CSE 6708 - Semantic Web

Assignment 2

Report on Paper Presentation

Paper Name Source Code Plagiarism Detection Method Using Protégé Built Ontologies

Samidhya Sarker
Student No. 1018052049
Group-2

September 18, 2019

Contents

List of Figures

1

Paper Summery

Extending the conclusion started in Assignment 1[1], we can see that the paper has some shortcomings. The main promising application of semantic web technologies is to create a knowledge base aiding information sciences. Although semantic indexing can aid traditional applications.

[1] `https://semantic-web.netlify.com/report/report.pdf`

2

Analysis of References present in Given Paper

2.1 List of Citation

1. M. K. Shenoy, K. C. Shet and U. D. Acharya. (2012, May). Semantic Plagiarism Detection System Using Ontology Mapping. Advanced Computing: An International Journal 3(3). [1]

2. The Protégé Ontology Editor and Knowledge Acquisition System. [2] (2013, July 1).

3. T. Bray, J. Paoli, C. M. Sperberg-McQueen, E. Maler and F. Yergeau. (2004, February 4). Extensible Markup Language (XML) 1.0. W3C Recommendation . Third Edition. [3]

4. F. Manola and E. Miller (2004, February 10). RDF Primer. W3C Recommendation. [4]

5. S. Harris, A. Seaborne. (2013, March 21). SPARQL 1.1 Query Language. W3C Recommendation. Available: [5]

6. J. Bao, D. Calvanese, B. C. Grau, et al. (2012, December 11). OWL 2 Web Ontology Language. W3C Recommendation. Second Edition. [6]

7. E Akin, Object Oriented Programming, Houston: Rice University Publishing House, 2001, pp. 33-34.

[1] http://airccse.org/journal/acij/papers/0512acij06.pdf
[2] http://protege.stanford.edu/
[3] http://www.w3.org/TR/2004/REC-xml20040204/
[4] http://www.w3.org/TR/2004/REC-rdfprimer-20040210/
[5] http://www.w3.org/TR/2013/RECsparql11-query-20130321/
[6] http://www.w3.org/TR/owl2-overview/

8. S. Kara, O. Alan and O. Sabuncu, "An ontology-based retrieval system using semantic indexing", Information Systems, vol. 37, no. 4, pp. 294–305, June 2012.

9. Pseudocode Standards, California Polytechnic State University Website. [7]

10. C. Liu, H. Wang, Y. Yu and L. Xu, "Towards Efficient SPARQL Query Processing on RDF Data", Tsinghua Science & Technology, vol. 15, no. 6, pp. 613–622, December 2010.

11. I. Ivan and C. Boja, Metode Statistice in analiza software. Bucharest: ASE Publishing House, 2004, pp. 218-224. Informatica Economică vol. 17, no. 3/2013

12. S. Russel and P. Norving, Artificial Intelligence: A Modern Approach (2nd edition). New Jersey: Pearson Education Inc., 2003, pp. 350-352.

13. P. Durusau, S. Newcomb and R. Barta (2007, November). Topic Maps Reference Model. International Organization for Standardization. [8]

14. D. Newman, T. Baldwin, L. Cavedon and E. Huang, "Visualizing search results and document collections using topic maps", Web Semantics: Science, Services and Agents on the World Wide Web, vol. 8, no. 2-3, pp 169–175, July 2010.

15. A. Hatzigaidas, A. Papastergiou, G. Tryfon and D. Maritsa, "Topic Map Existing Tools: A Brief Review", in Proc. The International Conference on Theory and Applications of Mathematics and Informatics, Thessaloniki, Greece, 2004, pp 185-201

2.2 Classification of Given Citation

- Reference 1 is a paper published in an international Journal.
- Reference 2 is the homepage of Protege editor.
- Reference 3 is World Wide Web Consortium (W3C) recommendation of XML.
- Reference 4 is World Wide Web Consortium (W3C) primer for RDF.
- Reference 5 is World Wide Web Consortium (W3C) recommendation for SPARQL.

[7] `http://users.csc.calpoly.edu/~jdalbey/SWE/pdl_std.html(2013,July1)`
[8] `http://www.isotopicmaps.org/TMRM/TMRM-7.0/tmrm7.pdf`

- Reference 6 is Wide Web Consortium (W3C) recommendation for OWL.
- Reference 7 is a book on OOP.
- Reference 8 is a Masters thesis from Middle East University (Turkey) on semantic indexing.
- Reference 9 is a dead link to a standard on pseudocode writing. Currently can be found on scribd [9].
- Reference 10 is an IEEE paper on SPARQL Query processing.
- Reference 11 is an book on Statistical methods in software analysis written in Romanian.
- Reference 12 is the most widespread book on AI.
- Reference 13 is the ISO reference model for topic maps.
- Reference 14 is a article from Elsevier on how to visualize data using topic maps.
- Reference 15 is an paper published in an international conference on tools required to implement topic maps.

2.2.1 Comment

As we can see that references 3-6, 13 are specifications from W3C, ISO. Reference 2 is a website. And References 11, 12 are excerpts from books. So, we shall analyze refences 1, 8, 10, 14, 15.

2.3 Reference 1: Semantic Plagiarism Detection System Using Ontology Mapping

In this paper, the authors discuss the applications of using Semantic web ontologies to detect plagiarism in text documents. The proposal is to train an ontology learner by feeding an ontology mapping algorithm (a) Main copies and (b) Probable plagiarizations.

The authors also experimented with a prototype TAO (Transitioning Applications to Ontologies) project but the implementation details are not given.

[9] `https://www.scribd.com/document/47856615/PSEUDOCODE-STANDARD`

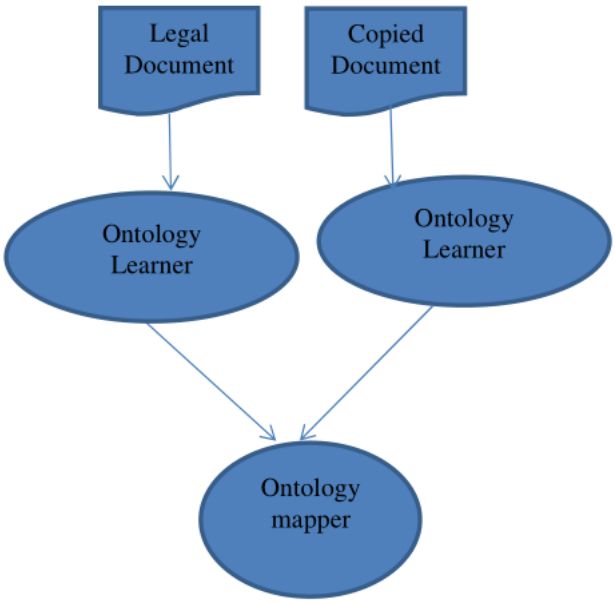

Figure 2.1: Architecture of the proposed detection system.

2.4 Reference 8: An ontology-based retrieval system using semantic indexing

This Masters thesis, can be divided into three parts. In chapter 2, Background information, the authors discuss general information science with an emphasis on information retrieval (IR). They also discuss optimizing the information retrieval or searching process implementing indexing, ranking etc. They also discusss evaluation metrices. Finally the concept of Semantic Web is introduced and the application to improve IR is discussed.

In chapter 3, various approaches including traditional, semantic approaches are discussed. The authors discuss semantic indexing and their proposed models.

In chapter 4, the authors conceptualize their own semantic retrieval process.

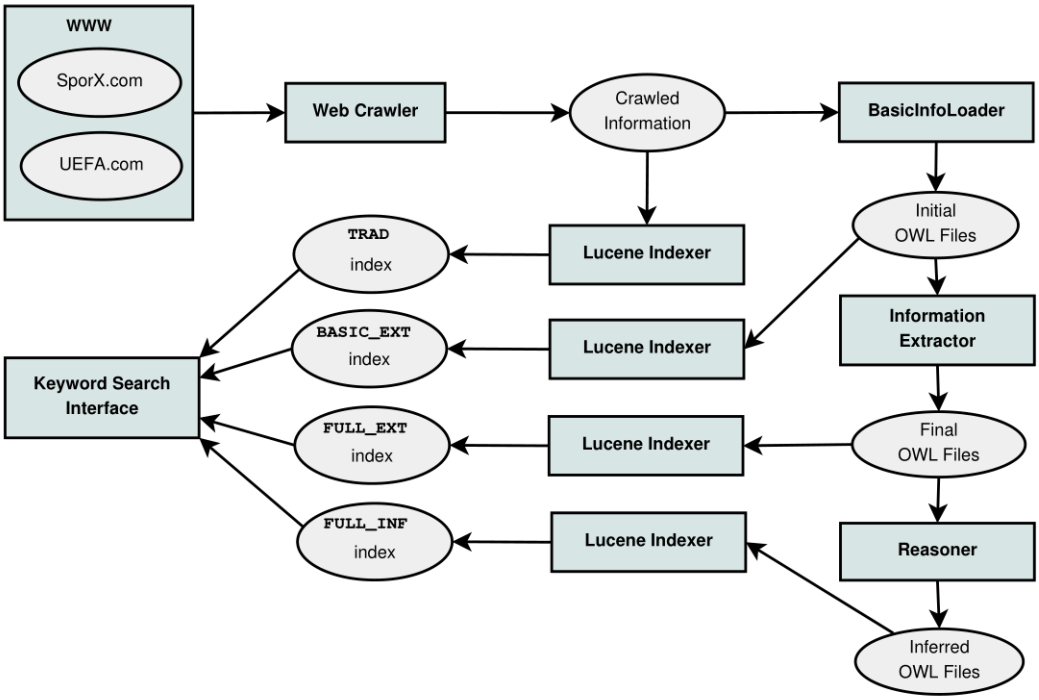

Figure 2.2: System design of the proposed semantic retrieval system system.

So, in this system, information is gathered from `www.uefa.com` and `www.sporx.com` which are not semantically readable. These information are then fed into Lucene Indexer and owl ontologies are formed from the gathered xml. In this way, both a indexed search engine and semantically vaiable Knowledge base is created which are interlinked. Semantic ruleset is created and a reasoner is used. So, information can be looked up efficently and accurately.

2.5 Reference 10: Towards Efficient SPARQL Query Processing on RDF Data

In this original research paper, the authors talk about SPARQL query optimiztion. The main claims of the authors are, in some RDF data storage systems (eg. APACHE Jena, SOR), SPARQL queries are often translated into SQL statements. But as SPARQL triples does not translate well enough into SQL statements, many self-joins and table joins are required which causes slow-downs.

2.6 Reference 14: Visualizing search results and document collections using topic maps

3

Analyis of Citations of Given Paper

4

Future Work

Bibliography

[1] Michel Goossens, Frank Mittelbach, and Alexander Samarin. *The LaTeX Companion*. Addison-Wesley, Reading, Massachusetts, 1993.

5

Appendix