

An Outline Implementation for Habitual Production of Ontology from RDF to Medical Data

Yan Zhigang

Inner Mongolia Normal University, China
3762533081@qq.com

Abstract

Information on the web is extremely growing in current years with a quicker velocity. This enormous or capacious data has driven intricate troubles for information recovery and information organization. As the data resides in a network with numerous forms, knowledge management on the web is a challenging task. Here the novel 'Semantic Web' concept may be used for understanding the web contents by the machine to offer intelligent services efficiently with a meaningful knowledge representation. The information recovery in the conventional web source is centered on 'page positioning' strategies, though in the semantic web the information recovery forms depend on the 'idea based learning'. The proposed work is gone for the improvement of another system for programmed age of cosmology and RDF to some continuous system information, removed from numerous storehouses by following their URLs and Text Documents. The enhanced altered ordering method is connected for the cosmology age and turtle documentation is utilized for RDF documentation. A program is composed for approving the extricated information from different archives by expelling undesirable information and considering just the report segment of the website page.

Keywords: *Semantic web, Resource description framework, Ontology, Improved inverted indexing, Knowledge management*

1. Introduction

Internet (WWW) is considered as a worldwide data storehouse that recognizes archives and other web assets by Uniform Resource Locators, interlinked by hypertext joins. Web indexes are utilized to recover the data from the web. Information overburden is the most concerning issue nowadays for the current framework. The advancement of the web incorporates the web renditions of web 1.0, 2.0 and so on. In this arrangement, the web rendition 3.0 is alluded to as semantic web [1] is advanced as an information administration bolster over the globe. Web indexes ought to be improved with semantic web abilities that break down site page content and give more important outcomes relating to the client question. Semantic web guidelines incorporate Resource Description Framework (RDF), web metaphysics, RDF Schema and Rule Interchange Format (RIF) for taking care of information. RDF gives a calculated depiction of data for speaking to the web assets like Turtle punctuation, N-Triples and so on. RDF portrays information on the Web in chart shape [2]. Ontologies comprise the limited arrangement of terms, connections, limitations and sayings [3]. Ontologies have ended up being valuable for compelling learning displaying and data recovery. The remaining paper is arranged as follows:

Article history:

Received (December 21, 2018), Review Result (January 18, 2019), Accepted (March 4, 2019)

In Section 2 the related work is presented. The proposed work and its methodology are discussed in Sections 3 & 4. The results are presented in Section 5. Comparative analysis is explained in Section 6 and Conclusions are given in Section 7.

2. Literature review

M.S.P.Babu et.al [4] provided an overview of some of the semantic search engines that yield unique search experiences for users. Wilkinson et.al [5] proposed an information retrieval system using document structure.

Amel Grissa Touzi et.al [6] suggested the Fuzzy Ontology of Data Mining (FODM) for processing automated generation of ontologies in the domain of data mining.

Amira Aloui et.al [7] implemented a plugin named “FO-FQ Tab plug-in,” which can be integrated with a protégé editor for building the fuzzy ontologies from large databases. To overcome the drawbacks of the existing system for accessing the related science information,

M.S.P.Babu et.al [8] proposed a new framework for the automatic generation of ontology and RDF for real-time web data.

Tahani Alsubait et.al [9] developed the e-learning suite, with the set of questions designed using ontological representation.

A.H.M.Rupasingha et.al [10] suggested that the performance of the ontology generation is always dependent on the specificity of the terms.

Seongwook Youn et.al [11] discussed the pros and cons of tools like protégé 2000, OilEd, Apollo, OntoLingua, Onto Edit, webODE, KAON, ICOM, DEO, webOnto that is used for ontology creation. Kgotatso Desmond Mogotlane et.al [12] presented a comparative study of plugins of protege tools like DB2OWL and Data Master.

3. Proposed methodology

Semantic web capacities like RDF and metaphysics are connected to advance learning. The proposed work is a usage of the system proposed by the creators [8]. The system is planned regarding the semantic web Stack. It is completed in two stages, specifically the Data extraction stage and Data portrayal stage. Web scratching is performed utilizing HTML parsing procedure in information extraction stage by giving example seek inquiry as a contribution to numerous stores. DOM parsing and HTML parsing procedures are connected to approve the information recovered from different storehouses by considering just the record area of the site page. Extensible Markup Language (XML) is the base for the semantic web portrayals; the approved data is changed over into semi-organized documentations by utilizing XSD statement from DOM tree and go as a contribution for the following layers of the proposed structure. XML documentation is given as a contribution to the information portrayal stage. RDF documentation is produced and spoken to in a graphical frame utilizing the Graphviz device. A printed portrayal of the RDF diagram is given utilizing Turtle, the Terse RDF Triple Language. Enhanced Inverted Indexing system is connected for the ontological portrayal of words by barring the stop words.

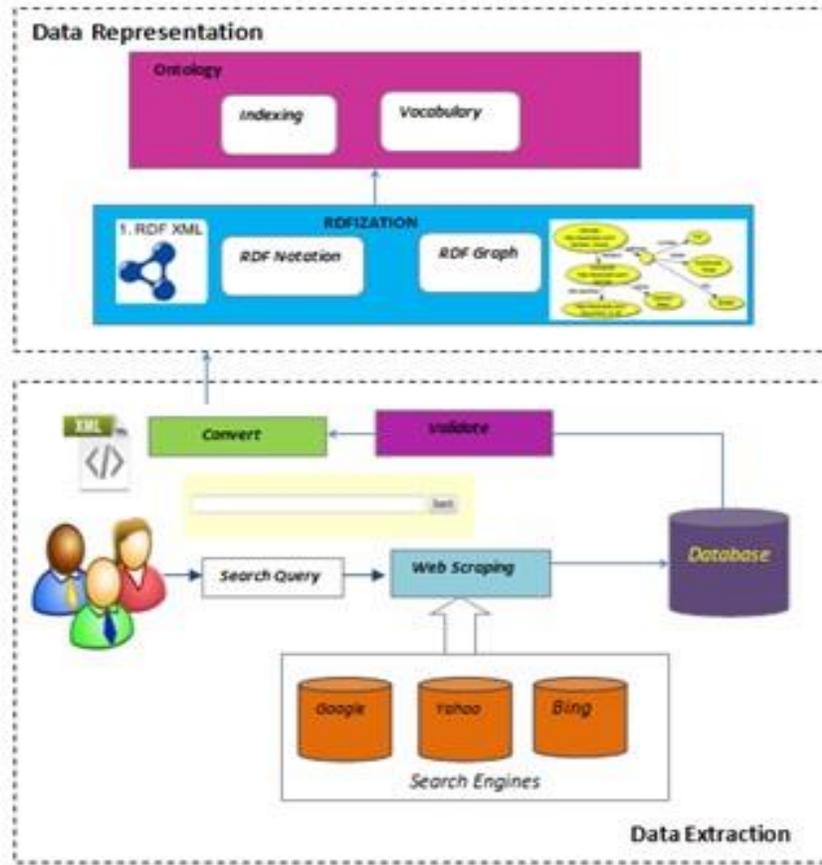


Figure 1. Proposed framework

4. Methodology

Implementation of the framework proposed in Section III will be carried out in two phases namely data extraction and data representation phases. The details are given below.

Phase 1: Data extraction

The data extraction phase is sub-divided into three steps namely web scraping, data validation, XML Conversion.

Step 1: Web Scraping

Web scraping also is referred to as screen scraping or Web harvesting is used to fetch and extract the data from a web document using HTML parsing techniques. Here Web pages are crawled and the content of the Web page is extracted, reformatted and stored in a database. The HTML parsing technique is used for scraping data from web documents.

Step 2: Data Validation

In the Data Validation Step, the data collected from step 1 is validated using HTML and DOM parsing techniques. Here unwanted data is removed and the necessary portion of URLs is retained. The validated data is stored in a database by considering Document Object Model (DOM). The Data validation process is explained in [Figure 2].

Step 3: RDF-ization

RDF is the basic building block in the semantic web, promoting conceptual modeling of web data [13]. It identifies things using Uniform Resource Identifiers (URIs) and describes

resources with properties and values. Each tuple in a relational database is decomposed as RDF triples.

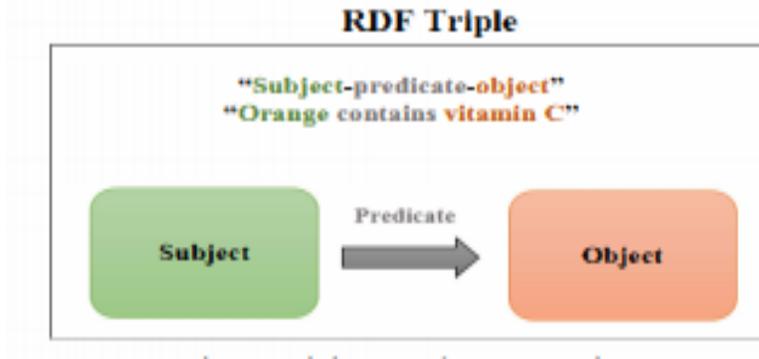


Figure 2. RDF triple

Decomposition of tuple creates a new blank node, corresponding to the row, and a new triple set is obtained. RDF can be viewed as a directed graph with labeled nodes and arcs. The triple is represented as a <subject, predicate, object> format by exploring the relationship among the nodes [14]. A node can be a URI reference, literal, or a blank node. The predefined data type used in RDF is “RDF: XML Literal,” for embedding XML in RDF.

5. Result analysis

A sample search query is given as an input and web scraping results are shown in [Figure 3] as follows,

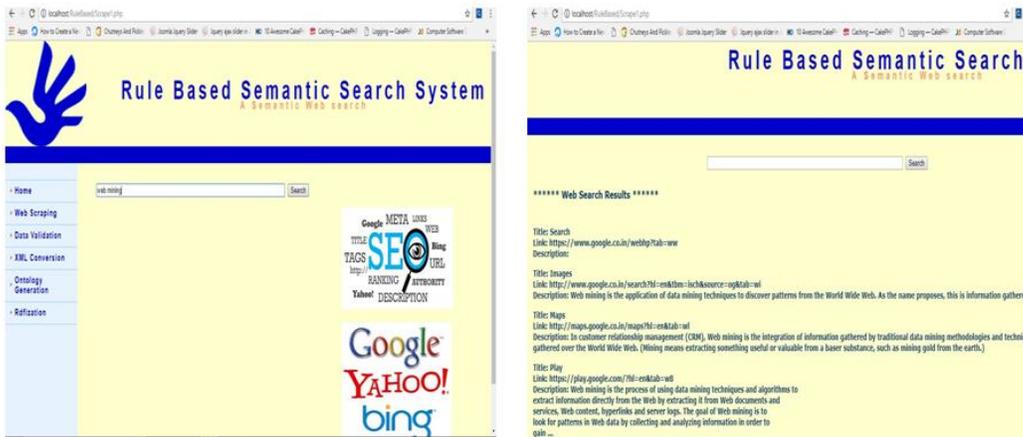
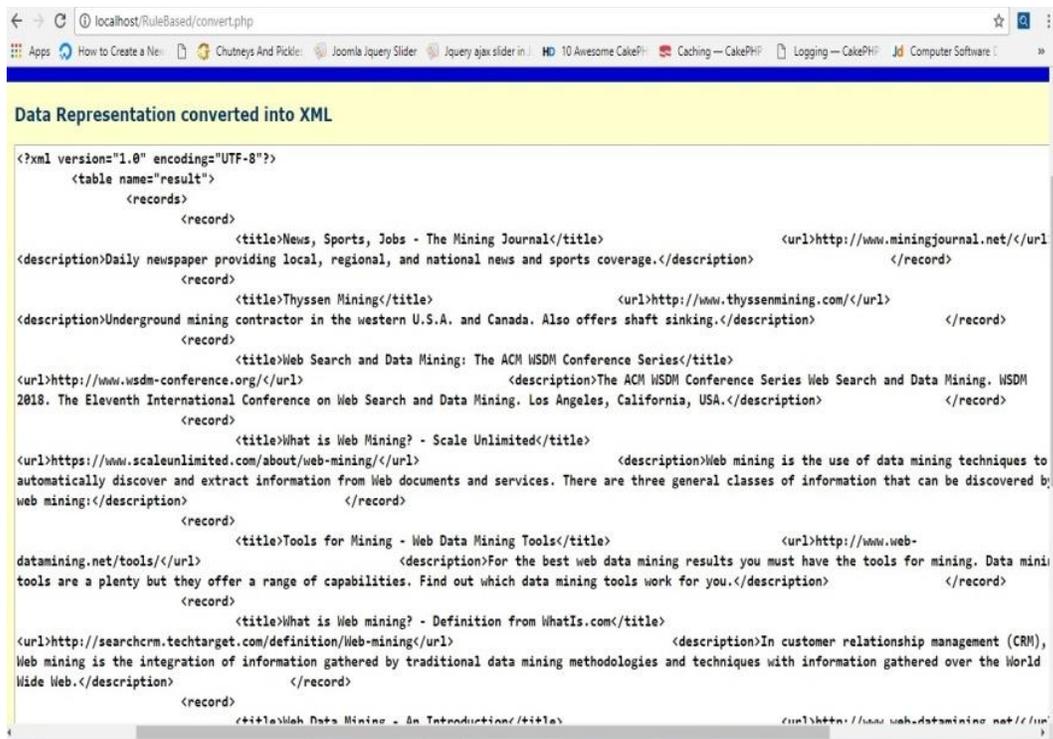


Figure 3. Web scraping results

Web scraping performance is evaluated by considering the following parameters like database size, number of URLs extracted, the number of keywords, response time, latency time. Scraped data from multiple repositories is given as an input to the data validation step. The validated data is obtained as an output to the data validation process by applying HTML and DOM parsing techniques. Data validation considers only the document section of a web page. The validated data stored in the database is converted into the XML notations by applying XSD declaration from the DOM tree.

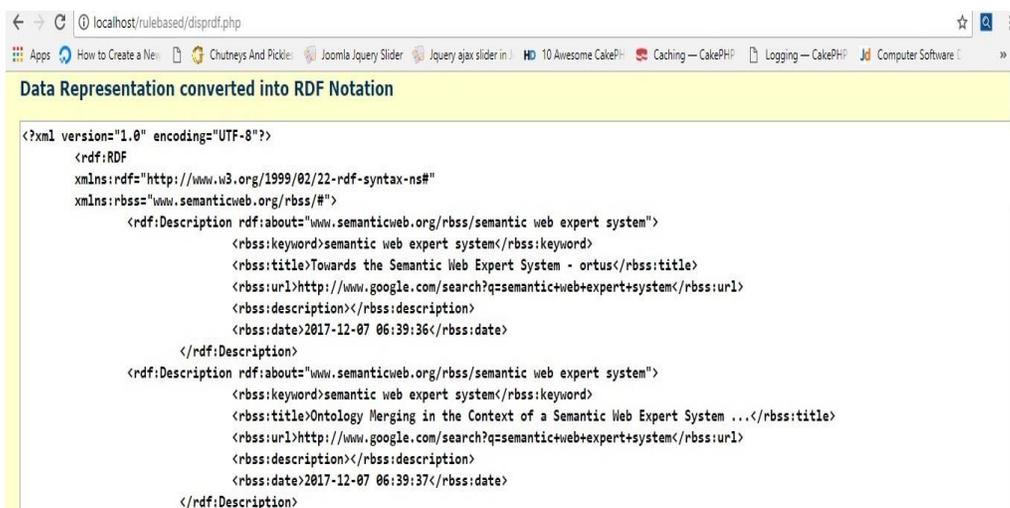


The screenshot shows a web browser window with the address bar at localhost/RuleBased/convert.php. The page title is "Data Representation converted into XML". The main content area displays an XML document with the following structure:

```
<?xml version="1.0" encoding="UTF-8"?>
  <table name="result">
    <records>
      <record>
        <title>News, Sports, Jobs - The Mining Journal</title>
        <url>http://www.miningjournal.net/</url>
        <description>Daily newspaper providing local, regional, and national news and sports coverage.</description>
      </record>
      <record>
        <title>Thyssen Mining</title>
        <url>http://www.thyssenmining.com/</url>
        <description>Underground mining contractor in the western U.S.A. and Canada. Also offers shaft sinking.</description>
      </record>
      <record>
        <title>Web Search and Data Mining: The ACM WSDM Conference Series</title>
        <url>http://www.wsdm-conference.org/</url>
        <description>The ACM WSDM Conference Series Web Search and Data Mining. WSDM 2018. The Eleventh International Conference on Web Search and Data Mining. Los Angeles, California, USA.</description>
      </record>
      <record>
        <title>What is Web Mining? - Scale Unlimited</title>
        <url>https://www.scaleunlimited.com/about/web-mining/</url>
        <description>Web mining is the use of data mining techniques to automatically discover and extract information from Web documents and services. There are three general classes of information that can be discovered by web mining.</description>
      </record>
      <record>
        <title>Tools for Mining - Web Data Mining Tools</title>
        <url>http://www.web-datamining.net/tools/</url>
        <description>For the best web data mining results you must have the tools for mining. Data mining tools are a plenty but they offer a range of capabilities. Find out which data mining tools work for you.</description>
      </record>
      <record>
        <title>What is Web mining? - Definition from WhatIs.com</title>
        <url>http://searchrm.techtarget.com/definition/Web-mining/</url>
        <description>In customer relationship management (CRM), Web mining is the integration of information gathered by traditional data mining methodologies and techniques with information gathered over the World Wide Web.</description>
      </record>
      <record>
        <title>Web Data Mining - An Introduction</title>
        <url>http://www.web-datamining.net/</url>
      </record>
    </records>
  </table>
</pre>
```

Figure 4. XML conversion

RDF is a recommended standard of the World Wide Web Consortium (W3C). RDF generation for sample relation named “testrdf” which has an attribute as <name, description, freq> is considered. The “testrdf” represents the relation name is considered as a class in an RDF graph and has a set of three nodes that are connecting the testrdf in a depth-wise manner represents the tuple of a relation. Ontology generation for the data obtained from multiple repositories as well as the text file.



The screenshot shows a web browser window with the address bar at localhost/RuleBased/disprdf.php. The page title is "Data Representation converted into RDF Notation". The main content area displays an RDF document in turtle form:

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rbss="http://www.semanticweb.org/rbss/#">
  <rdf:Description rdf:about="http://www.semanticweb.org/rbss/semantic web expert system">
    <rbss:keyword>semantic web expert system</rbss:keyword>
    <rbss:title>Towards the Semantic Web Expert System - ortus</rbss:title>
    <rbss:url>http://www.google.com/search?q=semantic+web+expert+system</rbss:url>
    <rbss:description></rbss:description>
    <rbss:date>2017-12-07 06:39:36</rbss:date>
  </rdf:Description>
  <rdf:Description rdf:about="http://www.semanticweb.org/rbss/semantic web expert system">
    <rbss:keyword>semantic web expert system</rbss:keyword>
    <rbss:title>Ontology Merging in the Context of a Semantic Web Expert System ...</rbss:title>
    <rbss:url>http://www.google.com/search?q=semantic+web+expert+system</rbss:url>
    <rbss:description></rbss:description>
    <rbss:date>2017-12-07 06:39:37</rbss:date>
  </rdf:Description>
</rdf:RDF>
</pre>
```

Figure 5. RDF-ization results in turtle form

Improved inverted indexing technique is applied for extracting the words with their frequencies discarding the stop words, in the order of highest precedence. The highest frequency word is considered as a frequent search term for rule framing using description logic. The rule mapping is done for the efficient retrieval operation which will be future work. RDF is a recommended standard of the World Wide Web Consortium (W3C). RDF representation of data in turtle form is shown in [Figure 5].

6. Conclusion

The advancement of information portrayal on the web is taking numerous varieties step by step which prompts the tremendous significance of the data portrayal and information stockpiling to help the top of the line data recovery frameworks, Hence the exploration on the semantic web is stimulated and numerous types of research introduced an approach to speak to information utilizing the semantic web. The current framework gathers the information from web crawlers is showing normal execution in recovery and this manner, the proposed system enhances the execution by applying semantic web abilities RDF-ization and metaphysics age which gives quicker recovery. Metaphysics age utilizing devices like Protege, Ontorat, RDF2OWL and so on. is manual handling of making classes and subclasses. Enhanced rearranged ordering strategy is adjusted for the cosmology age. Surmising guideline encircling can be connected to any sort of question given by the client and the database can be secured by applying proficient cryptography calculation which is done in our future work.

References

- [1] Sareh Aghaei, Mohammad Ali Nematbakhsh, and Hadi Khosravi Farsani, "Evolution of the world wide Web: From Web 1.0 To Web 4.0," *International Journal of Web & Semantic Technology (IJWesT)*, January, vol.3, no.1, pp.20-30, (2012)
- [2] Abdeslem DENNAI, Sidi Mohammed BENSLIMANE, "Semantic indexing of web documents based on domain ontology," *International Journal of Information Technology and Computer Science*, vol.2, pp.1-11, (2015)
- [3] Seema Redekar, Vishal Chekkala, Siddhapa Gouda, and Swapnil Yalgude, "Web search engine using ontology learning," *International Journal of Innovative Research in Computer and Communication Engineering*, March vol.5, no.3, pp.15-25, (2017)
- [4] G Sudeepthi, G Anuradha, and M Surendra Prasad Babu, "A survey on semantic web search engine," *International Journal of Computer Science Issues*, 3 IJCSI, vol.9, no.2, pp.241-245, (2012)
- [5] R. Wilkinson, "Effective retrieval of structured documents," *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (S.-V. New York, Ed.)*, pp.311-317, (1994)
- [6] Touzi Ag, H. Massoud, and A. Ayadi., "Automatic ontology generation for data mining using FCA and clustering," *Computer Science*, arxiv.org, no.1311.1764, (2013).
- [7] Amira Aloui, ENIT, Tunis, and Tunisia, "A fuzzy ontology-based platform for flexible querying," *International Journal of Service Science, Management, Engineering, and Technology*, July-September, vol.6, no.3, pp.12-26, (2015)
- [8] Babu, M Surendra Prasad, and S. Govathoti, "A semantic model for building integrated ontology databases," *IEEE International Conference on Software Engineering & Service Science IEEE*, (2017)
- [9] Tahani Alsubait, Bijan Parsia, and Ulrike Sattler, "Ontology-based multiple choice question generation," *Kunstl Intell*, vol.30, pp.183-188, (2016)
- [10] Youn, Seongwook, and D. Mcleod, "Ontology development tools for ontology-based knowledge management," *Encyclopedia of E-Commerce, E-Government, and Mobile Commerce*, (2006)

- [11] Kgotatso Desmond Mogotlane and Jean Vincent Fonou-Dombeu, "Automatic conversion of relational databases into ontologies," International Journal of Web & Semantic Technology (IJWesT) October, vol.7, no.3/4, pp.21-40, **(2016)**
- [12] Fateme Abiri, Mohsen Kahani, and Fatane Zarinkalam, "An entity based RDF indexing schema using Hadoop and HBase," 4th IEEE International Conference on Computer and Knowledge Engineering, **(2011)**
- [13] Faizan Shaikh, Usman A. Siddiqui, Iram Shahzadi, and Syed Uami, Zubair A. Shaik, "SWISE: Semantic web based intelligent search engine," 978-1-4244- 8003-6/10 IEEE, **(2010)**
- [14] Gopal Pandey, "The semantic web: an introduction and issues," International Journal of Engineering Research and Applications, Jan-Feb, vol.2, Issue 1, pp.780-786, **(2012)**

This page is empty by intention.