International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 10 | Oct -2016 www.irjet.net p-ISSN: 2395-0072

ONTOLOGY LEARNING

Sonam Mittal[1], Gaurav Sahu[2]
[1]*Assistant Professor, Dept. of Computer Science Engineering, BK Birla Institute of Engineering & Technology, Pilani, Rajasthan, India*
[2]*Assistant Professor, Dept. of Electronics & Communication Engineering, BK Birla Institute of Engineering & Technology, Pilani, Rajasthan, India*

Abstract - *Ontology defines a common vocabulary for researchers who need to share information in a domain. It includes machine-interpretable definitions of basic concepts in the domain and relations among them. Ontology can be defined as a formal explicit description of concepts in a domain of discourse (classes – sometimes called concepts), properties of each concept describing various features and attributes of the concept (slots – sometimes called roles and properties), and restrictions on slots (facets – sometimes called role restrictions). Ontology learning greatly facilitates the construction of Ontologies by the ontology engineer. The method described in this paper shows the ontology learning framework through ontology import, extraction, pruning and modeling.*

Keywords: Ontology, Corpus, Keywords Extraction, Relevance, Ontology Learning Algorithms

1. INTRODUCTION

"Ontology" is a term borrowed from philosophy that involves a branch of philosophy dealing with the nature and organization of reality. An ontology is an explicit, formal specification (i.e. machine readable) of a shared (accepted by a group or community) conceptualization of a domain interest [1]. The meaning of the terms included in the definition is as follows:

- Formal: refers to the fact that ontology must be understandable by the machine i.e. machine must be able to interpret the semantics of the information provided.
- Explicit: means that the type of used concepts and constraints on their using must be explicitly defined.
- Conceptualization: refers to an abstract model of some phenomenon in the world which identifies the relevant concept of this phenomenon.
- Shared: indicates that the ontology knowledge supports consensus and it is not restricted to certain individuals, but not accepted by a group.

In practical terms, developing ontology includes:

- defining classes in the ontology
- arranging the classes in a taxonomic (subclass-super-class) hierarchy
- defining slots and describing allowed values for slots
- filling in the values for slots for instances

Ontologies are built so that they can be shared anytime, anywhere and independently of the behavior and domain that uses them. Ontology learning is the process of instantiating the knowledge base. Ontology learning is concerned with knowledge acquisition.

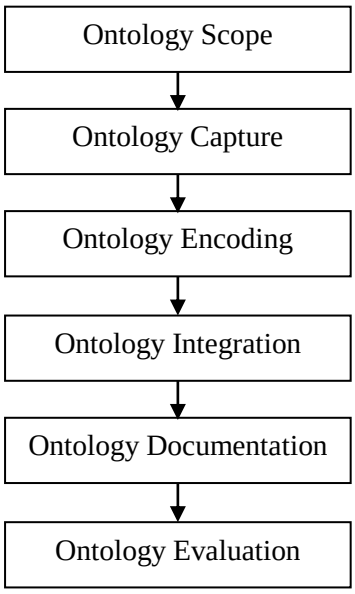

Fig. 1: Ontology Construction Methodology

2. ONTOLOGY LEARNING LAYERS

The ontology learning layers can be represented in the form of a stack [2]. The different layers of this stack

are processed in order to identify the terms (linguistic realization of domain specific concepts), finding out their synonyms, categorizing them as concepts, defining concept hierarchies, relations and describing rules in order to restrict the concepts.

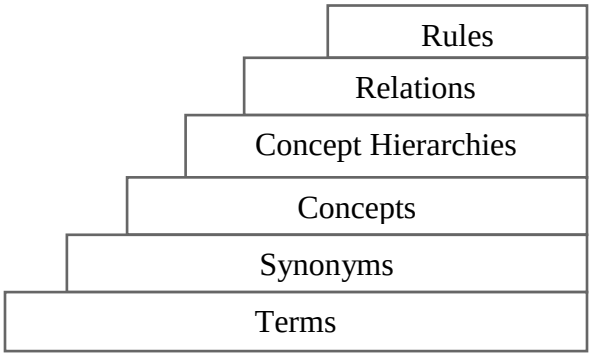

Fig. 2: Ontology Learning Cycle

- **Term:** Linguistic realizations of domain specific concepts
- **Synonym:** It allows the acquisition of the semantic term variations in and between the languages
- **Concept:** It includes the extraction of formal and informal definitions
- **Concept Hierarchies:** It includes extraction of relations of concept properties and relations with other concepts
- **Relations:** It explains the type of association between concepts of the domain
- **Rules:** It is used to infer knowledge in the ontology.

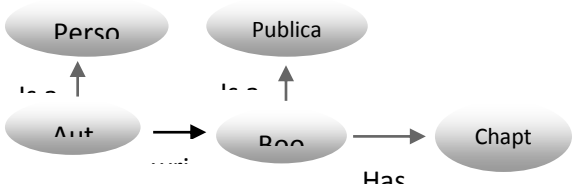

Fig. 3: Example of Small Ontology

2.1. ARCHITECTURE For ONTOLOGY LEARNING

Ontology learning architecture has 4 major components: Ontology Management Component, Resource Processing Component, Algorithm Library Component and Coordination Component.

- **Ontology Management Component:** It is used by ontology engineer for manually dealing with Ontologies. It allows inclusion of existing Ontologies, browsing existing Ontologies, validation [3], modification and evolution [4].
- **Resource Processing Component:** The techniques for discovering, analyzing, importing and transforming relevant input data are present in this component. The most important sub component is the natural processing system [5]. The main task of resource processing component is to generate a set of pre processed data as input for the algorithm library component.
- **Algorithm Library Component:** This component is the algorithm backbone of the framework. Different algorithms are there for the extraction and maintenance of the ontology model.
- **Coordination Component:** This component is used by ontology engineer to interact with the ontology learning components for processing of the resources and algorithm library. Graphical user interfaces are provided to the ontology engineer to select relevant data, apply processing and transformation techniques to start a specific extraction mechanism. Results are merged and presented to the ontology engineer with different views of the ontology structures.

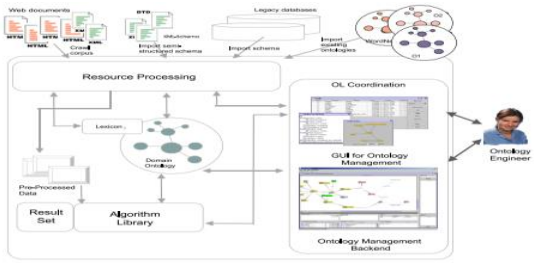

Fig. 3: Ontology Learning Architecture

2.2. ONTOLOGY LEARNING ALGORITHMS/METHODS

There are different ontology learning algorithms. Some of the algorithms are described here. They cover different parts of ontology definition – may be evaluated in isolation of each other [6].

2.2.1. EXTRACTION of TAXONOMIC RELATIONS

The extraction of taxonomic relations is done in various ways.

- Statistics-based extraction using clustering
- Statistics-based extraction using classification

Clustering

Clustering can be defined as the process of organizing objects into groups whose members are similar in some way based on the distributional representation [7]. The different types of clustering methods are:

a. **Agglomerative:** In this phase, each term is defined to constitute a cluster of its own. The larger clusters are iteratively generated by merging the most similar/least similar ones until some stopping criterion is reached.

b. **Partitional:** In the initialization phase, the set of all terms is a cluster. It produces a hierarchical description of terms. In common it uses K-means clustering.

c. **Conceptual:** It builds a lattice of terms by investigating the exact overlap of representing terms between two represented terms.

Classification

Classification can be defined as the refinement of the taxonomy using WordNet [8], by classifying new relevant terms into the given concept hierarchy. K Nearest Neighbor (kNN) and Support Vector Machines are the algorithms used for this purpose. .

2.2.2. LEXICAL ENTRY & CONCEPT EXTRACTION

The simple method for extracting relevant lexical entries that indicates concepts is counting frequencies of terms in a given set (linguistic pre processing) documents/corpus ***D*** This approach, in general, assumes that a frequent term is a set of domain-specific texts that indicates occurrence of a relevant concept. A standard approach is followed for term weighting based on the following measures:

- The **lexical entry frequency** $lef_{l,d}$ is the frequency of the occurrence of lexical entry, where, $l \varepsilon L$ in a document/corpus $d \varepsilon D$.
- The document frequency df_l is the number of documents in the corpus ***D*** that ***l*** occurs in.
- The corpus frequency cf_l is total number of occurrences of ***l*** in the overall corpus ***D***.

Domain	Method	Features used	Purpose
Free Text	Clustering	Syntax	Extraction
	Inductive Logic Programming	Syntax, Logic Representation	Extraction
	Association rules	Syntax, Tokens	Extraction
	Frequency-based	Syntax	Pruning
	Pattern-matching		Extraction
	Classification	Syntax, Semantics	Refinement
	Formal Concept Analysis	Syntax, Semantics	Refinement
Dictionary	Information Extraction Page Rank	Syntax Tokens	Extraction
Knowledge Base	Concept Induction, A-Box Mining	Relations	Extraction
Relational Schema	Data Correlation	Relations	Reverse Engineering

Table 1: Ontology Learning Approaches

3. CLASSIFICATION of ONTOLOGY LEARNING APPROACHES

The ontology learning approaches can be classified into different categories as mentioned in the Table below:

2.2.3 ONTOLOGY PRUNING

Pruning is needed if one adopts a (generic) ontology to a given domain. A frequency based approach is used to determine frequencies in a corpus. Entities that are frequent in a corpus are considered as a part of a domain. To determine domain relevance, ontological entities retrieved from a domain corpus are compared to frequencies obtained from generic corpus. Users can manually select relevance measures for frequency calculation. Ontology pruning algorithm uses the computed frequencies to determine the relative frequency of each concept in the ontology.

CONCLUSION

Ontology learning is introduced as an approach that may help the ontology engineer in construction of the ontology. The overall process is semi-automatic that requires human intervention. It greatly relies on interaction between the human modeler and the learning algorithm for construction of the ontology in a domain.

REFERENCES

[1] Tom Gruber, 1992, What is Ontology, Online,http://www-ksl.stanford.Edu/kst/what-is-an-ontology.html

[2] P. Buitleaar, P. Cimiano, and B. Magnini, Ontology learning from text: an overview: ontology learning from text: methods, applications and evaluation, pp 3-12, 2005

[3] Y. Sure, J. Angele, and S. Staab. OntoEdit: Guiding Ontology Development by Methodology and Inferencing In S. Spaccapietra, S. March, and K. Aberer (eds.). LNCS Semantics of Data, Springer, 2003 (to appear). (Extended version from ODBase-2002)

[4] A. Maedche, B. Motik, L. Stojanovic, R. Studer, and R. Volz: Ontologies for Enterprise Knowledge Management, IEEE Intelligent Systems, December, 2002

[5] Manning, C. and Schuetze, H.: Foundations of Statistical Natural Language Processing. MIT Press, Cambridge, Massachusetts, 1999

[6] A. Maedche and S. Staab. Measuring Similarity between Ontologies. In: Proc. Of the European Conference on Knowledge Acquisition and Management - EKAW-2002. Madrid, Spain, October 1-4, 2002. LNCS/LNAI 2473, Springer, 2002, pp. 251-263

[7] L. Kaufman and P. Rousseeuw: Finding Groups in Data: An Introduction to Cluster Analysis, John Wiley, 1990

[8] G. Miller. WordNet: A Lexical Database for English. *Communications of the ACM*, 38(11), pp. 3941

[9] A. Gomez-Perez, Ontology Engineering, Springer Verlag, 2002/2003

[10] M. Missikoff, R. navigli and P. Velardi: The Usable Ontology: An Environment for Building and Assessing Domain Ontology, Proceedings of the International Semantic Web Conference 2002. Springer, 2002, pp. 39-53