmlns:protege="http://protege.stanford.edu/plugins/owl/protege#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns="http://localhost/Institute#"
  xmlns:swrl="http://www.w3.org/2003/11/swrl#"
  xmlns:swrlb="http://www.w3.org/2003/11/swrlb#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xml:base="http://localhost/Institute">
<owl:Ontology rdf:about=""/>
<rdfs:Datatype rdf:about="http://www.w3.org/2001/XMLSchema#date"/>
<owl:Class rdf:ID="Research_Student">
  <owl:disjointWith>
    <owl:Class rdf:ID="Undergraduate_Student"/>
  </owl:disjointWith>
  <rdfs:subClassOf>
    <owl:Class rdf:ID="Student"/>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty>
        <owl:ObjectProperty rdf:ID="hasResearchInterest"/>
      </owl:onProperty>
      <owl:someValuesFrom>
        <owl:Class rdf:ID="Research_Interest"/>
      </owl:someValuesFrom>
      <owl:Restriction>
        <owl:subClassOf>
          <owl:Class rdf:ID="MStat"/>
        </owl:subClassOf>
      </owl:Restriction>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="MStat">
  <rdfs:subClassOf>
    <owl:Class rdf:ID="Postgraduate_Course"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="Magazine">
  <rdfs:subClassOf>
    <owl:Class rdf:ID="Periodical"/>
  </rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="Article">
  <rdfs:subClassOf>
    <owl:Class rdf:ID="Publication"/>
  </rdfs:subClassOf>
  <owl:disjointWith>
    <owl:Class rdf:ID="Thesis"/>
  </owl:disjointWith>
  <owl:disjointWith>
    <owl:Class rdf:ID="Proceedings"/>
  </owl:disjointWith>
  <owl:disjointWith>
    <owl:Class rdf:about="#Periodical"/>
  </owl:disjointWith>
  <owl:disjointWith>
    <owl:Class rdf:ID="Book"/>
  </owl:disjointWith>
</owl:Class>
<owl:Class rdf:about="#Undergraduate_Student">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:someValuesFrom>
        <owl:Class rdf:ID="Undergraduate_Course"/>
      </owl:someValuesFrom>
      <owl:onProperty>
4. WEKA Configuration

4.1 Connect

import weka.core.*;

/**
 * Generates code based on the provided arguments, which consist of
 * classname of a scheme and its options (outputs it to stdout).
 * The generated code instantiates the scheme and sets its options.
 * The classname of the generated code is <code>OptionsTest</code>.
 *
 * @author FracPete (fracpete at waikato dot ac dot nz)
 */
public class OptionsToCode {

    /**
     * Generates the code and outputs it on stdout. E.g.:<p/>
     * <code>java OptionsToCode weka.classifiers.functions.SMO -K
    "weka.classifiers.functions.supportVector.RBFKernel" &gt; OptionsTest.java</code>
     *
     */
    public static void main(String[] args) throws Exception {
        // output usage
        if (args.length == 0) {
            System.err.println("Usage: java OptionsToCode <classname> [options] > OptionsTest.java\n");
            System.exit(1);
        }

        // instantiate scheme
        String classname = args[0];
        args[0] = "";
        Object scheme = Class.forName(classname).newInstance();
        if (scheme instanceof OptionHandler)
            ((OptionHandler) scheme).setOptions(args);
    }
// generate Java code
StringBuffer buf = new StringBuffer();
buf.append("public class OptionsTest {\n");
buf.append("\n");
buf.append("  public static void main(String[] args) throws Exception {\n");
buf.append("    // create new instance of scheme\n");
buf.append("    " + classname + " scheme = new " + classname + "();\n");
if (scheme instanceof OptionHandler) {
    OptionHandler handler = (OptionHandler) scheme;
    buf.append("      \n"); buf.append("    // set options\n");
    buf.append("    scheme.setOptions(weka.core.Utils.splitOptions(" +
    Utilities.backQuoteChars(Utilities.joinOptions(handler.getOptions())) + "));\n");
    buf.append("  }\n" );
}    buf.append("}\n"); // output Java code
System.out.println(buf.toString());
}

4.2 Add arbitrary weights

Import weka.core.converters.ConverterUtils.DataSource;
import weka.core.converters.XRFFSaver;
import weka.core.Instances;
import java.io.File;
/**
 * Loads file "args[0]", sets class if necessary (in that case the last
 * attribute), adds some test weights and saves it as XRFF file
 * under "args[1]". E.g.: <br/>
 * AddWeights anneal.arff anneal.xrff.gz
 * @author FracPete (fracpete at waikato dot ac dot nz)
 */
public class AddWeights {
    public static void main(String[] args) throws Exception {
        // load data
        DataSource source = new DataSource(args[0]);
        Instances data = source.getDataSet();
        if (data.classIndex() == -1)
            data.setClassIndex(data.numAttributes() - 1);
        // set weights
        double factor = 0.5 / (double) data.numInstances();
        for (int i = 0; i < data.numInstances(); i++) {
            data.instance(i).setWeight(0.5 + factor * i);
        }
        // save data
        XRFFSaver saver = new XRFFSaver();
        saver.setFile(new File(args[1]));
        saver.setInstances(data);
        saver.writeBatch();
    }
}