A Survey on Ontology Approaches in Data Mining for Various Technologies

K. Poongothai[1], A. Suganthi[2]

[1, 2]*Assistant Professor, Department of Computer Science & Engineering, KGiSL Institute of Technology, Coimbatore, India*

***Abstract*- Ontologies play an important character in semantic web, internet and other social related application. It also plays an important role in engineering technology such as Data Mining and other applications. Normally ontologies focus on the ultimatum (demand) of high quality of ontology. The ontology is more expensive and incompetent of time process. Ontology learning is the process of assembling the ontologies by the ontology engineer. Ontology learning is a framework that proceeds with some of the fields which specified in the form of ontology import, extraction, pruning, refinement, and evaluation given by the ontology engineer. The problem that the ontologies learning technique deals with the difficulty to retrieve the knowledge relevant to the interest of the users or stakeholders. In this survey paper, some of the ontology techniques can be used in various fields like NLP, Text Mining, database integration and other fields like biomedical, medical and web technologies.**

***Keywords*- Ontology, Ontology Learning, NLP, Semantic Web.**

I. INTRODUCTION

A. Ontology

In the last decades, ontology plays an essential role in various fields such as Database Integration, Artificial Intelligence, Semantic Web, Natural Language Processing (NLP) and Web Technologies. The ontology has been developed to enable interoperability across multiple information systems. The ultimate role of the ontology is to support by knowledge sharing and reclaim the process. Ontologies play a vital role in various web and internet related applications to share knowledge in specific domain. Ontology can be specified or constructed as machine readable and understandable. The main role of the ontology is to express as the formal representation of knowledge by a set of concepts within a domain and their relationship. The researchers in ontology mainly focus on accelerating the use of time and reduce the effort of building the ontology by using the different information sources or databases. Normally, Ontology is a set of limitations, which affirm (announce) what should automatically hold in any possible or different world.

In Artificial Intelligence, the word ontology is defined as formal, explicit specialization of a shared conceptualization. Ontology should be machine-readable and improves the information organization, management and understanding. Ontology has a significant role in the areas dealing with vast amounts of distributed and heterogeneous computer-based information. Ontology concepts and relationships should be explicitly defined. Search engines will use ontology to find pages with words that are syntactically different but semantically similar. Researches try to build ontology automatically or semi-automatically to save time and the efforts of building the ontology.

Ontologies can be described in two main categories as Domain ontology and Term ontology. Domain ontology specifies about the formal description of the classes and the relationship between the classes whereas the term ontology specifies about the abstract and smaller which focus on the time, space, plans. Ontologies can be fabricated (constructed) by using three approaches and these approaches are needed to help in by providing the integration task to describe about the information sources. The three approaches are single ontology approach, multiple ontology approach and hybrid ontology approach. Normally, ontology creation is an iterative procedure. The ontology supports various standards like easy transmission and interpretation of data for various applications.

The main use of the term ontology is to provide a common understanding of specific domains or areas which can be used to communicate with the people and the application system. The most and important role of ontologies is to share a common information structure and enabling reuse of already existing domain knowledge instead of creating a new domain knowledge. Ontology is a backbone of the Semantic Web.

The term ontology can be expressed in the format as a tuple

$$v := (C,R,H_C, rel, A^v)$$

Where C-> set of ontology concepts

R-> set of non-taxonomic relationships

rel-> R->C x C

A^v-> set of axioms

Many of this success depends on the ability to share, reuse and personalize existing ontologies since designing and maintaining ontologies is deemed to be a time-consuming and labor intensive task. The reuse of existing ontologies can occur either while designing and maintaining ontology or development of new applications. Existing and widely used

ontology languages such as OWL do not support reuse of parts of other ontologies.

B. Ontology Work

The ontology and ontology engineering plays a vital role in Semantic Web and other fields such as Artificial Intelligence, Medicine, Database Integration, Information Retrieval. Ontology Engineering is one of the field that reconnoiters (explores) the methods and tools for handling the ontology life-cycle. Ontology is a domain-independent methodology which provides the guidance for the ontology-building, modification or alteration and estimation.

The ontology life-cycle can be diagrammatic represented by four main stages: the specification stage, the formalization stage, the maintenance stage and the evaluation stage.

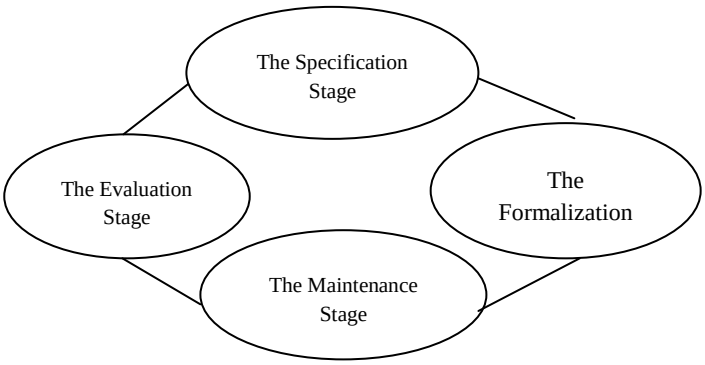

Fig: Ontology Life-Cycle

- *The Specification Stage* is a dependent approach which allows the identification of the purpose and the scope of the ontology. The Specification Stage mainly relies on the domain professionals (experts) and requires the explanation of capability questions that the ontology must be able to answer.
- *The Formalization Stage* provides a conceptual and formal model that mainly satisfies the specification stage of the ontology life-cycle.
- *The Maintenance Stage* mainly used to allow the ontology update and development or progress and checks the consistency.
- *The Evaluation Stage* analyzes the resulting ontology and checks the initial needs and satisfies the desired features of the ontologies.

C. Ontology Learning

Ontology leaning refers to the process of automatic or semi-automatic support for the assembly of ontology. The ontology learning is mainly concerned with the knowledge discovery from different data sources through the ontologic structure. The automatic or semi-automatic will support for the instantiation of a given ontology which mainly refers to the ontology population in the concept of ontology learning. The ontology learning involves two necessary characteristics. The first characteristics are used for the availability of prior knowledge and here the effort of knowledge transformed is difficult. The second characteristics are the input involved in the process of ontology learning. Here the input can be taken in the three different formats as:

a. Structured data which is used to extract the parts of ontology by using the available structural information as database organizations.
b. Semi-structured data is a data which is composed by some of the structural information that need to be developed from free text. E.g. Dictionaries.
c. Unstructured data is a method which is used to extract the ontology parts from unstructured data that do not depends on any structural information for improving the quality. E.g. Natural Language text documents.

The layers that are included in the ontology learning development process are

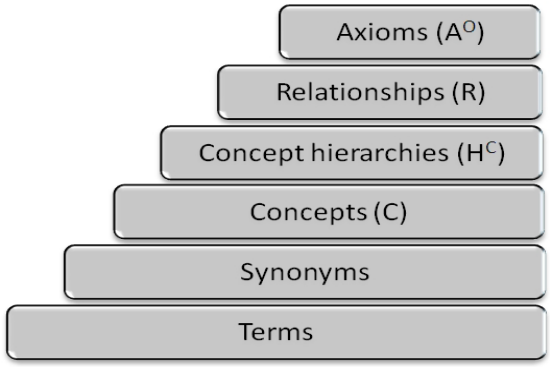

Fig: Layers in ontology development process

D. Ontology Modules

An ontology module is a reusable component of a larger or more complex ontology, which is self-contained but bears a definite association. The modules can be reused or extended by introducing new concepts and relationships, asserts that they are not isolated entities or disjoint from each other. The main goals of ontology modularization are partial reuse, complexity, ownership and customization, efficient reasoning and tooling support. The properties of ontology modules and modularization are size, correctness, completeness, localized semantics, correct reasoning, transitivity, safe reuse and decidability.

Techniques in Modular Ontologies

Logic-Based Approaches:

In this approach, formal algorithms based on sound logical foundations are developed for module extraction that are complete and correct. When the module is correctness, it satisfies trivial extracted by the set of axioms from an ontology about the ontology information. When it is

completeness, it is somewhat non-trivial and defined based on the conservative extensions of other ontology.

Graph-Theory Based Approaches:

It is simple approach to modularize ontology is to transverse ontology hierarchy and apply heuristics to identify a sub-graph. These algorithms are tractable, intuitive to a user and are widely used. It represents a set of graph traversal-based modularization algorithms. This approach is useful that do not take into consideration the underlying semantics of the ontology.

E. System Overview

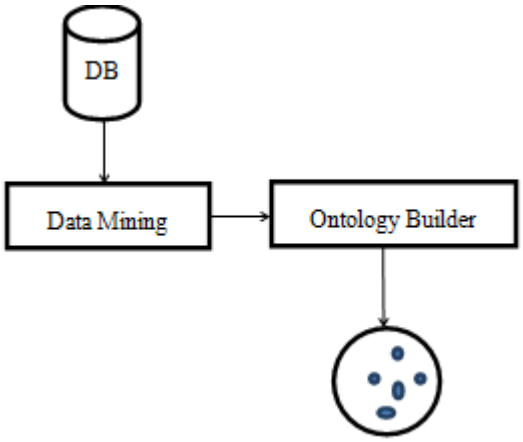

Fig: Overall Structure of the System

II. MOTIVATION

In this survey paper, Ontologies plays an important role which is used to retrieve the knowledge from the database and can be used for reusing the extract information. The ontologies can be used in various domains such as Natural Language Processing (NLP), Artificial Intelligence, Database Integration, Semantic Web, Text Mining and Web Technologies. Beforehand the establishing of the ontology learning process, the steps of ontology development have to be defined whether it should be manual or automatic. The ontology learning task has some of the issues regarding the three approaches such as structured, semi-structured and unstructured data. Here the clustering algorithms can be used for some of the ontology learning from text approaches. The main motivation for using the ontologies in various fields is to extract the information and can be reuse with existing data. Building ontology is a complex work, in order to build ontology that focus on the domain expert to declare all domain concepts and relationships between them.

III. ONTOLOGIES TECHNIQUES

Here the various ontologies techniques can be discussed which can be used in ontology design and ontology learning process that need to be motivated.

A. Wordnet:

One of the most and important well-developed ontologies is the Wordnet. It is an automatically assembled for lexical reference system. The central motivation for using the wordnet is a set of synonyms. Normally the wordnet refer to as the synset and it is organized as a hierarchy by superclass-subclass relationship i.e., hypernymy-hyponymy. The wordnet is a simple hierarchy that focused on the hierarchy of noun synset. The verbs in the wordnet are divided into 15 clusters according to their meaning with the primary relationship between the verbs in the cluster.

B. Generalized Upper Model:

The generalized upper model is a universal assignment and it is domain-independent interested ontology that supports sophisticated (refined) natural language processing in English and explained in many languages. Here the level of abstractions is in between the lexical (verbal) knowledge and theoretical (conceptual) knowledge. This technique also based on the hierarchy concept and these can be represented into three major subtypes as 1) **Configuration,** participate all elements in some activity or state of relationships 2) **Element,** single conceptual term and 3) **Sequence,** complex activities are connected by some of the relation which forms the sequence structure.

C. Unified Medical Language System (UMLS):

This system is mainly used to design for facilitate reclamation and can integrate or participate the multiple machine readable medical information resources. The main goal of this system is to facilitate the link between different sources and the users can use in different terminology. It is used to represent the semantic network which describes about the additional relationship between various categories. The hierarchies include for both entities and events includes like activity, progression and grievance.

D. Toronto Virtual Enterprise (TOVE):

This project is an example of domain-specific ontology for a specific task and it is an enterprise modeling. This project mainly focuses on the concept of creating the enterprise that can answer the questions to pertain the information in the model and gather the answer to queries. These questions justify the choice of concept and the relations among the ontology. The ontology engineers determine the problem that arises in the actual enterprise. Here axioms and affairs are used to link knowledge rom various clusters.

E. GENISM:

It is a generic simulation system which includes macromolecules for complex internal structures such as DNA and RNA. The GENISM includes two sub-ontologies as 1) the class knowledge base and 2) the process knowledge base. The general classes of objects are represented as frames. Frames

are used to represent instances of objects in simulations. The process of GENISM describes the potential behavior of objects in the system and uses the qualitative process theory approach to represent processes. It is used to simulate the reactions of information to specify the starting conditions.

F. Statistical Approach:

This technique mainly focus on building the ontology using some of the keywords that are very close to the ontology concepts and closely related to each other. This approach is used to retrieve the related pages that describe the importance of the specified domain in the search engine. Some of the attributes are that need to be focused are

- Number of different web sites for retrieving the concept
- Total number of presence by analyzed using web site
- Ratio between twofold preceding measures

G. Natural Language Processing (NLP) Approach

It uses a set of syntactic patterns to introduce the dependency relations between the different words. The dependency relation is defined as the asymmetric binary relation between the word called head and the word called modifier. It represents as the automatic approach for describing the taxonomies from a text document. It described the relationship between the objects and the set of attributes or the attributes themselves. It is closer to the concept hierarchy and produces the ontology in lattice form.

IV. CONCLUSION

This survey paper represents the novel approach of ontology and ontology learning framework in various fields like Semantic Web, Artificial Intelligence, Database Integration, Web Technologies, NLP, Text Mining and medical applications. The ontology plays an important role in Software Engineering too to extract the useful information from the large database and the data can be reuse for the existing domain. Here some of the ontologies techniques and the ontologies approaches were discussed to extract the important information from the large database that can be used in various fields which is used from the early years.

REFERENCES

[1]. C. Roussey, F. Pinet, M.Ah Kang, O. Corcho: “An Introduction to ontologies and Ontology Engineering”, Springer-Verlag London Limited, 2011.
[2]. M. Hazman, S. R. El-Beltagy, A. Rafea: “Ontology Learning from Domain Specific Web Documents”, International Journal of Metadata, Semantics and Ontologies”, 20008.
[3]. A. Zouaq: “A Survey of Domain Ontology Engineering: Methods and Tools”.
[4]. L. Drumond, R. Girardi: “A Survey of Ontology Learning Procedures”.
[5]. D. C. Wimalasuriya, D. Dou: “Ontology-based Information Extraction: An Introduction and A Survey of Current Approaches”, Journal of Information Science, pp. 306-323, 2010.
[6]. R. Subhashini, J. Akilandeswarai: “A Survey on Ontology Construction Methodologies”, International Journal of Enterprise Computing and Business Systems, Vol.1, Jan 2011.
[7]. N. Fridman Noy, C. D. Hafner: “The State of the Art in Ontology Design”, American Association for Artificial Intelligence (AAAI) Magazine, Vol. 18, No. 3, 1997.
[8]. J. Zhang: “Ontology and the Semantic Web”, Proceedings of the North American Symposium on Knowledge Organization, Vol. 1, 2007.
[9]. Z. Saba Syed, T. Finin, A. Joshi: “Wikipedia as an Ontology for Describing Documents”, Association for the Advancement of Artificial Intelligence, 2007.
[10]. M. Hazman, S. R. El-Beltagy, A. Rafea: “A Survey of Ontology Learning Approaches”, International Journal of Computer Applications, Vol. 22, No. 9, May 2011.