RESEARCH PAPER | Computer Science | Volume : 3 | Issue : 11 | Nov 2013 | ISSN - 2249-555X

Optimizing The Multi Ontologies Using A Hybrid Approach

KEYWORDS: Semantic index, hybrid ontology, similarity measure, user profile, tf-idf, probabilistic model

S.Vigneshwari
Research Scholar, Department of Computer Science and Engineering, Sathyabama University, Chennai, Tamilnadu, India

Dr. M. Aramudhan
Associate Professor and Head, Department of Information Technology,Perunthalaivar Kamarajar Institute of Science and Technology, Karaikal, Tamilnadu, India

ABSTRACT *Ontological optimization is very much important while developing ontologies. There are various mathematical models available for optimization. One such model is a probabilistic hybrid model , which integrates the mutual similarity measures of the local instance repository and the world wide web. A framework is proposed and the models are evaluated against time based and precision based metrics. The results prove that the hybrid model behaves peculiar and yields fruitful results.*

INTRODUCTION

Currently the research methods are widely keyword based. Sometimes query documents can be used for searching the relevant ones. However, the query document based search is more complicated and time consuming one. The document search and retrieval techniques use vector space models, latent semantic indexing or probability models. In the vector space model, term frequency (tf) and inverse document frequency (idf), is used for weighing terms, where tf refers term frequency and idf refers inverse document frequency. Latent Semantic Indexing is used for encoding the document semantics. Usually incident matrix of the documents and their terms are mapped into a latent representation. Then emerged the PLSI (Probabilistic Latent Semantic Indexing) approach, which implies the probabilistic models on the document vector. Afterwards, language based models were used for document clustering and search. Relying only on the term frequency of the document alone is unworthy, since it is lengthy and wont yield optimized results. In the proposed technique, a combination of vector space approach and probabilistic approach has been applied.

RELATED WORK

- Jung et al(2010), suggested the matching of concepts between two different ontology using WordNet. Different types of matching techniques have been applied here. They are Concept matching property matching, Logical Inference matching etc. These techniques have to be first implemented in a single ontology, then can be done between two different ontologies.
- X.Tao et al(2011) created a model which uses a World Knowledge Base (WKB) and user local repositories in order to capture user history and information needs of the user. This model is a contribution to the web information gathering system.
- Janez et al(2005) prepared a survey report on ontology evaluation techniques. In this work, a level wise evaluation has been done based on the ontologies. There is a variety of ontological evaluation levels like data level, taxonomy level, context level and syntactic level are done in this paper.
- Peter D. Karp et al,(1999), suggested XOL ,a XML-based ontology exchange language, which is best suited for sharing ontologies in a distributed system. The XOL can also be used for translating the SQL query of a relational database into XQuery of XML database.
- Vigneshwari and Aramudhan(2013), proposed a technique to extract the interesting measures using ontology mining. Here the balanced mutual information is used, to find out the similarity between two concepts in the same ontology.
- Yanhui and Chong (2010) proposed a mechanism to integrate ontologies in a multi ontology database. A framework for ontology integration, which combines both ontology similarity measures and ontology integration algorithms, was suggested. Then the integrated ontology was evaluated and checked for consistency.
- Ankayarkanni and Leni(2013), used an optimized BPN classification algorithm for classified satellite images. The same hybrid learning approach can also be applied for classifying the text.
- Mary Posonia and Jyothi(2013), proposed a semantic approach to group the XML documents without nested and repeated elements with respect to a fixed threshold.

MATERIALS AND METHODS

Local ontology (LO): The automated ontology called local ontology, which is generated from the used developed queries. Sometimes, this ontology may also be called as User Profiling Ontology, since it majorly concerns with the concepts, searched by the users. Local ontology is generated from the queries collected in the local database.

$$\text{Tf-idf(term, document, Document sets)} = \text{tf(term, document)} * \text{idf(term, document)} \quad (1)$$

Here tf refers to the term frequency, which is the number of times a term appears in a document. idf refers to the inverse document frequency, which means the number of documents that contain the particular term.

Global ontology (GO): The automated ontology constructed based on the documents available from the web. Global ontology is generated from the data available in the web.

Hybrid Ontology: The hybrid ontology is concerned with the mutual information between the concepts in the local and the global ontology. The hybrid ontology comprises the data from both the LO and GO.

Fig 1 represents the generation of hybrid ontology. Here the query document is mapped onto the local repository. If the match is not found then it is verified from the global repository and updated in the local repository. The corresponding ontologies, which are automatically

Volume : 3 | Issue : 11 | Nov 2013 | ISSN - 2249-555X

generated, are local ontology and global ontology. The optimized ontology, which emerges from the local and global ontology, is the hybrid ontology and it is developed based on the mutual information among the documents.

CALCULATION OF MUTUALLY SIMILAR INDICES

$M_{iGO}(C_i,C_j)=\log$ (2)

In equation(2) , M_{iGO} refers global mutual information index of global ontology. C_i ,C_j refers to two different concepts in the global ontology.

$M_{iLO}(C_i,C_j)=\log$ (3)

In equation (3), M_{iLO} refers global mutual information index of global ontology. C_i ,C_j refers to two different concepts in the local ontology.

$M_{ihybrid}(C_i,C_j)=\log$ (4)

In equation (4) ,$M_{ihybrid}$ refers hybrid mutual information. $C_{i\,(LO)}$, refers to i^{th} concept in the local ontology and $C_{i\,(GO)}$, refers to i^{th} concept in the global ontology.

$P(C_i) =$ (5),

where is the weight of the concept C_i in a particular document D_i , which is calculated based on the tf-idf values. , refers to the weight of the concept C_i in all the documents which are involved.

ALGORITHM FOR SIMILARITY SEARCH

Let K be the concept to be searched, LO be the Local ontology, GO be the global ontology and MI be the mutual information index.

Searching the repositories

IF K appears in LO

Then return

Else if K appears in GO

Then

Update K in LO also

Return.

End if

Generate the search time in milliseconds

Generate the memory usage in bytes.

Searching the Hybrid ontology

Generate the Hybrid ontology based on MI index

$M_{ihybrid}(C_i,C_j)=\log$

Generate the search time in milliseconds

Generate the memory usage in Kilobytes.

EXPERIMENTAL SETUP

On comparing the search time between the individual ontologies and the hybrid ontology, the search time is very less in the hybrid ontology when compared to the other case.

TABLE – 1
EVALUATION OF SEARCH TIME, MEMORY USAGE AND PRECISON OF LOCAL, GLOBAL AND HYBRID ONTOLOGIES

	LO	GO	Hybrid
Search Time(ms)	.5	.8	.4
Memory Usage (Kilobytes)	1.8	2.2	1.9
Precision rate(100 docs)	.58	.43	.62

Precision= (6)

Equation 6 gives the formula to calculate the precision rate, and this will be helpful for evaluating the search.

CONCLUSION

Hybrid ontologies reduce the search time and also optimize the search memory usage. The precision rate of the optimized hybrid ontology is also good. Utilizing both the Vector Space mechanism and the probabilistic approach, the hybrid mechanism works better in a cross ontology. On the same hybrid ontology after establishing the relationships among the concepts, query reformation techniques can be applied for efficient information retrival, which is considered to be the future part of this work.

REFERENCE

[1] JungAe Kwak, Hwan-Seung Yong(2010). "Ontology matching based on Hypernym, Holonym and Meronym Sets in WordNet. International Journal of Web and Semantic Technology,Vol 1, No. 2. | [2] X. Tao, Y. Li, and N. Zhong(2011) ,"A Personalized Ontology Model for Web Information Gathering", IEEE Transactions on Knowledge and Data Engineering, 23(4):496—511. | [3] Janez Brank , Marko Grobelnik , Dunja Mladenic(2005). "A survey of ontology evaluation techniques", In Proceedings of the Conference on Data Mining and Data Warehouses , SiKDD 2005 | [4] S.Vigneshwari, Dr.M.Aramudhan(2013). "Strengthening ontologies by mapping similarity indices with semantic relationships".Proceedings of the International Conference on Computer Networks and Information Technology, June 29,2013, Bangkok, Thailand,PP:60-63 | [5] Heasoo Hwang, Hady W. Lauw.Lise Getoor, AlexandrosNtoulas(2012). "Organizing User Search Histories".IEEE Transactions on Knowledge and Data Engineering, 24(5): 912-925. | [6] Yanhui Lv ,Chong Xie(2010). "A Framework for Ontology Integration and Evaluation." Intelligent Networks and Intelligent Systems (ICINIS). Proceedings of 3rd International Conference on Nov. 2010 Page(s): 521-524. | [7] Peter D. Karp, Vinay K. Chaudhri, and Suzanne M. Paley(1999)."A Collaborative Environment for Authoring Large Knowledge Bases". Journal of Intelligent Information Systems,Volume 13 Issue 3, Nov.-Dec. 1999 Pages 155-194. | [8] B. Ankayarkanni , Ezhil Sam Leni (2013). "Classification of High Resolution Satellite Images using Object based Segmentation".In the proceedings of NC Event 13,August 2013. | [9] A. Mary Posonia, V.L. Jyothi(2013)." Structural – based clustering techniques of XML documents".International Conference on Circuits, Power and Computing Technologies ICCPCT , Page(s): 1239 - 1242 |