A STUDY OF VARIOUS APPROACHES AND TOOLS ON ONTOLOGY

Sanju Mishra[1]
Department of Computer Applications
Teerthankar Mahaveer University,
Moradabad, India
sanju.tiwari.2007@gmail.com

Sarika Jain[2]
Department of Computer Applications
National Institute of Technology
Kurukshetra, India
jasarika@nitkkr.ac.in

***Abstract*— The increasing volume and unstructured nature of data available on the World Wide Web (WWW) makes information retrieval a tedious and mechanical task. Lots of this information is not semantic driven, and hence not machine processable, but its only in human readable form. The WWW is designed to builds up a source of reference for web of meaning. Ontology information on different subjects spread globally is made available at one place. The Semantic Web (SW), moreover as an extension of WWW is designed to build as a foundation of vocabularies and effective communication of Semantics. The promising area of Semantic Web is logical and lexical semantics. Ontology plays a major role to represent information more meaningfully for humans and machines for its later effective retrieval. This paper constitutes the requisite with a unique approach for a representation and reasoning with ontology for semantic analysis of various type of document and also surveys multiple approaches for ontology learning that enables reasoning with uncertain, incomplete and contradictory information in a domain context.**

Keywords*— *Ontology, Semantic Web, Ontology Learning, Concepts, Concept Hierarchy.

I. INTRODUCTION

Knowledge comes from different sources like unstructured, structured or semi structured text. Knowledge can be interpreted in the form of both machines and human readable forms by using ontology. Ontologies capture the domain knowledge in an inclusive way and provide a shared agreed upon understanding of a domain. Ontology usually encompasses modeling primitives such as terms, concepts, generic relations between concepts, and axioms. Ontologies represent and share the knowledge within an application domain. Manual construction and population of ontologies, is a very time-consuming and labor-intensive task. Present research in the field of automatic and semi-automatic ontology acquisition and development provides methods and solution to solve this problem. Hearst [Hearst, 1992] introduced several methodologies for ontology learning and ontology population for assisting to build ontologies.

Ontology learning systems can be categorised by data types, which they learned [Gomez and Macho, 2003]. Ontology learning system can take three types of input (i) unstructured data like any text file, books etc. (ii) Semi-structured data like HTML/XML files, (iii) structured data like databases. Ontology learning involves identifying ontology elements such as terms, synonyms, concepts, relations, properties and axioms from textual sources.

II. VARIOUS APPROACHES ON ONTOLOGY

Multiple surveys have been reported since 2000. In 2003, the first OntoWeb[Gomez and Macho, 2003] Consortium was performed by Gomez Perez and Macho. In this survey 36 approaches have been discussed about onto learning from text and emphasized on three things such as methodology, system for ontology learning and approaches for accuracy evaluation. In the same time [Shamsfard and Barfourash, 2004] presented the second survey on ontology learning and claimed for discussion of 50 approaches in this survey. The main aim of this survey was to introduce about the framework for ontology learning approaches comparison. The approaches are basically served as test cases in framework. The major area of this review was based on (1) finding the hierarchical relations while non-hierarchical relations were less observed (2) axiom learning was undiscovered (3) developing domain ontology. Current systems require domain specific patterns for different ontologies.

The third survey was presented in 2002 by [Ding and Foo, 2002] with 12 ontology learning projects and their review was based on different findings such as almost structured data for input, difficult task to discover relation and also it is a complex task to solve. In 2005 Buitlaar et al. [Buitlaar et al., 2005] represented a combined work in a workshop with their 10 papers. These papers are combined in their book. Author highlights the use of “Ontology Learning Layer Cake” to refine the different phases of Ontology Learning and observed the things such as axiom extraction tasks and the importance of evaluation platform to upgrade the progress of ontology learning.

It is proved that ontologies are beneficial for different applications such as heterogeneous data, information integration, information retrieval and question answering. Hence ontology nurtures the interaction between systems. According to Sowa [Sowa, 2000], ontologies are categorized in 3 types (i) formal ontologies, conceptualization of specific domains which are represented by rules and axioms, these are

978-1-4799-6023-1/15 $31.00 © 2015 IEEE
DOI 10.1109/CICT.2015.43

also stated by logic to represent the computations and complex inferences (ii) prototype based ontology: these are totally based on prototypes or instances on behalf of axioms and logics. (iii) Lexicalized ontology: In this type of ontology the subtype-super type relations are represented rather than prototyping of data.

Ontologies are applied in a variety of applications, including web service discovery [Paolucci et al., 2002], information integration [Alexiev et al., 2005], natural language processing [Nirenburg and Raskin, 2004] and dynamic composition of web services [Sirin et al., 2003]. According to Velardi [Velardi, 2011] there are two approaches to learn ontology, first is rule based ontology which is based on predefined rules or heuristic patterns introduced by Hearst 1992 in the form of lexico-syntactic patterns. In this paper author introduces OntoLearn Reloaded, a graph-based algorithm for learning taxonomy and develops the system without using Wordnet.

Maedche and Volz [Maedche and Volz, 2001] have developed a semi-automatic system that is part of ontology management infrastructure KAON. It was a learning system which uses linguistic and statistics based approaches. Text-To-Onto incorporated the OntoEdit system in itself for ontology engineering. But there was a problem in these systems that they are domain dependent specific ontology models. Cimiano introduced a ontology framework Text2Onto, an extended version of Text-To-Onto which represents ontological structures at Meta level hence called modeling primitives and developed a Probabilistic Ontology Models (POM). This model can be translated into different ontology representation languages such as RDFs, OWL and F-Logic.

Buitelar [Buitelar, 2005] presents the Ontology Learning Layer Cake in figure 1.

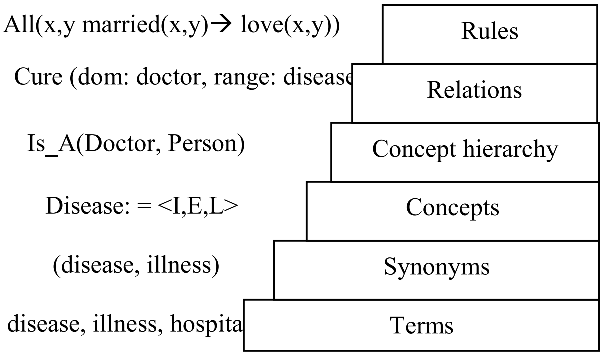

Fig 1: Ontology Learning Layer Cake

In this example, knowledge is defined for the concept disease and related concepts, there can be number of terms in different languages to refer or associated with a single disease. There is a hierarchical relation between the concept doctor and person, and non-hierarchical relations between the concepts doctor and disease. A rule can be formed, defined over the person and disease concepts.

Raja Mohan and Arumugam [RajaMohan and Arumugan, 2012] presented a methodology to develop and presume the Indian Medicinal Plant Ontology based on the usage of protégé 3.4.4. Medicinal Plant Ontology is one of the best ontology for people. It is implemented by using OWL and SWRL and a study of Medicinal Plants to categorize the things. The medicinal plants and their parts are good source for preventing the diseases. Medicinal Plant ontology is taxonomy of plant, flowering plant, non-flowering plants, and parts of plants, special names and subclass of flowering plants. Prior research on ontology design methodologies shows that, manual construction of ontology is a tedious task and also very hard for a designer to construct a consistent ontology.

Rene Robin [Robin and Uma, 2010] presented an ontology-based e-learning system for software risk management (SRMONTO). To accomplish this approach author uses the automatic extraction techniques to build domain ontology for e-learning purposes from heterogeneous sources. Author highlighted the design and construction of ontology-based e-learning system and hence evaluates the presented system. In their paper, the authors provide an efficient use of SRMONTO for the proposed e-learning system that enables to provide learners with a more attractive education support.

Srivastava and Bhattacharyya [Srivastava and Bhattacharya, 2008] presented a simple POS tagger based on Hidden Markov Models (HMM) for the task of POS tagging. They tried to exploit the morphological affluence of the languages without addressing to complex and expensive analysis. Angrosh and Shalini Urs [Angrosh and Urs, 2007] proposed domain ontology for ontology-based information retrieval system of Agricultural Electronic Theses and Dissertations (ETDs) which is available in Vidyanidhi Digital Library. They have successfully developed a prototype ontology-based information retrieval system for a specific case of Agri-Pest domain [Angrosh and Urs, 2006].

OntoLearn [Velardi, 2005] was the first system to automatically extract the taxonomy from documents and websites. It is the corpus based ontology learning system. This system has five main algorithms such as term extraction, concept extraction by using natural language definition, parsing of extracted concepts, semantic disambiguation (using wordnet), and identification of relationships between extracted concepts. First three algorithms are simple to evaluate but last two algorithms needs expertise of domain specification to evaluate the reliability of system.

III. NEED OF ONTOLOGY

The Web became an object of our daily life and the amount of information in the web is ever growing. Besides plain texts, especially multimedia information such as graphics, audio or video has become a prevalent part within the web's information transport. But how to find some useful information within this huge information space? How to achieve homogeneity in heterogeneous digital information resources? The increasing volume and unstructured nature of data available on the World Wide Web (WWW) makes information retrieval a complex and mechanical task. Lots of this information is not semantic driven, and hence not machine understandable, but its only in human readable form. Almost since the emergence of computers, one basic goal of computer science research has been to be able to process the meaning of symbols and not only the syntactic structures of the language. The World Wide Web is a system of interconnected hypertext documents that are accessed via the Internet. These web pages follow the HTML, XML languages which are unstructured heterogeneous resources so unable to make an efficient search. To overcome with this problem, ontology is used to conceptualize the domain and represent an area of knowledge. There is a need to analyse or conceptualize a domain to invent an effective knowledge representation system hence ontology is a solution to represent the knowledge and provide the vocabulary to share the information and to make machine interpretable concepts.

Ontology term is borrowed from philosophy, it's not new. According to Gruber "Ontology is a specification of a shared conceptualization of a domain [Gruber, 1993]". It is a data model to represents a set of concepts within a domain and the relationships between those concepts. According to W3C "Ontologies define the terms used to describe and represent an area of knowledge".

This research is plenteous for managing the unorganized data which is unstructured in nature. It is a time consuming and expensive task to remove the ambiguity on web document searching where search is based on keywords not conceptual and unstructured by nature. To overcome with this problem there is an ontological approach for conceptualization of the domain. Conceptualization means inventing an idea or explanation and formulation it mentally. The result of conceptualization is a set of concepts. According to Gruber "A conceptualization is an abstract simplified view of the world that we represent for some purpose". For the construction of information systems in the domain we need a knowledge yielding concepts in machine understandable format. This knowledge should be systematic, well defined and agreed upon main persons in the field. This knowledge which is always hidden between terms and mostly appears in an unstructured or semi structured form which makes the machine unsuitable for further processing. This knowledge can be represented in the form of concepts which are the basic building blocks for ontologies and serves as a key element in knowledge representation. Concepts are generally used by people; usually they don't have any unique definition for concepts. A general meaning of concept is abstract idea such as: House, Animal, Drinks etc. To develop ontology there is an important phase, Ontology Learning phase.

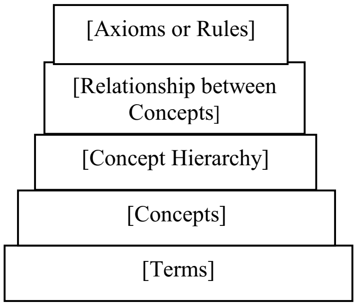

Fig 2. Ontology Learning Phases

Combining all the efforts in ontology learning, ontology can be defined as:

$O = \{(T, S), (C, CH), R, A\}$ Where O is Ontology.

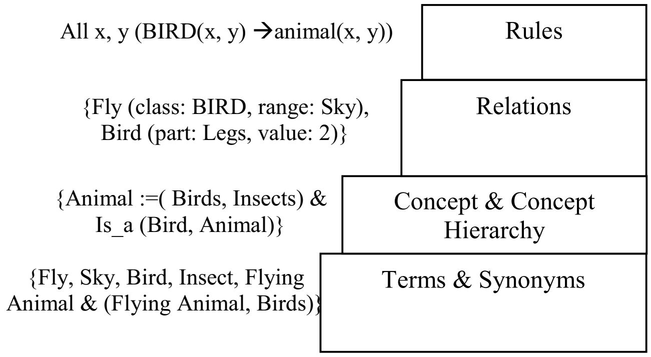
All x, y (BIRD(x, y) →animal(x, y))
Rules
{Fly (class: BIRD, range: Sky), Bird (part: Legs, value: 2)}
Relations
{Animal :=(Birds, Insects) & Is_a (Bird, Animal)}
Concept & Concept Hierarchy
{Fly, Sky, Bird, Insect, Flying Animal & (Flying Animal, Birds)}
Terms & Synonyms

Fig 3. Ontology Learning

According to Sowa [Sowa, 2000] "concept is a mediator which relates to its objects". There are number of approaches to represent the knowledge for machine process able but ontology is an approach for modern computer science as a formal representation of knowledge. Ontology Language [Bechhofer et al., 2004]. Buitelaar [Buitelaar et al., 2005] and Cimiano [Cimiano et al., 2006] propose that all of the aspects of ontology development can be organized into a layer stack with the more complex tasks being at the top. Ontology Learning remains an active area of research with considerable potential for facilitating the work of ontology engineers.

Schatz [Schatz, 1997] represents a timeline of searching evolution as shown in table 1. To improve the semantics based searching ontology is used. Ontology maps all type of data in their own concepts. For example: If there is a word like a table or board to search on google there are numbers of results but having the different meanings due to keyword matching only. Some results belong to furniture for table while others may for tabular format table. Hence it is time consuming process to search the appropriate data and remove the ambiguity and also a tedious task. If the data will be in conceptualized form then repetition of data can be reduced and hence searching process can be efficient.

Table 1: Timeline of Searching Evolution by Schatz (1997)

S. No.	Year	Search
1.	1960	Syntax (Text Searching)
2.	1970	
3.	1980	
4.	1990	Structure (Document Search)
5.	2000	
6.	2010	Semantics (Concepts Search)

IV. OBJECTIVES OF THE STUDY

In this work we present the concepts and knowledge required for the proposed approach. The objective of this study is to facilitate the knowledge representation by using Ontology from texts in real-world setting through several information extraction and learning approaches. There will be three things to do by this research:

- Extract the relevant terms from unstructured/semi structured data.
- Concept formation and developing concept hierarchy.
- Hence developing ontology based framework for knowledge representation.

All the available systems have used different representation schemes. We want to focus on some representation scheme which is unique and portable. So our main aim is to provide a framework for knowledge representation towards ontology learning.

V. CONCLUSION

As ontology becomes an important way to structuring the information in intelligent systems hence it is required to develop a new approach from existing approaches. Due to availability of inference engine in ontology, reasoning process is easy to do. In this paper we introduce about various surveys and tools to represent the knowledge of a domain which does not only highlight the concepts, but also the relation between concepts and level of abstraction to define relation. A comparison table of ontology existing tools has presented in table 2 which compared number of tools with their properties of previous work.

REFERENCES

A. Gomez-Perez and D. Manzano-Macho, A Survey of Ontology Learning Methods and Techniques, Deliverable 1.5, Onto Web Consortium, 2003.

A. Maedche, R. Volz 2001, The Ontology Extraction & Maintenance Framework: Text-to-Onto. In Proceedings of the IEEE International Conference on Data Mining, 2001.

A. Raja Mohan, G. Arumugam, Developing Indian Medicinal Plant Ontology using OWL and SWRL, ICDEM'10 Proceedings of the Second international conference on Data Engineering and Management, Pages 131-138, Springer-Verlag Berlin, Heidelberg, 2012.

B. Schatz, Information Retrieval in Digital Libraries: Bringing Search to the Net. Science, 275(5298):327–334, 1997.

C.R Rene Robin, G.V. Uma, Ontology Based Semantic Knowledge Representation for Software Risk Management, International Journal of Engineering Science and Technology, 5611-5617, 2010.

E. Sirin, J. Hendler, and B. Parsia, Semi-Automatic Composition of Web Services using Semantic Descriptions. In Proceedings of the ICEIS 2003 Workshop on Web Services: Modeling, Architecture and Infrastructure. Angers, France, 2003.

J. F. Sowa, Knowledge Representation: Logical, Philosophical, and Computational Foundations. Brooks Cole Publishing Co., Pacific Grove, CA, 2000.

L. Zhou, Ontology Learning: State of the Art and Open Issues. Info. Technol. Manage. 8, 3, 241–252, 2007.

M.A. Angrosh and Shalini R. Urs, Ontology-driven Knowledge Management Systems for Digital Libraries: Towards creating semantic metadata based information services, In: Proceedings of National Seminar on Knowledge Representation and Information Retrieval, Paper: N. Document Research & Training Centre, ISI, Bangalore, 2006.

M.A. Angrosh and Shalini R. Urs, Development of Indian Agricultural Research Ontology: Semantic rich relations based information retrieval system for Vidyanidhi Digital Library. International Conference on Asian Digital Libraries, ICADL 2007.

M.A. Hearst, Automatic Acquisition of Hyponyms from Large Text Corpora, Proceedings of the 14th International Conference on Computational Linguistics, 539–545, 1992.

M. Shamsfard and A. Barforoush, Learning Ontologies from Natural Language Texts. Int. J. Human Comput. Stud. 60, 1, 17–63, 2004.

M. Srivastava, P. Bhattacharyya, Hindi POS Tagger Using Naive Stemming: Harnessing Morphological Information without Extensive Linguistic Knowledge, International Conference on NLP (ICON08), Pune, India, 2008.

M. Paolucci, T. Kawamura, T.R. Payne, K. Sycara, Semantic Matching of Web Services Capabilities. The Semantic Web – ISWC, 333-347, 2002.

P. Buitelaar, P. Cimiano and B. Magnini, Ontology learning from text: An Overview In Ontology Learning from Text: Methods, Evaluation and Applications, Eds. IOS Press, Amsterdam 2005.

P. Velardi, R. Navigli, A. Cuchiarelli, and F. Neri, Evaluation of OntoLearn, a Methodology for Automatic Population of Domain Ontologies, 2005.

P. Cimiano, J. Volker, Text2Onto A Framework for Ontology Learning and Data-driven Change Discovery, 227-238, 2005.

R Navigli, P. Velardi, S. Faralli, A Graph-Based Algorithm for Inducing Lexical Taxonomies from Scratch, IJCAI, 1872-1877,2011.

S. Nirenburg, V. Raskin, Ontological Semantics, A prepublication draft, chapter by chapter, can be found on ontological semantics.com, Cambridge, MA: MIT Press, Boston, 2004.

S. Bechhofer, F. Van, Harmelen, J. Hendler, I. Horrocks, D.L. McGuinness, P. F. Patel-Schneider and L. A. Stein, OWL Web Ontology Language Reference. W3C Recommendations. Available at http://www.w3.org/TR/owl-ref/, 2004.

T.R. Gruber, A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, 5 (2), 199-220, 1993.

V. Alexiev, M. Breu, J.D. Bruijn, D. Fensel, R. Lara and H. Lausen Information Integration with Ontologies: Experiences from an Industrial Showcase. John Wiley & Sons, Chichester 2005.

Y. Ding and S. Foo, Ontology Research and Development: Part 1. J. Inf. Sci. 28, 2, 123–136, 2002.

Table 2: Comparison Table of Existing Tools

Tools	**Element Learned**	**Base Ontology**	**Type of Input**	**Input Language**	**Learning Approach**	**Representing Language**
ASIUM	Frames of verbs, Ontologies from parsing of text.	Conceptual-Clustering	Unstructured	French	Conceptual Clustering	Frame Based
TEXT-TO-ONTO	Dictionaries, Ontologies, NL Text	Conceptual-Clustering	Semi-structured(web data, HTML data)	German	Formal Concept Analysis + Clustering	F-logic based extension of RDF
TEXT2ONTO	Ontology learned Structure.	POM(Probabilist-ic Ontology Model)	Textual data	HTML,XML, KB,Ontologies, DTDs	NLP	RDFs, OWL, Flogic
WEB-KB	Instances of Classes and Relationships.	Ontology for which instances are learned.	Ontology+ examples of instances	HTML	Bayesian Learning, FOIL	First order rule logic
HASTI	Concepts, Taxonomic and Non-Taxonomic, Relations, Axioms	Seed Ontology(small kernels of primitives	Unstructured Data	Persian	Linguistic based, semantic analysis, logical reasoning	Subset of KIF
DODDLEII	Taxonomic and Non-Taxonomic Relations	WordNet	Domain Specific Text	English	Analysis of lexical co-occurences statistics	
SYNDICA-TE	Words, Concepts, Taxonomic and Non-Taxonomic	Knowledgebase	Unstructured NL text	German	Quality learning + semantic analysis	Description Logic Language
SVETLAN	Classify Noun Texts		Unstructured Data	French	Clustering based on distributional similarties	special format of structured domain
ONTO-LEARN	Documents and Data from Web		Unstructured/semi structured Data	French	NLP and Statistical Method	
TF/IDF	Concepts and Relations between them		NLText (Semi)		Text Mining, Statistical Approach	