Metadata Based Semantic Query In Indian Language Databases

Ch V S Satyamurty[1], Dr J V R Murthy[2], M Raghava[3]

1 ,3CVR College Of Engineering, Hyderabad
2 JNTU, Kakinada
*Corresponding Author: Ch V S Satyamurty

ABSTRACT

The information retrieval from semantically equivalent relational databases is a major challenge to the end users. More over the data retrieval from different languages and presenting to the knowledge workers is a major challenge. The query can be translated to various languages   from native language to the target language and retrieving the data from various database schemas which are structurally different is a problem. This paper proposed a framework with such capability and displaying the results using the translator, wordnet and domain knowledge databases.

KEYWORDS: Metadata, RDBMS, Semantic query, Structured query, Student, Translator, Wordnet.


I.   INTRODUCTION

The relational databases(RDBMS) play a major role in this information age. The RDBMS are using in numerous applications across various domains. The repetitive tasks can be implemented   through various technologies and tools available and producing enormous amount of data. The data generated by these applications are stored in a database. The databases will store and retrieve data more efficiently and effectively. The standards made by Unicode consortium helps data can be stored in various languages. The database vendors brought the feature of internalization and localization capabilities to cater the needs of various native languages. This helps the various departments of larger organizations to store data in their local native language databases. The schemas of these databases created by various departments are differ in structure but semantically they are equivalent. The retrieving the data from these semantically equivalent multilingual databases [1] data and presenting to the end user is not a straight away using a query step [2-4,6-7]. The metadata available in these database repositories is helpful to retrieve the data efficiently. The wordnets [5] are available   for most of the languages. The language wordnet helpful to find the semantic equivalent words of the given word. The task of retrieving the semantic data from   different Indian language databases is a topic of interest by researchers. This paper proposes a framework which accepts keyword from the end user and translates in to various native Indian languages using the industry standard tools. The translated keyword can be fed to a wordnet to find the different semantic equivalent words. Each word will be taken as an input look at the metadata of the database in search of a table for the translated word and tries to interpret the structure of the table. Once a suitable table is find then generate query and execute the   query, retrieve the information of interest. The results of the automatically generated query helpful the naïve users of their knowledge of the data model. The end user become aware of the centralized results.

II.   RELATED WORK

The authors in [8] presented a model to retrieve the information from of English language databases which are semantically equivalent and reported the results at the centralized location. The authors [9] used protégé tool to create sports ontology in the first step. Next with the help of Stanford parser to build the query tree of the user input. Using the wordnet find the semantic equivalent words for the key word present in the parse tree and generate semantic query and retrieve the data from the sports domain. The authors [10] used wordnet to explore equivalent semantic words of metadata which extracted from query logs. That helps the single query can be expanded as multiple semantic equivalent queries to extract the data from databases. The authors in [11] used key-mantic the natural language technique for processing of information using the key word from RDBMS. The key-mantic generates semantic words of the given keyword using the intrinsic weight matrix and aggregate functions to find the proximity of words to rebuild the query as semantic equivalent queries. In [12] the authors used Pseudo relevance feedback method for query expansion using wordnet to find semantic key words and

---

extract information from informational retrieval systems. In [13] The authors suggested method to implement wordnet for Malayalam language using MySQL data base.

III. PROPOSED MODEL
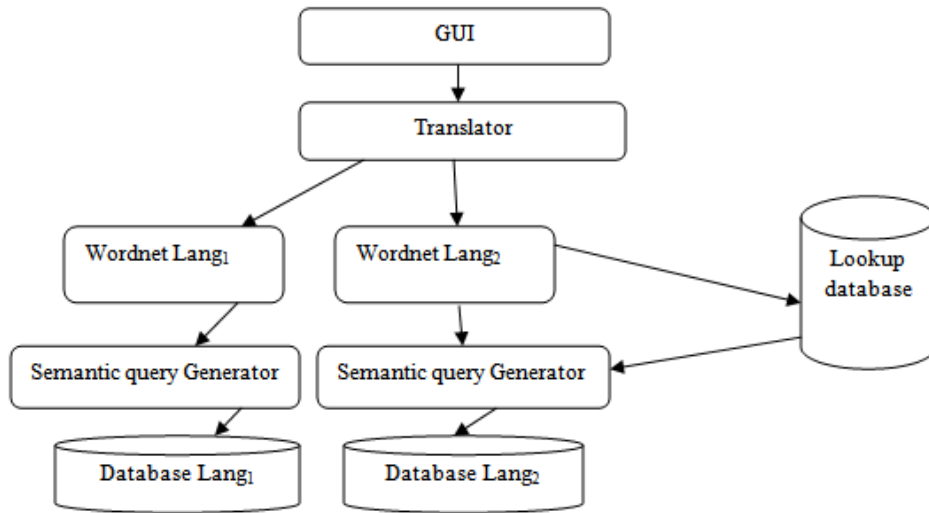
Fig.1 The Proposed framework.

The transactional data bases are developed using RDBMS data model, these databases are no longer confined to many operational inconsistencies across many manipulations. The designers of these databases done on the structure of the data model. The structure of the data model has an influence to derive some inference. Hence, the language of the schema does not have any manner in information extraction, if the schema is the same. The idea directed the designers of the database to include the localization and internalization issues in databases. The third-party software helps the end-user to retrieve the data from multi language. In the proposed model the end user works with the native language without any worry to the language of the data base implemented. The query generation is a straight forward step replacing the key word of the source query mapped on to the semantically equivalent target language query. The word net is helpful to find synonyms for the given input word. Wordnet is a lexical database for the language chosen. The wordnet have synsets, hypernym, holonomy, homonymy. The synsets are useful to find the synonyms for the given word, which are semantically equivalent.

3.1 Framework
The detail description of the proposed framework in Figure -1 given in this subsection. The first layer of the model is graphical user interface where user can enter the input keyword. The second layer is the translator which translates the keyword into different Indian languages. Occasionally there is a chance to given word may not be mapped to the target language In those cases the third layer consists of wordnets of native language which helps to find the semantically equivalent words for the given key word. Each semantic word of the wordnet is validated using the look up database and the semantic word does not exist then the word will be discarded. The semantic words given by wordnet are checked with the metadata of the target database to find the table to generate equivalent semantic query. Using the generated query access the data of the various data models of the target language and retrieve information from these data models and presented to the end user.
Graphical user interface
The interface is prompts the end user to input the keyword of the information extraction and display the results of the queries executed to the end user.
Translator
This layer translates the input key words into the multilingual words using the Microsoft translator. The translator finds the equivalent word in target language maintained. Then the model validates translated word with the lookup database. Then the metadata of the target database to find the table for the translated word and the availability of the table name for corresponding to the translated word. It is evident that the database schema consists of semantically equivalent information then next step is involved.
Wordnet

The next layer is wordnet which is a lexical database for the language chosen and extracts the synonyms of the given keyword. The wordnet acts as mapper to map the given input keyword into equivalent words which are semantically equivalent words.

Query Generator

Then the framework checks the target language database   metadata for the availability of the table with the synonym and recreate the query to execute in the target language database. If the search for each keyword is success, then framework generates the query corresponding to the given word.

Databases

The databases of different institutions holds the semantic data of the key word in different structures.

IV. EXPERIMENTAL RESULTS

The proposed experiments done for the education domain. The data stored in the databases of the different institutions across India. If the user interested in student information maintained by different colleges in their local languages. The translator translates the given input key word in to three target languages namely Telugu, Hindi and Kannada. The fig 2 the input word is 'छात्र' in hindi the synonyms for   the given keyword are विद्यार्थी   , शिष्य   , चेला   , अनुयायी   . The keyword is translated in to telugu and and kannada and find the synonyms of the target language database using wordnet or dictionary.   The sysnonyms extracted from the dictionary are useful to finding the table existence in the database of any college using the metadata available in the database. Once the table existed in the database , the information is extracted using the semantically equivalent query and presented to the end user shown in the fig3 .

The f ig 4 the input key word is 'ECE'

Fig 5 the result of the input keyword 'ECE' and slected from these   semantically equivalent databses and preseted the results to the end user.
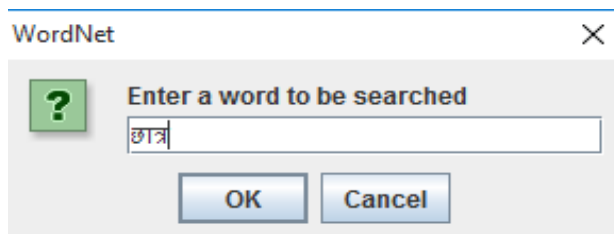

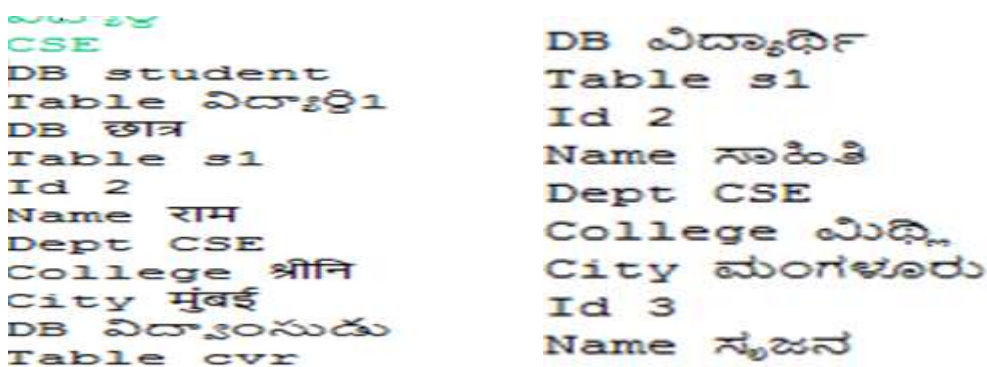Fig 2: Input string Student equivalent word in Hindi as an input

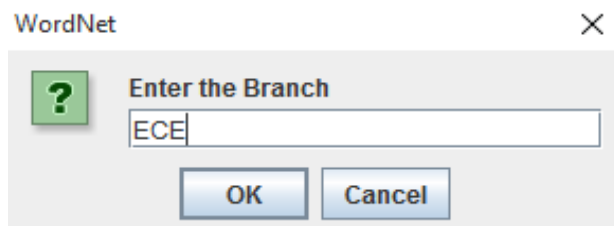Fig3: Output for the word in different Indian languages

Fig input key word "ECE'

```
student
ECE
DB విద్యార్థి
Table s1
DB విద్యాంసుడు
Table cvr
Id 2
Name కపిల్
Dept ECE
College నారాయణ
City వరంగల్
```

```
Id 3
Name
శ్రీను
Dept ECE
College నారాయణ
City ఖమ్మం
DB ಛಾತ್ರ
Table s1
DB విద్యార్థీ
```

Fig-5 Output for the key word 'ECE'

V. CONCLUSION

This paper presented an architectural framework for extraction of information from Indian language databases. The structural stability of the SQL is exploited by this framework to translate the native language query into the target language SQL query through an exhaustive search mechanism. This paper considered Hindi as the native language and Kannada and Telugu as target languages while translating one SQL statement into another. The results of the generated query are presented at a centralized location.

REFERENCES

[1]. ChVS Satyamurty, Dr JVR Murthy,M Rragahva. Meta databased semantic query in Multilingual databases. International Conference FICTA-2017,Informtaion and decision sciences, Advances in intelligent systems and Computing Springer 701, pages249-253.
[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 multi database systems, Knowledge and Data Engineering, 12(6):959-.978, 2000.
[3]. Kumar, P., Mohan, Vaideeswaran, J. Semantic based Efficient Cache Mechanism for Database Query Optimization, International Journal of Computer Applications, Volume 43. As No.23, page(s): 14-18, April 2012.
[4]. C. Fellbaum. Wordnet: An electronic lexical database. 1998.
[5]. A. Simitsis, G. Koutrika, and Y. E. Ioannidis. Précis: from unstructured keywords as queries to structured databases as answers. VLDB J., 17(1):117–149, 2008.
[6]. M. Wattenberg and F. B. Viégas. The word tree, an interactive visual concordance. IEEE TVCG'08, 14(6):1221–1228, 2008.
[7]. ChvS Satyamurty ,Dr JVRMurthy ,M.Raghava Metadata based semantic query in relational databases,India-2017 Danang,2017.
[8]. M Umadevi and Meera Gandhi ,WordNet and Ontology based query expansion for semantic information retrieval in Sports domain, Journal of Computer Science 2015, 11 (2): 361.371,DOI: 10.3844/jcssp.2015.361.371.
[9]. Christina Christodoulakis, Eser Kandogan, Ignacio G TerrizzanoVIQS- visual interactive exploration of query semantics,ESIDA,ACM ,4503-4903 ,March2017.
[10]. Mariana Soller Ramada, Joao Carlos da Silva, Plınio de Sa Leitˆao-Junior, Data Extraction from Structured Databases using Keyword-based Queries,SBBD Proceedings,ISSN 2316-5170 Oct 2014.
[11]. Ramakrishna Kolikipogu, B. Padmaja Rani, and N. Swapna Pseudo Relevance Feedback by linking WordNet for
[12]. Expanding Queries in Information Retrieval Process, International Journal of Modeling and Optimization, Vol. 3, No. 5, October 2013, pp462-467.
[13]. P. C. Reghu Raj, P. C. Reghu Raj, Malayalam Wordnet: A Relational Database Approach, International Journal of Latest Trends in Engineering and Technology (IJLTET), Vol. 3 Issue2 November 2013, PP 75-79.