Conceptual model of Ontology in Education Era

Pravin Metkewar
Associate Professor,SICSR, Symbiosis International University, Atur Centre, Gokhale Cross Road, Deep Bungalow Chowk, Model Colony, PUNE (M.S.), India

Shrikant Mapari
Assistant Professor,SICSR, Symbiosis International University, Atur Centre, Gokhale Cross Road, Deep Bungalow Chowk, Model Colony, PUNE (M.S.), India

Sachin Naik
Assistant Professor,SICSR, Symbiosis International University, Atur Centre, Gokhale Cross Road, Deep Bungalow Chowk, Model Colony, PUNE (M.S.), India

Abstract - The Ontology Problem is a fundamental challenge of the emerging Semantic Web. This problem is comprised of three key sub-problems that are higher level Ontology Problem, the domain ontology problem and the Ontology Integration Problem. Key issues are OWL (web based ontology) reasoning is limited and has performance issues, Multi-threaded access is unstable and some API's are confusing. This paper derives current status of ontology in education domain also focusing on how to develop a conceptual and strategic model which helps us to develop an ontology in terms of Natural language parsing (NLP) , XML (extended markup language), OWL and RDF (resource descriptor framework).

Keywords: Ontology construction, NLP, OWL, RDF

I. INTRODUCTION

Evolution of Ontology-

According to the W3C, "The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries."

It can be defines the Semantic Web as "a web of data that can be processed directly and indirectly by machines. The semantic web is a vision of information that can be readily interpreted by machines, so machines can perform more of the tedious work involved in finding, combining, and acting upon information on the web. The Semantic Web, as originally envisioned, is a system that enables machines to "understand" and respond to complex human requests based on their meaning. Such an "understanding" requires that the relevant information sources be semantically structured.

The Semantic Web is a web that is able to describe things in a way that computers can understand.

Statements are built with syntax rules. The syntax of a language defines the rules for building the language statements. But how can syntax become semantic?

"If HTML and the Web made all the online documents look like one huge book, RDF, schema, and inference languages will make all the data in the world look like one huge database" Tim Berners-Lee, Weaving the Web, 1999

The RDF (Resource Description Framework) is a language for describing information and resources on the web. Putting information into RDF files, makes it possible for computer programs ("web spiders") to search, discover, pick up, collect, analyze and process information from the web.

In this paper, we have present our approach to extract relevant ontology concepts and their relationships from a knowledge base of heterogeneous text documents. We also have given the architecture of the implemented system.

II. LITERATURE REVIEW

Ayesha Ameen (2012) has stated that Ontology is an integral part of semantic web. Ontology can be design and create metadata elements required for developing semantic web applications. The evolution of semantic web has encouraged creation of ontologies in many domains. In this paper they have describe various steps involved in creation of **university ontology**. University otology can be applied to particular university. They have used

protege 4.0 alpha tool. And finally they have conclude the result with successful creation of ontology of University and it can be reused or integrated into any university to faciliate efficient access and retrival of information that can be automatically processed by machine.

Paul Buitelaar and et al (anonymous) have observe says that most of the work in ontology learning has been directed towards learning ontologies from text. As human language is a primary mode of knowledge transfer, ontology learning from relevant text collections seems indeed a viable option as illustrated by a number of systems that are based on this principle, e.g. TextToOnto and Ontolearn. All of these combine a certain level of linguistic analysis with machine learning algorithms to find potentially interesting concepts and relations between them.

Gao Shu, and Nick J.Avis (2008) have defines that ontologies are a central building block of the Semantic Web. Ontologies define domain concepts and the relationships between them, and thus provide a domain language that is meaningful to both humans and machines. In this paper, the analysis and design of a prototype ontology for visualization are discussed, whose purpose is to provide the semantics for the discovery of visualization services, support collaborative work. Relevant taxonomies are also reviewed and a formal framework is developed.

According to study of Matteo Gaeta and Stefano Paolozzi (2011) ontologies have been frequently employed in order to solve problems derived from the management of shared distributed knowledge and the efficient integration of information across different applications. However, the process of ontology building is still a lengthy and error-prone task. Therefore, a number of research studies to (semi-)automatically build ontologies from existing documents have been developed.

Tieli Sun and et al (2012) state that ontology offers great potentials as a rich source of information and has been applied to a wide area like semantic web , digital library ,etc. But creating a desirable ontology is still quite challenging. They have propose a two-layer ontology model (TLOM) , and a novel five step method for ontology building base on TLOM. This method can facilitate the organization and the evolution of ontology. Creating a formal ontology requires not just creating a conceptual vocabulary but also specification how terms in that vocabulary should be combined and what conclusion should be drawn from given set of assertions. They have stated that for ontology building most of the cases they used semi automated means with the help of tool like Protege, OntoEdit .Automatic ontology building is still unavailable yet, and that is future work.

Amal Zouaq and Roger Nkambou (2008) have presents a semiautomatic framework that aims to produce domain concept maps from text and then to derive domain ontologies from these concept maps. This methodology particularly targets the e-learning and Artificial Intelligence in Education (AIED) communities as they need such structures to sustain the production of e -learning resources tailored to learners' needs. This paper details the steps to transform textual resources, particularly textual learning objects (LOs), into domain concept maps, and it explains how this abstract structure is transformed into a formal domain ontology.

Recent research of Sara Tonelli and Marco Rospocher (2011) in Technology-Enhanced Learning (TEL) demonstrated several important benefits that semantic technologies can bring to the TEL domain. An underlying assumption for most of these research efforts is the existence of a domain ontology. The second unspoken assumption follows that educators will build domain ontologies for their courses. However, ontologies are hard to build, especially for not-tech-savvy users. Tools for ontology extraction from text aim to overcome this problem. We have conducted an empirical study with educators, both from Information Technology (IT) and non-IT domains, where they used current ontology extraction tools to build domain ontologies for their courses from their existing course material. Based on the obtained study results they have drawn conclusions about the existing ontology extraction tools and provided recommendations for their future development so that they can be beneficial for the TEL domain.

Huei-Zhen Gu (2012) states that this research is aimed to explore a methodology of building an ontology-based knowledge base in relational database. It is proposed that a metadata model of ontology is the essential element for ontology design of knowledge. The extended entity relationship model is also offered as the design method for modeling structural knowledge. Finally, the knowledge of data integrity of relational database is adopted as the example to implement a simple ontology-based knowledge base on Oracle 11gXE database for verification.

III. PROBLEM DEFINITION

The Ontology Problem is a fundamental challenge of the emerging Semantic Web. This problem is comprised of three key sub-problems, the higher level Ontology Problem, the Domain Ontology Problem, and the Ontology Integration Problem, described in detail below:

A Higher level Ontology Problem

When representing the world with ontologies, we need certain basic "building block" concepts before we can ontologically define higher-level concepts. For example, before we can really define what we mean by a "university region" we first need basic definitions of building block concepts such as "planet," "geographic location," "set," "boundary," "container" and "content of," and perhaps "elevation," "longitude," "latitude" and so on. We can call a set of building-block concepts an “Higher level Ontology."

B The Domain Ontology Problem

The nature of ontology changes domain to domain. Steps will be taken up into consideration for building ontology for Education domain, same steps probably would not consider for building ontology for some other domain like agriculture, banking, health care etc because the nature of domain in some cases top to bottom or vice versa. Meta-ontology that maps all the concepts in the most common higher level ontologies to one another so that it doesn’t matter which one is used. But this hasn’t happened yet.

Furthermore, they need to somehow connect together via a solution to the above higher level Ontology Problem so that they can be normalized and mapped to one another easily. Until that happens the Semantic Web will still be incredibly useful in case of web personalization, but only for representing and accessing knowledge base or working with domain-specific concepts that are defined by the small set of currently existing domain ontologies. Our major focus towards building ontology for Education domain in terms of parsing natural language through *nlp.stanford.edu:8080/parser/* to generate xml file.

C The Ontology Integration Problem

As discussed, it is one thing to develop an ontology but quite another to make it compatible with other existing ontologies. This is the Ontology Integration Problem. This problem turns out to be far more subtle than most people who currently write about the Semantic Web have noted as of yet. Integrating ontologies is not as simple as just mapping classes in one ontology to corresponding classes in ontology. Because it turns out that it is not merely the names and properties of classes that are significant to defining their meanings and mappings, but also their inheritance paths in their respective ontologies.

For example 1: If we mapped two ontologies to one another simply by virtue of mapping "Person" in Ontology A to the class "Person" in Ontology B. However there is big difference in actual meaning between what these two ontologies mean by "Person." Ontology A uses "Person" to mean a human with legal status in some legal system. Ontology B says that a Person is simply some type of "Living Thing" but not necessarily a legal entity. In other words, in Ontology A, "All Persons are Legal Entities" while in Ontology B, "Some Persons may be Legal Entities," while others may not be.

Another approach might be to instead develop a general semantics for expressing inter-ontology mapping concepts Use meta-ontology to create instances that express mappings between classes and properties in various ontologies. Difficulty of such integration increases exponentially to the number of concepts being integrated.

IV. CONCEPTUAL MODEL OF EDUCATION DOMAIN ONTOLOGY

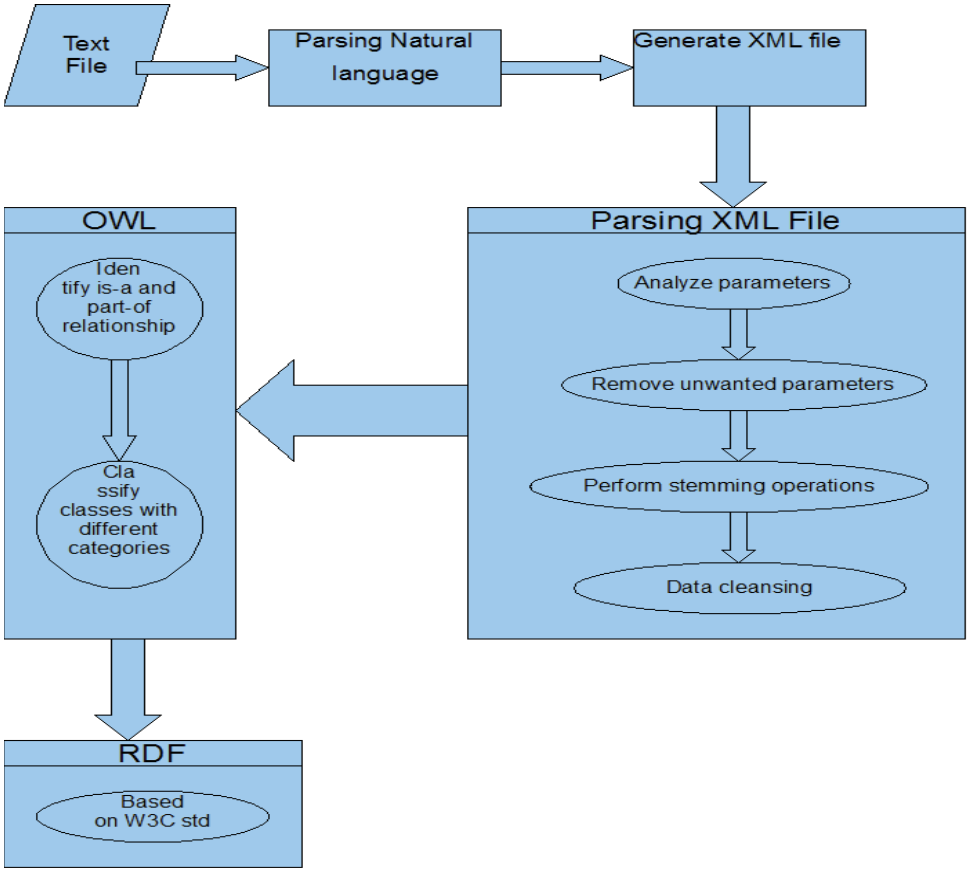

Figure 1: Conceptual model for building Ontology

V. PROPOSED METHODOLOGY

45 different categories of natural language parsing converted into 11 categories i.e., Noun, Pronoun, Demonstrative, verb, Adjective, Adverb, postposition, conjunction, particles (defaults), quantifiers, residuals (symbol, punctuation). Also in order to build ontology following steps to be taken into consideration and discussed one by one.

- Understand how to parse XML file in order to extract noun, adjectives, verb and it's is-a or part-of relationships and also interdependencies.
- Construct and understand some of the OWL language elements, and their explicit semantics.
- Use informal modeling techniques to unsolved issues relating to knowledge capture given in Assumptions 1 to 5.
- Construction and exploration of Resource Descriptor Framework (RDF).

Steps for building Education Domain Ontology:-

A XML

- Parsing natural language through *nlp.**stanford**.edu:8080/**parser**/* to generate xml file.
- It is based on so many different parameters like noun, verb, adjective, pronoun etc. Analyse the existing parameters.
- Remove parameters which are not useful in terms of ontology
- XML conversion to csv
- Performing stemming operations
- Perform pre-processing for cleaning the data.
- Identify class and its corresponding properties and methods

- Identify pronoun and merge its properties and methods with the earlier found noun or class.

B OWL

- It allows you to specify far more about the properties and classes. To form a relation between classes in terms of is-a and part-of relation
- Part-of relation are little bit domain specific e.g., mau has parents and reversely parent has child(mau). It is necessary to take care of such things.
- Classes extraction to be categorised in terms of equivalent classes, super classes, inherited classes, disjoint classes etc.
- Data Properties to be defined in terms of its range, data type , width, equivalent property, super property, disjoint property etc
- Methods to be defined in terms of its return type, types of parameter, data type , equivalent methods, super methods, disjoint methods etc

C RDF

The *Resource Description Framework* (*RDF*) is a general framework for how to describe any Internet resource such as a Web site and its content. It follows World Wide Web Consortium (W3C) standard. RDF was designed to provide a common way to describe information so it can be read and understood by computer applications.

VI. STRATEGIC MODEL BASED ON ASSUMPTIONS, CURRENT PROBLEMS AND PROPOSED SOLUTIONS

Assumption 1: Automatic ontology building is still unavailable yet and this is the need of the day [4].
Existing Problem: Ontologies which are available over the globe are semi automatic. Even in case of Protégé 4.0 alpha version, we need to manually analyse the things, relationships and construct domain ontology through tool.
Proposed solution: Parsing natural language is a big and complicated task which is the first step of ontology building. In our proposed scenario we would parse the text file through nlp.stanford.edu:8080/**parser** which will generate xml file that gives details about noun, verbs, adjectives, pronouns etc. Xml file will be used for further processing to construct OWL and RDF. So as much as possible would try to atomise it.

Assumption 2: There is a clear need for tools that would enable users to manipulate the generated ontology in terms of Ontology concepts and relationships [5].
Existing Problem: In existing tools there is no such facility to transfer object properties from one class to another because of that nature of ontology becomes complex.
Proposed solution: We would propose facility to transfer object properties from one class to another and even we can change the existing mapping as per user requirement.

Assumption 3: Avoid too many useless vocabularies.
Existing Problem: In NLP parsing, it will cover entire text which is given because of that size of vocabularies which is more and difficult to manipulate it. Ultimately it increases the model complexity.
Proposed solution: In order to resolve above problem, we would propose to parse xml file in such a way that it will remove basic words which are no more required, perform stemming operation to obtain common name (remove ing, ed, es from the given word). Also we can remove nouns which are no longer required.

Assumption 4: Don't think too much about Java classes when you define OWL classes
Existing Problem: While dealing with classes, its details are missing in terms of meaningful class name, behaviour etc. Also defined property scope is also not clear.
Proposed solution:

- When defining the method (property) of a Java class, would specify its name and *behavior*
- When defining the property of an OWL class, would specify its name and *restrictions*

Assumption 5: Rule Based Inferences.
Existing Problem: Problems related with noun and its corresponding pronoun and also multiple occurrences of nouns.
Proposed solution:

- Rule 1: Pronoun which comes in next sentence followed by verbs and adjectives to be merged with noun in first sentence followed by verbs and adjectives.
- Rule 2: Occurrences of same noun followed by verbs and adjectives again merged with earlier found noun.

VII. CONCLUSION

This paper derives current status of ontology building in education domain. Also have more emphasis on current problems, its implications and limitations. Solution will be proposed in terms of providing conceptual model in the form of n – tier architecture. On the basis of analysis so far, tools available in the globe are not fully automised and in order to automise it fully, what are the steps to be taken into consideration that had discussed in this paper in the form of Assumtion 1 – 5.

REFERENCES

[1] Matteo Gaeta, Stefano Paolozzi, and Saverio Salerno "Ontology Extraction for Knowledge Reuse: The e-Learning Perspective" IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A: SYSTEMS AND HUMANS, VOL. 41, NO. 4, JULY 2011.
[2] Tieli Sun, Fengqin Yang, Dapeng Zhang ,Lin Yang, "Ontology Building Based on Two-layer Ontology Model" 2012 International Conference On Industrial Control and Electronics Engineering.
[3] Sara Tonelli, Marco Rospocher, Emanuele Pianta and Luciano Serafini "Boosting collaborative ontology building with key-concept extraction" 2011 Fifth IEEE International Conference on Semantic Computing.
[4] Marek Hatala, Dragan Ga_sevi_c, Melody Siadaty, Jelena Jovanovic, and Carlo Torniai "Ontology Extraction Tools: An Empirical Study with Educators" IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES, VOL. 5, NO. 3, JULY-SEPTEMBER 2012.
[5] Huei-Zhen Gu "Research on Building Computation of Ontology-Based Knowledge Base" 2012 12th International Conference on ITS Telecommunications 978-1-4673-3070-1/12/ ©2012 IEEE 437
[6] Ayesha Ameen , Khaleel Rahman Khan, B.Padmaja Rani , "Construction of University Ontology" IEEE World Congress on Information and communication Technology ,2012
[7] Gao Shu, Nick J.Avis, Omer F. Rana "Investigating Visualization Ontologies"
[8] Amal Zouaq and Roger Nkambou, "Building Domain Ontologies from Text for Educational Purposes," IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES, VOL. 1, NO. 1, JANUARY-MARCH 2008
[9] Paul Buitelaar1, Daniel Olejnik1, Michael Sintek2 "A Protégé Plug-In for Ontology Extraction from TextBased on Linguistic Analysis".
[10] http://en.wikipedia.org/wiki/Semantic_Web
[11] http://www.w3schools.com/web/web_semantic.asp.