

Metadata-Based Semantic Query in Relational Databases

Ch. V. S. Satyamurthy¹(✉), J. V. R. Murthy², and M. Raghava¹

¹ CVR College of Engineering, Hyderabad, India
satyamurthy@cvr.ac.in, satyamurtycvs@yahoo.co.in, raghava.m@cvr.ac.in
² JNTUK, Kakinada, India
mjonnalagedda@gmail.com

Abstract. The retrieval of data from various semantically equivalent databases having different schemas is long been an important issue. In this context, the proposed WordNet-based model demonstrates the semantic data retrieval capabilities from different databases using metadata available with them and publishes the results.

Keywords: Semantic query · WordNet · RDBMS · Metadata

1 Introduction

Relational databases are playing more important role in industry. The data collected are stored in a well-designed schema of the database. Different organizations expected to follow different schema's despite the data presented in these different schemas of these databases are likely to be semantically equivalent [1–4]. Consolidation of the information pertaining to different organizations and generating a comprehensive report from the consolidated data is a challenging task in the absence of proper metadata processing. Clear indexing and preparation of mapping tables us as a collaborative activity of domain experts of the respective organizations, and it is typically one of the classical ways to retrieve valid information from the databases. However, these approaches will have subjective influence and more error prone when original designer or design document is not available. Hence, it is imperative to develop adoptive methods for addressing the semantic gaps in the query and the data present in the databases.

This paper proposes a framework which prompts the query from the user in a raw form, explores the metadata of respective databases using Semantic WordNet, and builds the query automatically. The results of the automatically constructed query prove to compensate the naive knowledge of the user on the data model.

1.1 Related Work

Jarunsree Salee and Veera [4] used query-graph-based approach to extract data from the relational databases. The graph consists of relational lists, attribute lists, joining conditioning, and selection condition. Using this, ranking condition applies to tuples in the database. Jyotimor et al. [5] concerned with semantic query optimization using inductive learning approach, and they concentrated more on join order and parallelization of query. The inductive learning approach is implemented in SQL using SQL hints. This keyword-based query processing in relational databases returns tuples as a connected components based on the way they are associated. DBXplorer [10] and DISCOVER [11] implemented the Candidate Networks approach and BANKS [12] applied the Steiner Tree approach. These approaches have some drawbacks. Mariana Soller Ramada, Joao Carlos da Silva, Plínio de Sa Leitao-Junior [6] implemented query-based approach by semantically analyzed before applying to databases for this they computed intrinsic weight computation, using synonyms as keywords, Weight normalization for sub-matrix and proximity between the keywords. Lipyeow Lim, Haixun Wang, Min Wang [9] implemented query as a graph structure using ontology-constructed graph structure. The graph consists of nodes as concepts which are interlinked. Using this extract, the features and from these features they are learning semantic queries using SVM classification technique.

2 Proposed Model

WordNet [7] is a thesaurus for the English language based on psycholinguistics studies and was developed by Princeton university. It consists of a set of interconnected nodes as concepts and the links connecting the nodes as various types of relations between the concepts, such as synonymy, homonymy, holonymy, and hypernym. It contains lexical semantics relation of the words called synsets. The synsets are very helpful for obtaining lexical items with similar meaning. For example, a word father and a word begetter are grouped in the father, begetter and father also grouped in don as the godfather.

2.1 Frame Work

In this section, we will present the overall view of the model as shown in Fig. 1 and subsequently details of the model. The user interacts with graphical user interface, which is realized by the first layer of the model. The users' keyword is converted into SQL queries using the WordNet as this is the second layer of the model. Then in turn, the access to various databases is implemented through the corresponding data model interface, retrieves the information, and presents it to

the end user. This is the third layer of the model. In our model, the first layer prompts the user to enter the keyword, and using that keyword, it interacts with the WordNet API of English language. WordNet is responsible for the retrieval of the synonyms of the keyword. The WordNet will be acting as a mediator to convert the raw SQL query into semantically equivalent multiple queries based on the synonyms and is evaluated sequentially until one query instance is successful. The synonyms extracted from the WordNet are considered to generate SQL query of the keyword, for each word extracted from the WordNet. Then it checks the database metadata for the availability of a table with the synonym. If the check is successful, then it constructs the query based on the synonyms also in a sequential manner and retrieves the data. It repeats the same exercise for every synonym until it either succeeds or ends up with an error.

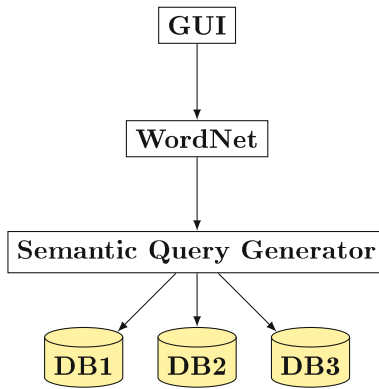


Fig. 1. Proposed framework

2.2 Graphical User Interface

It allows user to enter the keyword and displays the returned results.

2.3 WordNet

This contains lexical data relationships of English. Using this, we can extract synonyms of the given keyword.

2.4 Query Generator

This takes the keyword or synonym of the keyword and generates the query corresponding to the word.

2.5 Database(s)

The databases of different institutions contain semantically equivalent data of the keyword at different formats.

3 Experimental Results

We considered educational domain as our interest, and the institution we consider has many departments. Each schema of the database maintains data in different formats, and the names of the tables also may look different but are semantically equivalent. The keyword we have considered for experimentation is ‘student’ and the extracted synonyms to it from the WordNet such as scholar, book man, and Educated are considered for generating the find related SQL query. The keyword student is given in the graphical user interface which is shown in Fig. 2.

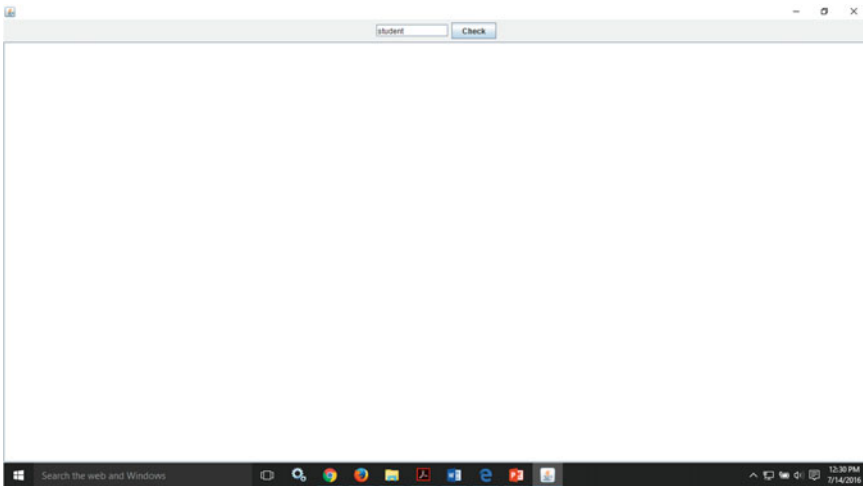


Fig. 2. Keyword specification by the user through the GUI

For each of the keyword, the model which searched the database metadata and the table corresponding to the synonym is accessed. The sample results of a query related to the keyword student is shown in Fig. 3.

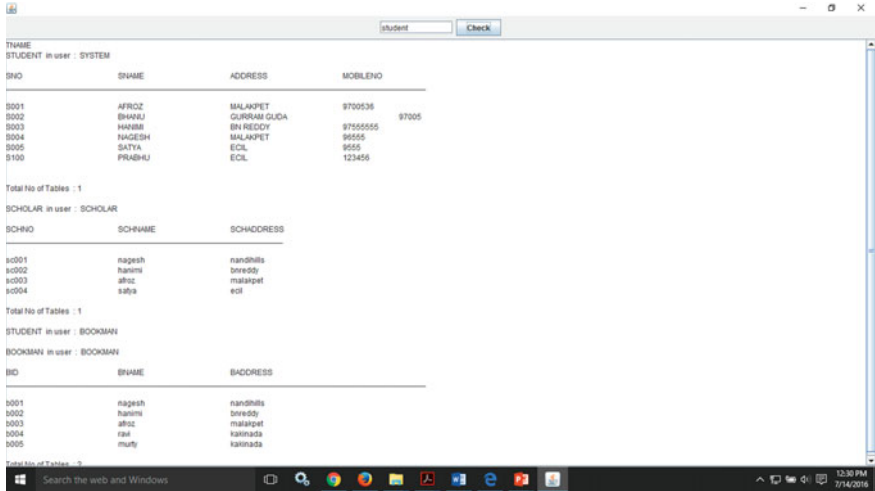


Fig. 3. Experimental results showing various semantically equivalent information retrieved from the database

4 Summary

This paper describes the architectural framework and each component of the framework. It describes middle layer about WordNet or dictionary to know about synonyms. Using this layer, a semantically equivalent query is generated for the input query and helps to retrieve the data from various databases. The results of the generated query processing are presented at a centralized location.

References

1. Kumar, P., Mohan.,Vaideeswaran, J. Semantic based Efficient Cache Mechanism for Database Query Optimization, International Journal of Computer Applications, Volume 43 No. 23, page(s): 14–18, April 2012.
2. Saini, M., Sharma D., Gupta, P., K. Enhancing Information Retrieval Efficiency Using Semantic-based-Combined-Similarity-Measure, International Conference on Image Information Processing (ICIIP 2011), IEEE Computer Society, 2011.
3. Hsu, C., Knoblock, A. Semantic query optimization for query plans of heterogeneous multidatabase systems, Knowledge and Data Engineering, 12(6):959978, 2000.
4. Jarunee, S., Veera, B., A Metadata Search Approach to Keyword Query in Relational Databases, IJCA vol. 69 may 2013
5. Jyoti, Mor., Indu Kashyap ., RK pathy. Implementing Semantic Query Optimization in Relational Databases, IJCA vol. 52 No 9 Aug 2012.
6. Mariana, S, R., Joao, C, da, Silva. Data Extraction from Structured Databases using Keyword-based Queries, 29th SBBD, ISSN 2316-5170 , October 2014
7. Miller, G. Nouns in WordNet: A Lexical Inheritance System, International Journal of Lexicography, vol. 3, no. 4, 1990.

8. Majid Khan and Khan, M., N., A. Exploring Query Optimization Techniques in Relational Databases, *International Journal of Data-base Theory and Application* Vol. 6, No. 3, June, 2013
9. Lipyew, L., Haixun, W., Min, W. Semantic query by Example, *EDBT 13*, March 18-22 2013.
10. Agrawal, S., Chaudhuri, S., and Das, G. Dbxplorer: A system for keyword-based search over relational databases. In *ICDE*, pages 516. IEEE Computer Society, 2002.
11. Hristidis, V., and Papakonstantinou, Y. Discover: Keyword search in relational databases. In *VLDB*, pages 670681, 2002.
12. Aditya, B., Bhalotia, G., Chakrabarti, S., Hulgeri, A., Nakhe, C., Parag, and Sudarshan, S. Banks: Browsing and keyword searching in relational databases. In *VLDB*, pages 10831086, 2002.