International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-5, January 2020

Development of Background Ontology for Weather Systems through Ontology Learning

Ramar Kaladevi, Appavoo Revathi, Sivakumar Sridhar

Abstract: Background or reference ontology is a common vocabulary for a system to share knowledge and support information integration. Weather system has more domain specific words, which are not fully covered by generic knowledge source like web and WordNet. For example, Temp is a word related with temperature in weather system, this kind meaning is not available in WordNet. Secondly, many new technical and scientific words are used and existing words also carry different senses. Thesauri usually cannot capture these new senses and words in time. Available background knowledge is insufficient to overcome the existing challenges and issues. This paper focuses on developing background ontology for weather system by enhancing existing knowledge bases. Finally the comparison is made between manually developed ontology and semi automatically developed ontology.

Keywords : Knowledgebase, Reference ontology, Semantic web, Ontology mapping.

I. INTRODUCTION

Development of background knowledge is essential for reference, mapping and integration between concepts. The idea is to make use of knowledge from web, WordNet, Linked Open Data (LOD), upper ontologies and domain ontologies [1]. Some of the few existing weather related ontologies such as JPL's Semantic Web for Earth and Environmental Technology (SWEET 2.0), Sensor observation ontology, DAML Weather ontology are still incomplete. To have perfect ontology mapping there is a need for complete and correct background knowledge.

During matching systems evaluations, Ontology Alignment Evaluation Initiative - OAEI as well as individual evaluations in [2] [3] show that lack of reference knowledgebase, especially domain centric knowledge is one of the core problems in matching systems.

Manuscript received on January 02, 2020.
Revised Manuscript received on January 15, 2020.
Manuscript published on January 30, 2020.

* Correspondence Author
Ramar Kaladevi* , Bachelor of engineering degree, Computer Science and Engineering, Bharathidasan University, Trichy.
Appavoo Revathi, Ph.D, Anna University, Chennai.
Sivakumar Sridhar, Master degree, Computer Science and Engineering, Anna University, Chennai

Many domain ontologies are developed in practice. DICE is an ontology developed and maintained for medical analysis purpose. This medical ontology implemented in OWL DL with more than 2000 concepts that are defined by more than 5000 lexically related terms. The concepts are connected to other concepts using 4300 relational links of 50 different relation types. Gene ontology (GO) is a bioinformatics knowledgebase to unify the representation of gene product and gene attributes of all species. The Unified Medical Language System is main reference ontology in biomedical domain. It considers around 20 lakhs lexical names for some 9lakhs concepts collected from more biomedical related libraries and vocabularies, It also uses 120 lakhs relations to relate concepts.

In this paper, a semi-automatic background ontology development method is proposed for weather system and it is compared with manually developed ontology. The paper is organized as follows: Chapter II illustrates the related works relevant to background ontology sources and ontology mapping using background knowledge. Chapter III explains the various methods of ontology development and Chapter IV discusses flow of proposed work with architecture and retrieval results. Chapter V concludes the proposed work.

II. LITERATURE REVIEW

Many ontology mapping and merging methods are depending on the use of the information in the ontologies such as class, subclass, axioms, instance, and property. Other ways are to make use of external background or reference knowledge in the process of matching. The knowledge for reference ontology is gained from different sources. Commonly two methods are used when refereeing background knowledge. In first method input ontologies are matched via background and the second is directly adapting the knowledge of background source.

Semantic web is used as a reference source that is the combination of number of background ontologies useful for the mapping process [4]. The domain ontologies are termed as semantic web for experimental set up. In Zhang & Bodenreider [5], mapping is done using this experimental ontology. Interoperability between domain ontologies is obtained by relating them with corresponding background ontology. In this case entities are matched with background ontology and find matched entities instead compare each other.

An automatic heuristic approach is proposed to identify and effectively utilize the missed knowledge while performing matching operation [6]. Missing background concept or knowledge is found iteratively and a highly matched concept pair is noticed as candidate match and if its entities are not correctly matched, then the remaining sub concepts in the hierarchy below are considered for matching. Identified missing knowledge is included to the existing knowledge source for reuse in further analysis.

LOD is used as background knowledge in BLOOMS technique which uses bootstrapping information present in the LOD cloud [7]. BLOOMS consider schema information of given ontologies for correspondence. It constructs concept based forest in accordance with Wikipedia and the forests are used to reach a decision on which concept names are to be aligned. Post processing is depends on the underlying reasoner and API. The upper ontologies are taken as reference or semantic bridges in the alignment process [8]. It has a systematic analysis of the relationships among mapped ontologies features (no. of simple and composite concepts, ontology depth, top level concepts, stems common english prefixes and suffixes), upper ontologies, mapping algorithms and experiment results. Mora et al. (2013) [9] enhanced the active learning framework of ontology matching proposed in [10]. The results are enhanced by graph propagation algorithm, upper ontologies and user feedback. The limiting factor of this method is the available reference alignments are less for further calculation of evaluation metrics.

Corpus is considered as background knowledge in [11]. This approach increases the mapping accuracy of algorithms by using a corpus of schemas and mappings in a specific domain. Corpus can be used as to increase the evidence of each entity by relating evidence from corpus for relevant entities. Second, statistics about entities and their relationships then use them to infer constraints that are to prune candidate mappings.

Domain ontology is used as background knowledge to map input ontologies in [12]. Anchoring matching technique connects concepts of input ontologies to concepts of background ontology. The input concept can be matched to more anchors in the background ontology reveal more relationships that represent distinct properties.

The above mentioned systems used different domain background knowledge for ontology mapping. These sources are essential for domain related conceptual mapping, integration and information retrieval. Available knowledge source for weather system is not carry all domain specific and lexical terms.

III. METHODS FOR ONTOLOGY DEVELOPMENT

Three main methods have been commonly adapted for acquiring domain or task specific ontologies. One is by integrating available ontologies. Second is by developing ontology from scratch, or by enhancing or extending an available ontology, which usually rely on knowledge derived from a relevant domain. Last method is by generic ontology is to be specialized to adapt it to a particular domain.

Manual ontology development is a time consuming process which requires more effort to acquire domain knowledge and for knowledge domain modelling. There are some advantages and difficulties in this ontology development. The merits are that the expertize level of human and their knowledge of a relevant domain help the creation of the concept hierarchy and the identification of respective properties. It also motivates the inclusion of required concepts into specific position of the ontology. In the same way, the cons are that the manual ontology construction is the complete effort of the ontology designers. Indeed, to construct a complex knowledgebase system, he has to spend lots of time, put effort to read documents/ databases, extract meaningful keywords (concepts) from documents without any conflict. In addition, he has to insert the keywords in proper position of ontology without duplication, properly distinct classes and instances and infers relation between the elements of the ontology. The overall requirements for manual ontology construction are the ontology developer is a specialist in the ontology construction process, needs domain knowledge along with ontology engineering skills and also to be experience in handling ontology development/ editor software.

In order to address the above difficulties many techniques have been proposed, including tools and systems that semi automatically or automatically use machine learning and text mining methods to develop ontologies[13]. These methods utilize dictionaries, thesaurus, unstructured text data, knowledge bases, semi-structured schemata or relational schemata. The process of automated or semi-automated ontology development, enhancement and adaptation of ontologies is termed as ontology learning. Ontology learning can be described by six major subtasks [14]. These are identification of terms (concepts), Synonym notification, Concept extraction, hierarchical and Non-taxonomic relation formation, and Rule acquisition.

IV. PROPOSED ONTOLOGY DEVELOPMENT METHOD

The proposed work follows the second approach by enhancing and extending an available ontology using domain knowledge extracted from concerned sources and experts in a semi-automatic way. The ontology learning steps are used when adding or modifying new concepts. The approach iterates the feedback from users and terms from various sources in the ontology learning task and acquired knowledge from the feedback is updated for further iterations

The existing domain ontology is taken into consideration and has extended the ontologies with relevant concepts, its synonyms, instances and corresponding relations. Conceptual clustering is the approach that collects concepts based on their semantic similarity to construct hierarchies from documents. The semantic similarity between concepts can be calculated by various methods. For instance, it is calculated by term distribution approach: distance between the linguistic terms of words lesser implies they are conceptually similar [15]. Concept clustering is mainly used to cluster terminologically related terms.

After relevant concept identification the ontology engineer selects the concepts and their relations (e.g. hyponyms/homonym) to be modelled. Then senses (meaning) related to the terms are identified. If there are many synonyms for the same word, closest one for the domain is retained. WordNet is a huge English lexical database.

International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-5, January 2020

It groups nouns, adjectives and verbs into Synsets (cognitive synonyms).Synonyms denote the same concept has different interpretations in many contexts. Synsets (totally 117000) are linked to other Synsets by means of a small number of conceptual relations. Other semantic relations such as more general, less general are also identified using Oracles (WordNet). If a concept is more general than another one then there is at least one sense from first concept has a sense in second concept as hyponym or meronym. Similarly, if a concept is less general than another concept then there is at least one sense in first one has a corresponding sense in second concept as hypernym or holonym.

In the conceptual stage, semantic concepts and relations are represented using label of the same in a normalized manner. The task is interactive because responsible expert is finalizing which classes, subclass and properties are essential for the domain ontology and iterative in order to improve accuracy. The process has two stages. In the first stage, documents are clearly analyzed and at the end analysis a new element list is generated and new contents are added into ontology using reverse engineering. At the second stage, the constructed ontology is refined using user queries by adding or deleting attributes to create new entities etc.

The ontology construction work has been based on information extracted and inferred from the online domain related sources and also from the web sources of the important weather systems. For instance, information captured from existing ontologies, lexical database (WordNet) and domain knowledge is from National Oceanic and Atmospheric Administration (NOAA), National Weather Service (NWS), METAR and 40 other weather web systems. The background weather ontology is constructed by reusing the existing ontologies /vocabularies into a coherent structure and extended wherever necessary, using domain knowledge. Background ontology building is the extension of an existing ontology. K-means clustering algorithm is used to group new concepts from corpus. Before constructing an ontology non-redundancy of classes, completeness of the knowledge, consistency between classes and extendibility of ontology (without altering the existing semantics) need to be checked. The ontology is organized in a tree-structured taxonomy. Top down approach is used for ontology development where the most generic classes are first identified and then decomposed into more specialized sub classes. If a class or concept is available with more lexical terms, most relevant term is considered as preferred term, and the remaining terms as synonyms[12].

The documents recovery is obtained by two phase strategy. During first phase, the documents are selected by means of the results retrieved from domain search queries. In this phase, much knowledge on weather concepts has not been identified and the search queries are not specific. Initially few documents of the domain are retrieved using this search strategy. After reading them, more domain knowledge is acquired and that allowed evaluating the relevance of documents found by search queries resulting in change of query strategy becoming more precise. In the second phase, queries using the existing classes inserted into the ontology in order to retrieve more relevant and specific documents. In fact, the process starts browsing a list of websites that are 'authoritative sources 'on weather issues and does a way of 'human crawling' i.e, it starts from the initial page of the sorted sources, recovering other documents and information.

The primary concepts listed are related to weather domain such as 'Temperature', 'Pressure', and 'Wind' and so on. Secondary level concepts are identified with the help of primary concepts such as 'Min temp', 'Max Temp', 'Wind speed', 'Wind direction' etc. Likewise concept hierarchy is formed.

The domain related information is derived from the identified documents and web sources, to enhance the existing ontology using Protégé ontology editor (http://www.protégé.stanford.edu). This editing software is chosen based on its features such as its extensible knowledge model, OWL classes, Properties, Individual, Customizable user interface, Provision to import ontologies in different formats, Support data entry, Ontology authoring and management support, Extensible architecture that enables integration with other application, Availability of Java Application Programming Interface (API), Support to use queries and rules. Knowledge base can be queried with SWRL, SQWRL (Query Language for OWL), and SPARQL.

Figure 4.1 shows snippet of the background ontology, in which the ellipses indicate classes (e.g., 'Pressure', 'Temperature', 'Wind' and 'Precipitation' etc.), the dashed ellipse shows the subclasses (e.g., 'Speed', 'Direction' etc.), the dashed rectangles indicate properties (e.g., 'date', 'time', 'country' etc.), the lines with arrowhead indicate relation between two classes (e.g., 'has Speed', 'has Minimum', 'typeOf', 'is a' and 'has Measurement' etc). The solid rectangle indicates the descriptive information related with class and property such as ID, label and comments. The compound outlined ellipse shows the class which contains more semantically and terminologically related terms (e.g., 'Minimum temperature'). From the available domain knowledge, the background ontology has a total of 232 classes and these were described by nearly 464 lexical terms. To identify the position of concepts in background ontology, position numbers are assigned to background ontology using breadth first traversal technique. A position is assigned to every node in the background ontology. For example root is in position 1 and the children of root as 1.1, 1.2, 1.3 etc. from left to right and the children of 1.1 as 1.1.1, 1.1.2 and so on.

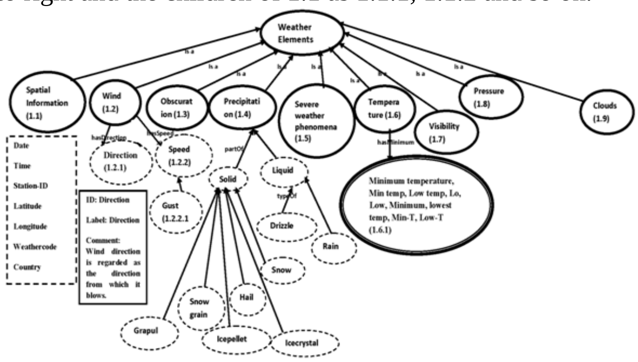

Fig 4.1. Snippet of background ontology

Figure 4.2 also shows the instance of the background ontology. It includes the sub concepts of the concept 'Cloud' such as 'High', 'Low', 'Middle'. Again, the concept 'High' lists its sub concepts 'Cirrus',' Cirrocumulus','Stratus','Cirrostratus'. Similarly, the other concept hierarchies are listed.

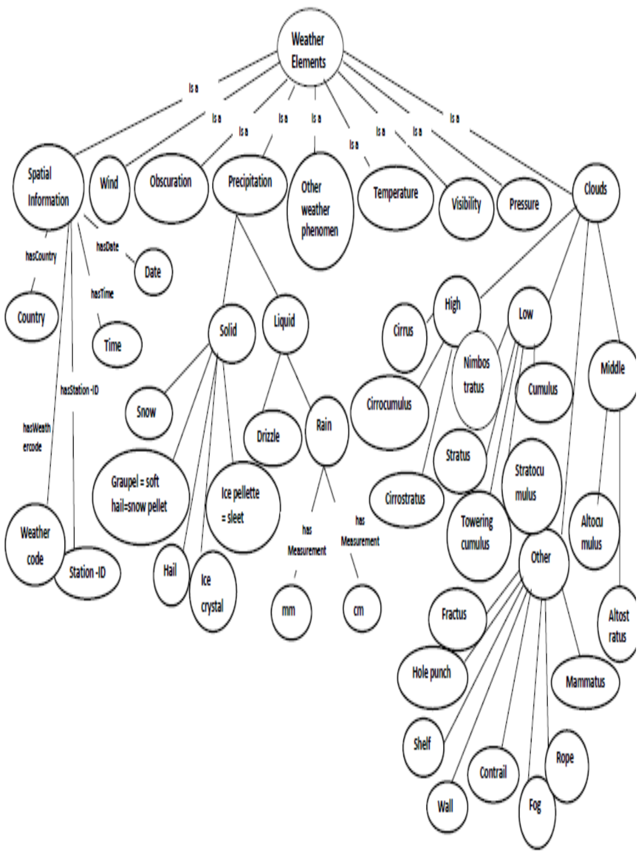

Fig. 4.2. Instance of background ontology

A. Results and Discussions

The semi automatically created ontology is compared with manually created ontology. For that precision, recall, F-measures are used as mentioned in [16], is one of the standard way to evaluate ontology development as well as information retrieval. In this scenario, the precision measures how many correct concepts are developed on the total of the semi-automatically created concepts. Recall reflects numbers of concepts are developed versus total relevant concepts. The performance evaluation is done on the Temperature class i.e on portion of the weather ontology. They are calculated as follow.

Precision (which is a measure of correctness) is shown in equation (4.1)

$$Precision = \frac{No.of Correct Found Mappings}{No.of Total Found mappings} \quad (4.1)$$

Recall (which is a measure of completeness) is shown in equation (4.2)

$$Recall = \frac{No.of\ Correct\ Found\ Mappings}{No.of\ total\ Correct\ Mappings} \quad (4.2)$$

F-measure is defined as a metric that combines both the recall and precision results as given in equation (4.3)

$$F - measure = \frac{2 * Precision * Recall}{(Precision + Recall)} \quad (4.3)$$

Table.4.1. Comparison between manual Vs semi-automatically developed ontologies

Method	Performance Metrics		
	Precision	Recall	F-measure
Manual method	60	78	67.8
Semi-Automatic method	72	84	77.5

Table 4.1 shows the comparison between manual and semi automatically developed ontologies. Reusability and enhancement of existing ontologies using domain knowledge by semi-automatic method contains more relevant concepts with respect to expert's suggestion than manually developed ontology

V. CONCLUSION

Proper information management and retrieval depend on heterogeneities management and interoperability between related systems. In order to meet this requirement, domain specific systems need a proper reference knowledge base. In this paper, background ontology in weather domain has been developed, with all the required concepts, attributes and relations by extending the existing knowledge using related knowledge sources and expert knowledge. Semi automatically developed background knowledge is compared against manually developed background knowledge. Performance results show that semi-automatic ontology development method is more efficient due to the reuse of existing knowledge, domain knowledge and supporting tools than manual development. In enhanced background ontology the class which has more semantically and terminologically related terms is grouped in a single node. It improves retrieval performance and reduces search time.

REFERENCES

1. P.Shvaiko and J. Euzenat, 2013, “Ontology matching: State of the art and future challenges”, IEEE Transaction on Knowledge and Data Engineering, 25(1), 2013, pp.158-176.
2. R. Agrawal, A. Ailamaki, P. Bernstein, et al ,“The Claremont report on database research” ,SIGMOD Record, vol. 37, no. 3, 2008, pp. 9–19.
3. S. Spaccapietra and C.Parent, “Conflicts and correspondence assertions in interoperable databases”, SIGMOD Record, volume 20, No. 4, 1991, pp. 49–54.
4. M. Sabo, M. d Aquin and E. Motta , "Exploring the semantic web as background knowledge for ontology matching”, Journal on Data Semantics, XI,2008, pp156–190.
5. S. Zhang and O. Bodenreider , “Alignment of multiple ontologies of anatomy: deriving indirect mappings from direct mappings to a reference”, In Proceedings of AMIA Annual Symposium, 2005, pp. 864-868.
6. F. Giunchiglia , P. Shvaiko and M. Yatskevich , “Discovering missing background knowledge in ontology matching”, In ECAI, 2006, pp. 382-386
7. P.Jain, P. Hitzler, A. Sheth , K.Verma and P. Yeh P, “Ontology alignment for linked open data” Proceedings of 9th International Semantic Web Conference (ISWC), 2010, 402–417.
8. V. Mascardi, A. Locoro and P. Rosso, “Automatic ontology matching via upper ontologies: A systematic evaluation,” IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 5, 2010, pp. 609–623.
9. M. Mora, R. Ernesto and R. Ichise , “Ontology matching by actively propagating user feedbacks through upper ontologies”, Academic Journal, Vol. 10, Issue 2,2013, pp.85-93E. H. Miller, “A note on reflector arrays (Periodical style—Accepted for publication),” IEEE Trans. Antennas Propagat., to be published.
10. F.Shi , J. Li , J. Tang , G. Xie and H. Li , “Actively learning ontology matching via user interaction”, In Proc. 8th International Semantic Web Conf, 2009, pp. 585–600.
11. J. Madhavan , P. Bernstein , A. Doan and A. Halevy , “Corpus Based schema matching,” 21st International Conference Proceedings on Data Engineering (ICDE), 2005, pp. 57–68.

12. Z. Aleksovski , M. Klein , W.Ten Kate and F.Van Harmelen, “Matching unstructured vocabularies using a background Ontology”, In Proceedings of Knowledge Engineering and Knowledge Management (EKAW), 2006, pp. 182-197.
13. A. Maedche and S. Staab , “Ontology Learning” IEEE Intelligent Systems, Special Issue on the Semantic Web 16(2), 2001.
14. P. Buitelaar , P.Cimiano , M. Grobelnik , M. Sintek , “Ontology Learning from Text”,Tutorial at ECML/PKDD, Porto, 2005.
15. D. Faure and T. Poibeau, “First experiments of using semantic knowledge learned by ASIUM for information extraction task using INTEX”, proceedings of the ECAI Workshop on ontology learning, 2000, pp.7-12.
16. A. Gangemi, C. Catenacci, M. Ciaramita and J. Lehmann , “A Theoretical Framework for Ontology Evaluation and Validation”, Proceedings of SWAP- Semantic Web Applications and Perspectives, 2nd Italian Semantic Web Workshop, University of Trento, Trento, Italy, 14-16 December 2005.

AUTHORS PROFILE

Dr. R. Kaladevi received her Bachelor of engineering degree in Computer Science and Engineering from Bharathidasan University, Trichy. She obtained M.E in Computer Science and Engineering from Anna University, Chennai. She obtained her Ph.D degree from Anna University, Chennai. Her research interest includes information management, Information retrieval and IoT. She is the life member of ISTE.

Dr.A.Revathi received her Bachelor and Master degree in Computer Science and Engineering from Madras University. She obtained her Ph.D degree from Anna University, Chennai. Her research interest includes Network Security, Online Learning Management, Machine learning.

Dr.S.Sridharan received his Master degree in Computer Science and Engineering from Anna University, Chennai in the year 2006. He obtained his Ph.D degree from Anna University, Chennai in the year 2017. His research interest includes Network Security, information management and Cloud computing.