SEMANTIC PLAGIARISM DETECTION SYSTEM USING ONTOLOGY MAPPING

Manjula Shenoy K[1] Dr. K.C.Shet[2] U.Dinesh Acharya[1]

[1]Department of Computer Engineering, Manipal University, Manipal
manju.shenoy@manipal.edu,dinesh.acharya@manipal.edu

2Department of Computer Engineering,National Institute of Technology, Suratkal
kcshet@rediffmail.com

ABSTRACT

Plagiarism detection can play an important role in detecting stealing of original ideas in papers, journals and internet web sites. Checking these manually is simply impossible nowadays due to existence of large digital repository. Ontology is a way of describing documents semantics. Ontology mapping can resolve semantic heterogeneity in documents. Our paper proposes an automatic system for semantic plagiarism detection based on ontology mapping.

KEYWORDS

*OWL, SEMANTICS, ONTOLOGY, ONTOLOGY MATCHING, ONTOLOGY MAPPING*

1. INTRODUCTION

Today's huge volume of digital information is both advantageous as well as disadvantageous too. Advantageous in the sense that we can get each and every information on the net freely available for reference and hence searching time for required information has reduced a lot. Also it makes more acquaintance of information to people. Disadvantage in the sense that finding duplicate documents has become difficult. Manual detection of them takes more time and hence we seek the help of plagiarism detection systems.

Ontologies is a computational model of some domain of the world[7]. This describes the semantics of the terms used in the domain. It is often captured in the form of DAG (directed acyclic graph) consisting of a finite set of concepts + properties + relationships. Given two ontologies checking how similar they are is ontology matching. Figure 1. denotes two different ontologies and their partial mapping.

2. PREVIOUS SYSTEMS

There have been many plagiarism detection systems in literature[2]. Few approaches use sentence based matching of documents with the legal and original copies[1]. But these systems could not be applied to large volume of data that is existing. Also if sentences are modified using synonyms detection bypasses the documents. Few systems use word net and other similar resources to detect copying. Some systems used water marking[1]. Our idea is to use ontologies and matching to detect plagiarism
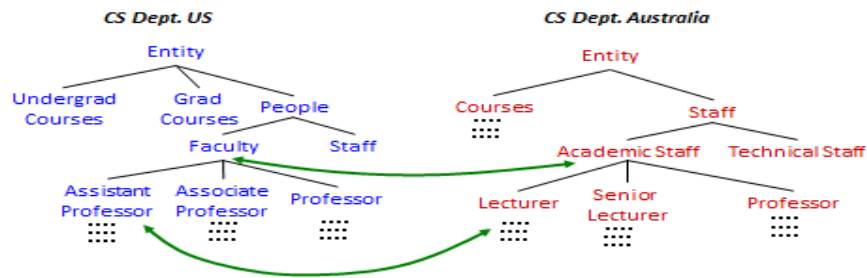
Figure 1. Ontology mapping Example

3. PROPOSED SYSTEM

Here we take original or legal document and so called copied document as input. We learn ontology from both these documents and then apply an ontology mapping algorithm to detect correspondences between the various entities. Ideally we should get 100% correspondence. But we may take the decision if mapping is above some threshold. The architecture of the proposed system is as shown in Figure 2.
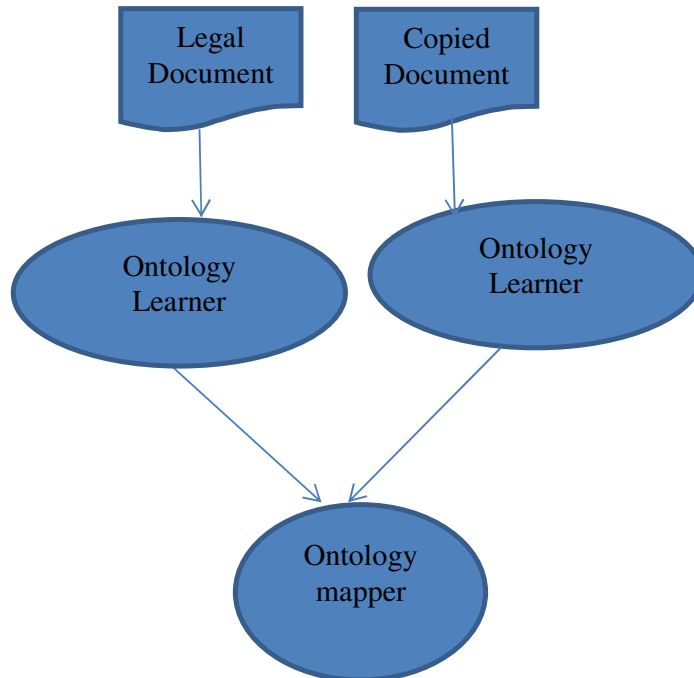

Figure 2. Architecture of detection system

Ontology learner user needs to learn ontologies from blocks of text and express it in the form of one of the representation language for ontology OWL. Outputs of both learners in OWL are fed

to ontology mapper to map concepts, properties and relationships. Based upon number of concepts, properties and relations in each ontology, the mapping discovered is expressed as a fraction of whole. If this fraction is above threshold we can assume documents are similar in meaning. Otherwise not.

4. PROTOTYPE EXPERIMENT

We used a demo version of OntoGen ontology learner which is a part of TAO project to generate ontology. And from this we exported OWL file and gave this to our own ontology mapper software. The results were generated for various documents containing small chunk of texts which are dissimilar in sentence wise but mean the same.

5. CONCLUSION

From the above experiment we concluded that we can build semantic plagiarism detection systems using ontology learner and mapper. The accuracy of the system is dependent on accuracy of ontology learner and mapper. Also the performance of the system too depends on the learner and mapper.

ACKNOWLEDGEMENTS

REFERENCES

[1]   Antonio Si, Hong Va Leong,Rynson W.H.Lau,"CHECK:A document plagiarism detection system" SAC'97Proceedings of the 1997 ACM symposium on Applied Computing.

[2]   Lancaster Thomas,"Effective and Efficient plagiarism detection" PhD thesis

[3]   T.R. Gruber, *Toward Principles for the Design of Ontologies Used for Knowledge Sharing* International Journal of Human-Computer studies, Vol. 43,Issue 5-6,pp907-928,Nov 1995.

[4]   T.R.Gruber, *A translation approach to portable ontology specifications* , Knowledge Acquisition,Vol. 5,Issue2 pp199-220,1993.

[5]   Natalya F. Noy , Deborah L. McGuinness , *Ontology Development 101: A guide to creating your first ontology* [Online]. Available: http://www.protege.stanford.edu/publications/Ontologydevelopment/ontology101-noy-McGuinness.html.

[6]   Tim Berners-Lee, James Hendler , Ora Lasilla , *The Semantic Web*, Scienticfic American, 2001.

[7]   Manjula Shenoy.K, K.C.Shet, U.DineshAcharya,." A simple ontology mapping method" in Proceedings of 5th International MultiConference on Intelligent Systems, Sustainable, New and Renewable Energy Technology and Nano Technology(IISN-2011), Klawad, Haryana, India. Febrauary 2011

.

Authors

Manjula Shenoy.K is currently working as a   Assistant professor (Sel.Grade) at CSE   Department, Manipal Institute of Technology, Manipal   University,Manipal.

Dr. K.C. Shet is a Professor in Department of Computer Engineering, National Institute of Technology ,Suratkal. He has several Journal and Conference papers to his credit.
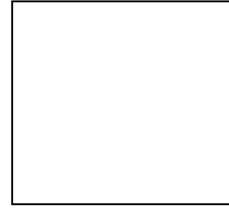
Dr. U.Dinesh Acharya is Professor and Head of the department of Computer Science and Engineering, Manipal Institute of Technology , Manipal University,Manipal. He has several Journal and Conference papers to his credit.