Volume 5, Issue 6, June 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Semantically Structured Ontology based Image Retrieval

Nirmitee N. Kurhekar, Prof. L. J. Sankpal
Department of Computer Engineering,
Sinhgad Academy of Engineering,
Kondhwa, Pune, India

Abstract — Ontology is a set of concepts and their relationships with each other which is used to present an abstract view of any application domain. Domain related information extraction method is used to extract certain kind of information from natural language text by processing them. Ontology extraction has recently emerged as a sub-field of information extraction. Ontology is now accepted as an important component of information systems and processing system. They indicate the particular domain. Ontology which is already constructed can be used again. Use of already constructed ontology offers enormous benefits by measuring and comparing ontology. A fundamental operation is to compute similarity and dissimilarity among ontological entities such that one can establish certain level of association between ontological entities used in different ontology by predefined measures and semantic comparisons. In this paper, Using GDR (Graph Derivation Representation) based Approach, semantic structure of two or more ontology can be compared and measured. The recent Graph Walk Based Method is used to compare and measure the semantic structure of the ontology. It re-ranks the path followed by graph. The graph walk method and weight tuning method effectively model concept relations and re-ranking gives more flexibility in adapting the generated similarity measure to different tasks using high-level and task-specific information. Finally it gives the appropriate image related to the query concept.

Keywords: Ontology, Semantic Structure, Re-ranking, Ontology Measures

I. INTRODUCTION

Ontology can be defined as the Explicit Specification of conceptualization. In other words, we can say that Ontology is a description of things that exist and how they relate to each other. It gives the semantic/linguistic structure of any concept related to the specific domain. In the recent years, web popularity demands for the service which helps user to get relevant information by skipping irrelevant one. We can say that, information retrieval is one of the important things. For proper information retrieval, knowledge representation and knowledge management must be done precisely. For this purpose ontology under natural language processing can be used efficiently. Ontologies and Natural Language Processing (NLP) can often be seen as two similar concepts. An Ontology Model is the classification of entities and modelling the relationships between those entities. The purpose of NLP is to identify the entities and to understand the relationship between those entities.

Ontology is a body of knowledge describing some domain, typically common sense knowledge domain. In the concept of information retrieval, user should retrieve accurate domain-specific information relevant to the query specified. Hence the domain representation is given by ontology. A domain ontology (or domain-specific ontology) represents concepts which is included as part of the world. Particular meanings of terms applied to that domain are provided by domain ontology. It is the graphical structure of domain. It gives the structure of concepts which are semantically associated to each other. Ontology considers the generalize concept in starting and proceed towards more specific concept relevant to the users need.

In the retrieval process, construction of ontology plays an important role. But ontology construction is very difficult and cumbersome job. In ontology construction various algorithms can be used, e.g. Graph Derivation (GDR) Based Approach. To avoid the problem of ontology construction the concept of Use of already constructed Ontology is evolved. This means that the existing ontology of relevant concept can be taken into consideration. Also according to the user's need some modifications can be done in that structure. This will reduce the work and time in generating new ontology due to which retrieval process becomes faster.

Objective is to enhance the performance by applying re-ranking to recover the graph walk results and to overlook the problems of polymorphism of ontology representation and the accumulation of implicit semantic knowledge. Goal of this project is to compare and measure structural similarity of different ontology. By comparing, we are going to find the relevant association of concepts. Also according to the specific relevant concepts generated from the system, images are retrieved.

In this paper, method of using already constructed ontology is proposed. Here, two or more already existing ontology are compared for their similarity and dissimilarity in the hierarchical structure. It starts with the generalise concept which lies at the top of the semantic structure. When other semantically related words or images occurred, they are attached to the structure according to their relevance. Re-ranking using graph-walk method is done according to the comparison of ontology. According to the formatted concept from ontology, Images are retrieved as an output of the system.

II. RELATED WORK

An ontology based measure with deployment of taxonomical features was proposed by A. Maedche and S. Staab without using tuning parameters to weight the contribution of potentially scarce semantic features. It is collection of two popular semantic similarity calculation approaches [3]: feature-based measures and graph-based measures. In the context of computing semantic similarity, adopted a similarity function to determine similar entity classes by a matching process based on synonym sets, semantic neighbourhoods, and differentiating features.

Text classification and clustering proposed a new similarity measure for measuring the similarity between two documents. The method takes three cases into consideration [1]: 1) features of both documents, 2) features of only one document & 3) features of none of the documents.

A cluster-based measure combines the smallest amount path length and the taxonomical depth and gives group for each of the branches in the hierarchical structure with respect to the root node. It explores the new methodology to measure semantic similarity between biomedical concepts using collection of number of different ontologies. It proposed a new ontology-structure-based technique for measuring linguistic similarity in single ontology and across abundant ontology in the domain of biomedical within the framework of integrated medical language system (UMLS) [2]. The proposed method is based on three features: 1) cross-linked path length between two concepts; 2) a new feature of common specification of concepts in the ontology; and 3) confined granularity of ontology clusters.

"Representing ontology using description logics, description graphs, and rules" presented an interesting work that uses the objective of description graph model to explore the problem of insufficient expressivity of describing exactly structured objects in ontological knowledge bases. It proposed an extension of Description languages with description graphs - a knowledge model construct that can correctly describe objects with parts connected in arbitrary ways. Moreover, to make modelling easy the restricted points of structured objects, it extended Description Languages with rules [5]. Deep study of the computational properties of such formalism is presented here. In particular, the sources of undesirability of the general and unrestricted formalism were identified.

"Fuzzy measures on the gene ontology for gene product similarity" presented several recent measures for obtaining the similarity of two products with graph-based ontology terms defined by common taxonomy terms. The fuzzy measure similarity [4] (FMS) has the advantage that it considers the context of both complete sets of annotation terms when computing the similarity between two products. When the two products are not illustrated by common taxonomy terms, it proposes a method that avoids a null similarity result. To account for the variations in the illustration reliability, it proposes a similarity measure based on the Choquet integral. These measures provide additional tools for the biologist in search of functional information for gene products.

"A Graph Derivation Based Approach for Measuring and Comparing Structural Semantics of Ontology" includes the GDR based approach, a three-phase process [6] to alter ontology to its GDR. Also analyze important properties of GDRs based on which stable semantic measurement and comparison can be achieved successfully and compare Graph Derivation Representation based approach with existing graph based methods using a many real world ontology as a training dataset.

Ontology based I R concerns the problem when an end-user is faced with a repository of images [8] whose content is composite and partly unknown to the user. To retrieve digitized art images using ontology based approach proposed a framework for retrieving art images using an ontology-based method. The retrieval scheme makes users more naturally find visual information and experimental implementation demonstrates good potential on retrieving art images in a human-centred manner [7].

"Mining frequent sub graphs" find out the problem of mining uncertain graph data and especially focuses on mining frequent sub graph patterns on an undecided graph database.[10] A new method of uncertain graphs is presented, and the frequent sub graph pattern mining problem is dignified by introducing a new measure, called expected support. The algorithm uses ordered methods to determine whether a sub graph pattern can be output or not.

III. IMPLEMENTATION

EXISTING SYSTEM

Existing System was developed to compare and measure the linguistic/semantic structure of Ontology. For that purpose, GDR (Graph Derivation Representation) Based Approach was used. By hypothetical study of the properties of GDR [6], it was shown that the representation of an ontology is semantic conserving and unique in terms of labels, connecting structure and isomorphism, which guarantees steady semantic ontology measurement. Three conditions were considered according to the GDR based approach as follows:

1. Define the premise of semantic and steady measurement and developed a Graph Derivation Representation, which cyclically transforms ontology to its GDR by number of derivation rules.
2. Initiate two classes of GDR treatments to polymorphism of ontology representation for continual and dependable measurement and comparison of the structural semantics of ontology.
3. Present the prearranged testing of important properties of GDR and experimentally compare GDR approach with existing graph derivation based process in terms of ontology measurement entities over frequent real world ontology as training dataset.

A Graph Derivation Representation based graph isomorphism approach is also used to stably compare the similarity in hierarchical structures of two ontologies.

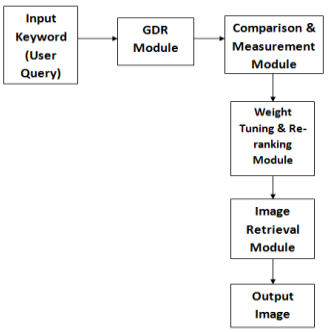
LIMITATIONS

There are mainly two restrictions of existing system; it can't compare the similarity between ontology with uncertainty and toponomy and second, it can be anticipated and improved to its quality for specific application domain of interest. Our proposed system will overcome these restrictions by using various methods like re-ranking, weight tuning etc....

PROPOSED SYSTEM

As proposed system uses the already structured ontology, it reduces time and work which is required in generating a new Ontology. It re-ranks the path covered in graph walk. The graph-based representation is paradigm and it is simple to broaden the graph schema to include other information sources available. While the graph walk and weight tuning successfully model concept relations and associations, re-ranking allows more flexibility.

SYSTEM OVERVIEW



Proposed System will overcome the restrictions of existing system and enhance the methodology in different way. Proposed System will work in four steps:

1. GDR based Comparison

Axioms & assertions are referred to convert empty ontology into the GDR which makes the comparison easy & modular.

2. Graph walk based method

As the name suggest, one can visit each node of the graph using this method. Also it will find the pseudonym similarity by comparing ontology. After the comparison, initial re-ranking is done using path covered in graph walk.

3. Re-ranking

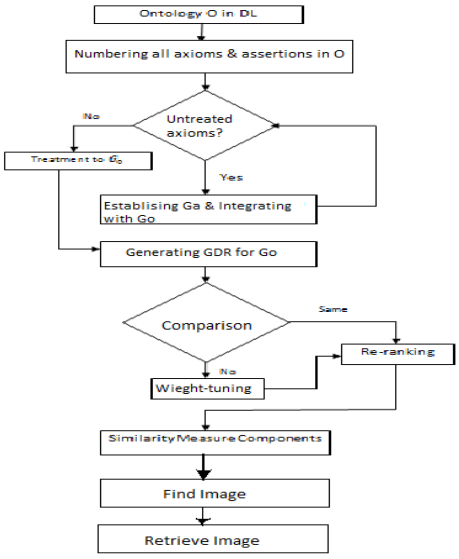
Ontology will be compared to ensure to what extent two ontology are equivalent. If the compared ontology is alike then re-ranking is done to clarify the result.

4. Image Retrieval

Retrieve the image on the basis of contents drawn from Ontology Comparison.

In the proposed System, a novel concept of CBIR will be used. CBIR (Content Based Image Retrieval) is a scheme which uses visual contents to search images from large scale image databases according to users' interests. CBIR includes two main parts: involuntarily semantic concepts extraction and image retrieval. Visual content may be universal or may be domain specific. Universal visual content include colour, texture, shape, spatial relationship, etc. A domain specific visual content, for example faces, is application dependent relative and may involve domain knowledge. Semantic content is obtained either by documented justification or by composite conclusion measures based on visual content. The main advantage of this is it integrates the documentary and the structured elements of the data.

SYSTEM FLOWCHART



IV. ALGORITHMS REQUIRED

1. *Non direct Transitive Relation*

Require: $G_{\mathcal{O}} = (V_{\mathcal{O}}, E_{\mathcal{O}}, \rho, \lambda, \eta)$
$MV \leftarrow \emptyset$
for all $i \in V_{\mathcal{O}}$ **do**
  for all $j \in V_{\mathcal{O}}^{C} \setminus \{i\}$ **do**
    if $(i,j) \in E_{\mathcal{O}}$ **and** $\eta(i,j) \cap \{R\} \neq \emptyset$ **then**
      $MV \leftarrow findIntermVertices(i,j)$
      if MV ! = $\emptyset$ then
        $E_{\mathcal{O}} \leftarrow E_{\mathcal{O}} \setminus \{(i,j)\}$
      end if
    end if
  end for
end for

2. *Elimination of Cyclic Inheritance*

Require: $G_{\mathcal{O}} = (V_{\mathcal{O}}, E_{\mathcal{O}}, \rho, \lambda, \eta)$
$S \leftarrow Circle(G_{\mathcal{O}})$
while S ! = $\emptyset$ **do**
  $v_k \leftarrow \min\{v_i | v_i \in S \wedge 1 \leq i \leq |S|\}$
  for all each $v \in S \setminus \{v_k\}$ **do**
    $\eta(v_k) \leftarrow \eta(v_k) \cup \eta(v)$
    if $(v_k, v) \in E_{\mathcal{O}}$ **then**
      $\eta(v_k, v_k) \leftarrow \eta(v_k, v_k) \cup (\eta(v_k, v)$
      $E_{\mathcal{O}} \leftarrow E_{\mathcal{O}} \cup \{(v_k, v_k)\}$
    end if
    if $(v, v_k) \in E_{\mathcal{O}}$ **then**
      $\eta(v_k, v_k) \leftarrow \eta(v_k, v_k) \cup (\eta(v, v_k)$
      $E_{\mathcal{O}} \leftarrow E_{\mathcal{O}} \cup \{(v_k, v_k)\}$
    end if
    for all each $v' \in V_{\mathcal{O}} \setminus S$ **do**
      if $(v, v') \in E_{\mathcal{O}}$ **then**
        $\eta(v_k, v') \leftarrow \eta(v_k, v') \cup \eta(v, v')$
        $E_{\mathcal{O}} \leftarrow E_{\mathcal{O}} \cup \{(v_k, v')\}$
      end if
      if $(v', v) \in E_{\mathcal{O}}$ **then**
        $\eta(v', v_k) \leftarrow \eta(v', v_k) \cup \eta(v', v)$
        $E_{\mathcal{O}} \leftarrow E_{\mathcal{O}} \cup \{(v', v_k)\}$
      end if
    end for
  end for
  $V_{\mathcal{O}}^{C} \leftarrow V_{\mathcal{O}}^{C} \setminus \{S \setminus \{v_k\}\}$
  $S \leftarrow Circle(G_{\mathcal{O}})$
end while

3. *Image Retrieval*:

input Q; */Q is the query image, segmented into closed contours form
find F_q; */ calculate the feature vector of query image Q
for */ N is the number of images in database having original images
compare Q; */search for the most similar images in database with ID & Components */ (compare number of components and feature vectors)
report similarities; */report partially or completely similar cases.
retrieve image;
end;

V. EXPERIMENTAL SETUP

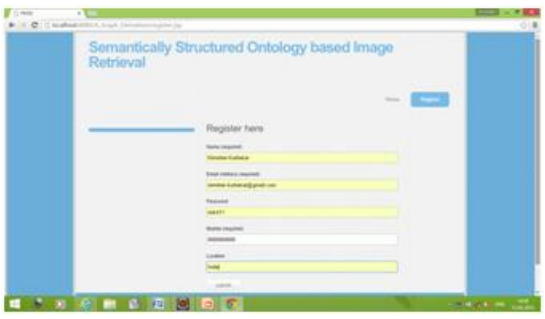
Dataset: The experiment can be applied on actual data set. Actual data set is the assortment of already constructed ontology. Divide the actual data set into a training set and a testing set. Generate the GDR (graph derivation representation) structure using the training data and then compile it on the testing data. Retrieval of the images is done in accordance with the concept by using training dataset.

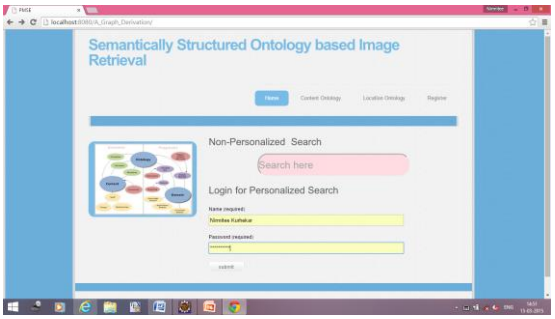
Evaluation Procedure: In this approach firstly similarity between given two or more ontology is evaluated. So that it is applicable for further processing to get appropriate concepts. After that algorithms are applied for weight tuning and to find out the ranking of a concept by traversing the graph. Then similarity is again calculated by comparing distance between same concepts in two different ontology.

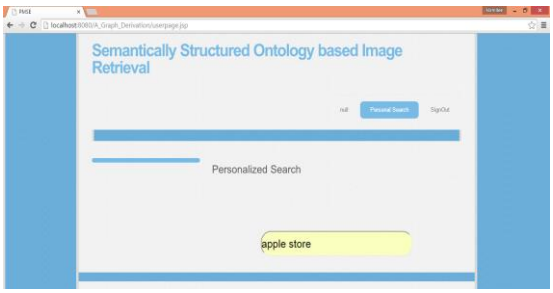
For the comparison, maximal common sub graphs are taken into consideration.

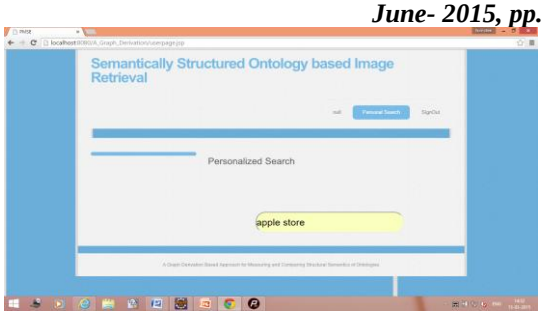
From the training dataset, concepts are compared which are outcome of the similarity measure. These concepts will be used for the retrieval of images. These images are the actual outcome of proposed system.

VI. RESULTS









The experimental results for ontology comparison are based on ontology's distance similarity. Distance Similarity represents same domain with same knowledge if distance is 0. It represents different domain with different knowledge if distance is 1. It represents partial semantic knowledge with overlapped domain if distance is between 0 and 1. According to these values similarity is checked and concept is drawn from which images are retrieved.

VII. CONCLUSION AND FUTURE WORK

A GDR derivation based approach is used to stably measure and compare ontology. By theoretical analysis, the GDR of ontology guarantees stable semantic ontology measurement as it is semantic-preserving and restricted in terms of labels, connecting structure and isomorphism. A general system structure is developed for suggesting adaptive similarity measures in dissimilar data represented as an entity-relation graph. It is developed on graph-walk based examples that generate measures of structural similarity between entities in the graph. Here to rank graph nodes that can use high-level information about the graph walk process, re-ranking is suggested. In particular, general features are proposed that describe the set of paths traversed in reaching a target node from the query of the user.

This paper can be extended with the following features. The system can be enhanced by including more specific image based features and work will be leads to develop a method to improve the outcome of system.

ACKNOWLEDGEMENT

I'm thankful to many persons who contributed to the completion of this research. Particularly I like to thank my guide Prof. L. J. Sankpal to help in research and Head of department Prof. B.B. Gite to allow me to continue this topic. Lastly I would like to Department of computer engineering, Sinhgad Academy of Engineering, Kondhwa to share their knowledge with me during my research work.

REFERENCES

[1] Ke Lu, Beihong Jin, Yinglong Ma, Ling Liu and Xiangjie Liu. "A graph derivation based approach for measuring and comparing structural semantics of ontology." IEEE Transactions On Knowledge And Data Engineering, Vol. 10, 2014.

[2] Jung-Yi Jiang, Yung-Shen Lin and Shie-Jue Lee. "A similarity measure for text classification and clustering." IEEE Transactions On Knowledge And Data Engineering, vol. 26, July 2014.

[3] A. Maedche and S. Staab. "Measuring similarity between ontology" Proceedings of EKAW02, 2002.

[4] J.M. Keller, M. Popescu and J.A. Mitchell. "Fuzzy measures on the gene ontology for gene product similarity," IEEE Transactions, vol.39,, 2006

[5] Yung-Shen Lin, Jung-Yi Jiang, and Shie-Jue Lee, "A Similarity Measure for Text Classification and Clustering" IEEE Transactions on knowledge and data engineering, vol. 26, 2014.

[6] H. Al-Mubaid and H. Nguyen. "Measuring semantic similarity between biomedical concepts within multiple ontology". IEEE/ACM Trans. Comput. Bio. and Bioinfo., vol.3,, 2004.

[7] Wen Gao Shuqiang Jiang, Tiejun Huang. "An ontology-based approach to retrieve digitized art images," Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, 2004.

[8] Avril Styrman, Eero HyvAonen, and Samppa Saarela. "Ontology-based image retrieval." University of Helsinki,, 2004.

[9] D. Fensel, "Ontology-based knowledge management," IEEE Comput., vol. 35, no. 11, pp. 56–59, Nov. 2002

[10] Z. Zou, J. Li, H. Gao, and S. Zhang, "Mining frequent subgraph patterns from uncertain graph data," IEEE Trans. Knowl. Data Eng., vol. 22, no. 9, pp. 1203–1218, Sept. 2010.

[11] A. Doan, J. Madhavan, P. Domingos, and A. Y. Halevy, "Learning to map between ontology on the semantic web," in Proc. 11th Int. Conf. WWW, Honolulu, HI, USA, 2002, pp. 662–673

AUTHOR PROFILE

Nirmitee N. Kurhekar received the B.E. degree in Computer Science & Engineering from H.V.P.M's College of Engineering & Technology, INDIA in 2012 and is pursuing M.E. in Computer Engineering at Sinhgad Academy of Engineering, Pune.

Prof. L. J. Sankpal received the B.E. degree in Computer Engineering from Marathwada University in 1991, M.E. degree in Computer Engineering from Shivaji University in 2001 and persuing Ph D from Bharti University. She is currently working as a Assistant Professor in Sinhgad Academy of Engineering, Pune.

