# Mining the knowledge Generated from Electronic Textbook by using Learning domain ontology Algorithm

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

**Abstract** —*Since Most of the organizations are large and hence they requires knowledge. But gathering of knowledge is not an easy task because it requires domain topic from which knowledge will be gathered and hence, for gathering of knowledge we have to provide some resource and that resource is nothing but our PDF.After uploading PDF we are gathering knowledge from it in the form of relationships. So, why there is need to identify such relationships.*
*Ex: Suppose we want to attend any conference then what will be our requirement that which topic will be choose? in which topic we should master so that we can teach another topic based on that topic? after teaching that topic on which topic we should jump so these requirements can be satisfied by using IsA, PartOf, Prerequisite, Next.*

**Keywords**- *Ontology design; Domain engineering; knowledge acquisition; semantic engineering; process and resource*

## I. INTRODUCTION

Mining the knowledge generated from electronic textbook by using LDO algorithm means that since,we are finding out relationships that exists between the topic and subtopic so it is called as Learning Domain Ontology because Ontology is defined as finding the relationships between words and entities.hence, we can say that it is concept of ontology design which provides structures from unstructured text.Since generation of knowledge called as the domain module is very difficult in which we need to select the domain topic and relationships that are exist among those domain module.Learning Domain ontology algorithm is domain independent and only processed electronic textbook is required and the knowledge genraed from it.Electronic textbook is considering as a source to generate knowledge providing behavior average teachers while preparing for their seminars.

In Knowledge acquisition process rules and ontologies required for a knowledge-based system. Different kinds of phrase was first used in collaboration and conjunction with expert systems which is used to to describe the initial tasks associated with developing an expert system, includes namely finding different domain experts and processs of interviewing domain experts and gathering of their knowledge which includes relations via rules, objects.

In semantic engineering primary Web resource are mined that is being mined is called as record of the requests made by client to a Web site,which are most often collected in a Web server log. z The real performance of those who use these resources may use additional structure than the one in structure mining. z It is also used for static site improvement by identifying navigational pattern of the user inside a Website z It is used to combine Web usage mining with content and structure analysis for identifying the relationships in order to "make sense" of observing frequent relations and the pages on which these relations are count.

## II. LITERATURE SURVEY

T. Leidig [8]says that extraction of useful information from web requires some techniques which has been applied but such techniques requires keyword extraction and hence such kind of extraction can be achieved by identifying relations so, finding out relations includes ontologyand hence,learning is necessary.

So[7]Paul Buitelaar says that for learning ontology extraction method is applied in which extraction of PDF can be done in form of text for which NLP techniques has been used.

Hence, A. Conde [5]says that if there is learning system which provides similar functionality according to the reuse of learning object the user will be satisfied and this reuse will be more feasible and hence,by using application program interface and simple indexing techniques it can be achieved.

Due to this problem [4]Sophia Annidout proceed that there is need of multiple words which is to be extractedwhich can be achieved by using C value and NC value method.

Hence, X. Ochoa [2]explained that there is need of semantic web for generation of knowledge so,knowledge is extracted by semi automatic evaluation which provides approach for for ontology evaluation which further reflects use of evaluation,all application in ontology based on this evaluation and all prerequrements for this.

## III. PRAPOSED SYSTEM

Here we provide electronic textbook and preprocessing is done by using basic structural analysis in which all Next relationships are identified and sequential analysis in which all partOf relationships are identified and heuristics are applied for identifying isA.Whatever the relationships that we have gathered are generated by using the Learning domain ontology algorithm that we have presented below:

1.Basic structural analysis:In this case PDF is provided and the relationships that are exist between topic and topic is identified which is called as Next relationship.

2.Basic sequential analysis:In this case PDF is provided and the relations that are exist between topic and subtopic is identified called as partOf relationships.

3.Heuristics:In this case all kinds of isA relationships are identified by using different Heuristics like isAcronym,isPossesiveGenuitive.In which we have included different heuristics techniques:

isAcronym:If subtopic contain acronym of Topic then it is a isA relationship

Ex:Topic is=>this is a extended markup language

Subtopic is=>this is XML. Then it shows that it is a isA relationship.

isPossesiveGenuitive: If subtopic contain acronym+of preposition then it is a isA relationship.

Ex: Topic is=>this is a extended markup language

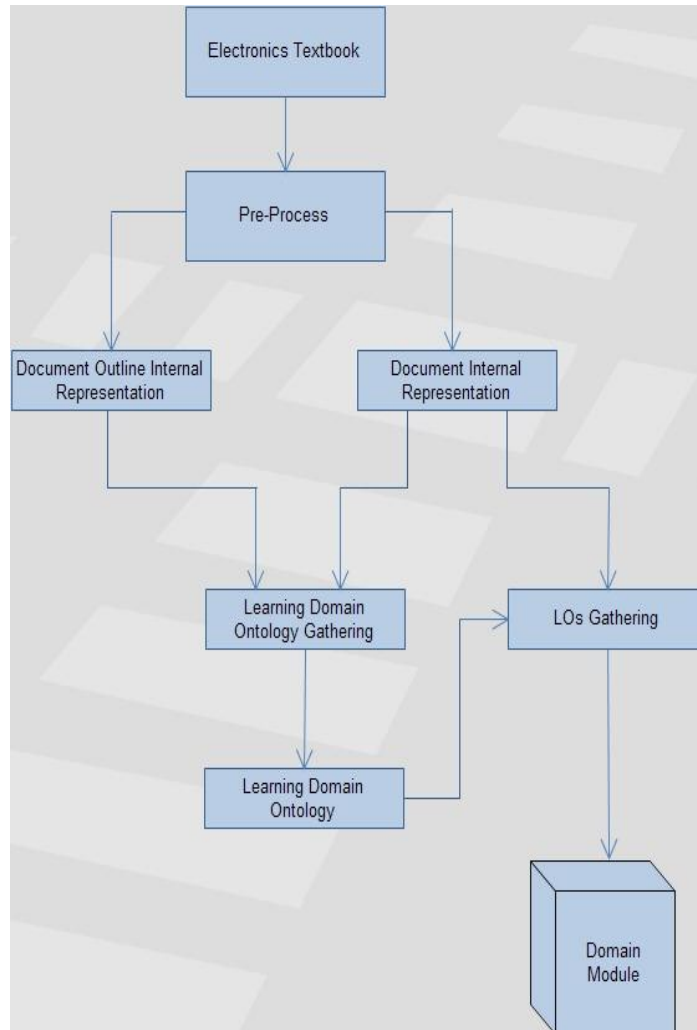Subtopic is=>this is XML of discovered by scientist then it shows that it is a isA relationship.



*Figure 1. Knowledge Mining process*

## IV.     PRAPOSED LEARNING DOMAIN ONTOLOGY ALGORITHM

Mining For identifying IsA, partOf, and Next relation

Step1 :

Initially all relationships that exist between topic and topic are called as Next relationship.

```
for(int i=1;i<topicList.size();i++)
{
        RelationshipBinder rb=new RelationshipBinder(topicList.get(i-1), topicList.get(i),Relationship.Next);
}
```

Step2:

Relationships that exist between topic and subtopic will be called as partOf.

```
for(int i=0;i<t1.getSubtopics().size();i++)
{
        Topic t2=basicStructuralAnalysis(t1.getSubtopics().get(i));
        RelationshipBinder rb=new RelationshipBinder(t1,t2,Relationship.partOf);
}
```

Step3

:isA relationship can be identified using AcronymHeuristics

```
{
        String subItem=relation.getTopic2().getName();
        If(subItem.matches("^[A-Z]+"))
        return true;
        return false;
}.
```

## V. CONCLUSION AND FUTURE WOK

In this way we have identified different type of relationships that are based on natural language processing techniques and ontology because building the domain module is very difficult task which require domain topic and all the pedagogical relationships that are exists among those topic hence,it save time because it provide all prerequirement for particular task.

But future work is done because it do not support multiple PDF at the same time because uploading of another PDF overwrites the content of old PDF and there is need to upload it again and again and also relationships that are identified are limited so work can be done on identifying more number of relationships.

## REFERENCES

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

[2] 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.

[3] 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.

[4] 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.

[5] 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

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

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

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

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

[10] 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.

[11] 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.