International Journal of Modern Trends in Engineering and Research
www.ijmter.com
e-ISSN No.:2349-9745, Date: 2-4 July, 2015

Evaluation of Knowledge Generation from Electronic Textbook

Rasika A. Rode[1], Prof. Gitanjali Shinde [2]

[1] *Dept. of Computer Engineering, Sinhgad Technical Education Society's, Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India, rasikarode222@gmail.com*
[2] *Assistant Professor, Dept. of Computer Engineering, Sinhgad Technical Education Society's, Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India, gr83gita@gmail.com*

Abstract—We know that organizations are growing because of tremendous increase in demand of knowledge. So from huge amount of data which is available in different organization regarding to technical system we need such kind of data which helps in preparation of any task which will help in saving time.

For example : if we want to take class regarding to seminar then knowledge which we are discovering from this large data will help us for preparation and this knowledge is called as relationships which are called as is-A, Next, Prerequisite, Part-Of.

Keywords- Domain engineering, Ontology design, Semantic engineering, Knowledge acquisition, Learning object.

I. INTRODUCTION

It is acting as tool for generating knowledge from electronic textbook. Electronic textbook is acting as PDF. So, from this tool we are getting relationships in this textbook with the help of natural language processing techniques which provides area of computer sciences for interacting between computers and human languages but while interacting there is large amount of data and hence we need to identify the relation which are acting as knowledge and if we store knowledge our cost and time will save.

For this purpose we provide processed electronic textbook which should be start from content, index and it should not split.

Ex: Contents
1. Introduction
1.1 Introduction to java
1.2 Inheritance
2. Class

So from above example there can be number of contents. Hence if we find relationships between them it helps teacher for preparation like I have to master in Introduction then I can teach the class which is called as prerequisite relation. Also introduction and introduction to java contain same words and hence it exhibit isA relation, after Introduction I must teach class and hence it is called as Next relation.

II. LITERATURE SURVEY

Erik Duval states that since there is need to maintain metadata which is acting as data about data because without the metadata learning content cannot be reused and it will difficult or impossible to identify and retrieve them and hence there is need of automatic generating of learning object metadata[1].

Hence Amal zouaq says that to use this metadata ontology should be extended and updated for discovering new entity and reuse those entities and hence manual creation for those entities are not sufficient because of lack of flexibility and hence for adding those entities knowledge is build as project[2].
But knowledge is build as preen coded hence Bernado Magnini explain that for this ontology learning is necessary and hence methods and evaluations are must. But building ontology from sentences are similar to building array of methods by using natural language processing techniques and hence acting as major area of research[3].
Hence further improvement is done by Johanna Volker added that to produce ontological outputs in terms of concepts, attributes and hierarchical relationship among those concepts and attributes[4].
So Armando Stellato further proceed with the the concept of discovering processes and resources in which he has developed wiktinary and from that he developed information regarding to words, sentences and large amount of relationship that are exists between these words[5].

III. SYSTEM ARCHITECTURE

In this architecture user will enter Domain name which will be of PDF then the request will go to whois server which will check whether it is valid and will fetch PDF from the database. Once the PDF is uploaded token extraction is takes place from which number of relations are extracted and in order to identify the identity between them TF/IDF algorithm is applied and then the ontology creation is takes place and result is reported to user.

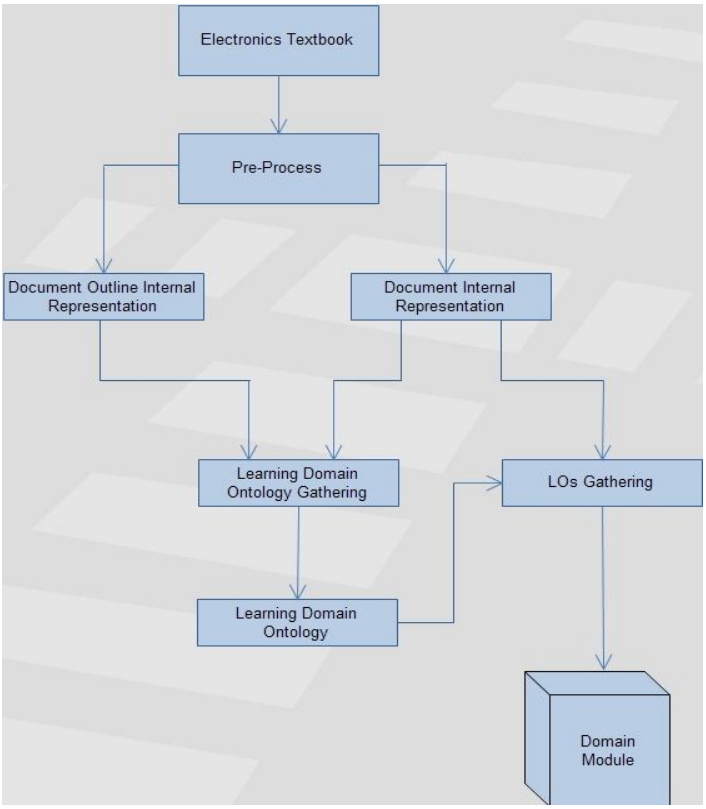

Figure 1. System architecture

IV. MATHEMATICAL MODEL

Suppose we have term frequency table for a collection consisting of only two documents are listed below consist of term and term count then calculation of tf-idf for term "this" is

Table 1. Doccument1

term	Term count
this	1
Is	1
another	2
example	3

Table 2. Doccument2

term	Term count
term	Term count
this	1
Is	1
A	2
example	1

Suppose we have term frequency table for a collection consisting of only two documents are listed below consist of term and term count then calculation of tf-idf for term "this" is
So,

$$idf(this,D) = \log\{n/d£D:t£D\} \quad (1)$$

$$idf(this,D) = \log(2/2) = 0 \quad (2)$$

So tf-idf is zero for "this" term and hence it is appear in all document.

1] Identify the isA relationships

$$I = \{i1, i2, i3 \ldots iN\} \quad (3)$$

Where 'I' is main set of isA relationships.

2] Identify the partOf relationships

$$P=\{p1,p2,p3,\ldots,pN\} \quad (4)$$

Where 'P' is main set of part of relationships.

3] Identify the next relationships

$$N = \{n1, n2, n3, \ldots\ldots..,nN\} \quad (5)$$

V. SECTIONS

Suppose we have term frequency table for a collection consisting of only two documents are listed below consist of term and term count then calculation of tf-idf for term "this" is

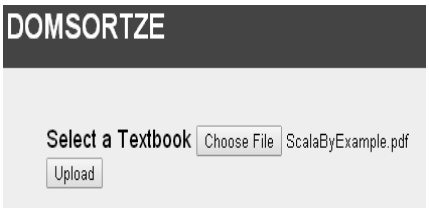

Figure 2. uploading file

Figure 3. Processing data

From fig. 2 and fig 3 first we have uploaded PDF named as "scale by example file" and after clicking on the process outline we will get the following result in the form of the relationship.

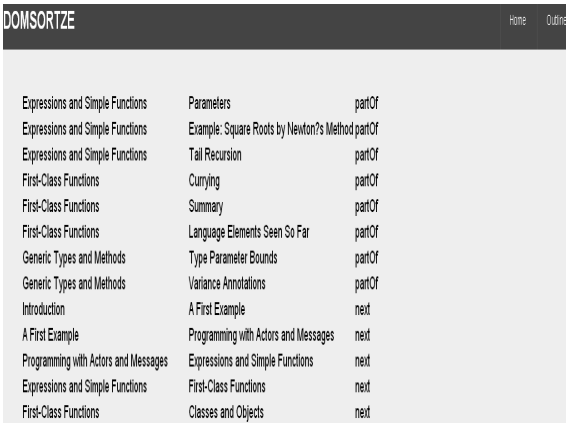

Figure 4. Generating relationship

As shown in fig. 4 first we have uploaded PDF named as "scale by example file" and after clicking on the process outline we will get the following result in the form of the relationship includes isA and next.

CONCLUSION AND FUTURE WORK

In this case we have uploaded electronic textbook which is called as processed electronic textbook and then we have found the relationships like isA, partOf, Prerequisite, Next between them and because of those relationship it saves the considerable time for preparation of task regarding to organization by using the ontology and natural language processing techniques. But problem with this is that when one PDF is uploaded the other PDF get overwritten and hence at a time only one

PDF can be supported which must be start from Content, Index and must be end with the Bibliography like terms.

Hence Future work is on identifying more number of relationship so that time complexity can be reduced considerably and also It should support the multiple PDF which can be called as preen coded knowledge so that new user need not to identify new relationship he can make use of already found relationships.

REFERENCES

[1] S. M. Metev and V. P. Veiko, Laser Assisted Microtechnology, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.

[2] Bowman, M., Debray, S. K., and Peterson, L. L. 1993. Reasoning about naming systems. .

[3] Ding, W. and Marchionini, G. 1997 A Study on Video Browsing Strategies. Technical Report. University of Maryland at College Park.

[4] M.A. Hearst, "Automatic Acquisition of Hyponyms from Large Text Corpora," Proc. 14th Conf Computational Linguistics (COLING '92), pp. 539-545, 1992.

[5] K.T. Frantzi, S. Ananiadou, and J. Tsujii, "The C-Value/NC-Value Method of Automatic Recognition for Multi-Word Terms," Proc. Second European Conf. Research and Advanced Technology for Digital Libraries (ECDL '98), pp. 585-604, 1998.

[6] M. Meire, X. Ochoa, and E. Duval, "SAmgI: Automatic Metadata Generation v2.0," Proc. World Conf. Educational Multimedia, Hypermedia, and Telecomm. (ED-MEDIA '07), pp. 1195-1204, June2007.

[7] A. Zouaq and R. Nkambou, "Evaluating the Generation of Domain Ontologies in the Knowledge Puzzle Project," IEEE Trans. Knowledge and Data Eng., vol. 21, no. 11, pp. 1559-1572, Nov. 2009.

[8] M. Larran˜ aga, A. Conde, I. Calvo, A. Arruarte, and J.A. Elorriaga,"Evaluating the Automatic Extraction of Learning Objects from Electronic Textbooks Using Erauzont," Proc. 11th Int'l Conf. Intelligent Tutoring Systems (ITS '12), pp. 655-656, 2012

[9] Semi-Automatic Ontology Development: Processes and Resources, M.T. Pazienza and A. Stellato, eds., IGI Global, 2012.

[10] Ontology Learning from Text: Methods, Applications, and Evaluation, P. Buitelaar, P. Cimiano, and B. Magnini, eds., IOS Press, 2005.

[11] T. Leidig, "L3-Towards an Open Learning Environment," ACM J. Educational Resources in Computing, vol. 1, no. 1, pp. 5-11, 2001.

[12] K. Verbert, "An Architecture and Framework for Flexible Reuse of Learning Object Components," PhD dissertation, Faculteit Ingenieurswetenschappen, Katholieke Univ. Leuven, Feb. 2008.

[13] B. Liu, C.W. Chin, and H.T. Ng, "Mining Topic-Specific Concepts and Definitions on the Web," Proc. 12th Int'l Conf. World Wide Web (WWW), pp. 251-260, 2003.

[14] M. Larran˜ aga, J.A. Elorriaga, and A. Arruarte,"A Heuristic NLP Based Approach for Getting Didactic Resources from Electronic Documents," Proc. European Conf. Technology Enhanced Learning (EC-TEL '08), pp. 197-202, 2008.

