ORIENTAL JOURNAL OF
COMPUTER SCIENCE & TECHNOLOGY
An International Open Free Access, Peer Reviewed Research Journal
www.computerscijournal.org

ISSN: 0974-6471
December 2011,
Vol. 4, No. (2):
Pgs. 451-454

Semantic Based Smarter Natural Language Interface for Database

KESHAV NIRANJAN

Department of Computer Science and Information Technology,
Singhania University, Pacheri Bari, Rajasthan (India).
*Corresponding author: E-mail: keshav.niranjan@gmail.com

(Received: September 11, 2011; Accepted: October 20, 2011)

ABSTRACT

The paper presents semantics based search paradigm to be embedded in Natural Language Interface (NLI) systems. The classical Information Retrieval (IR) models were based on lexical mapping and approximation based searches which suffered from obvious weaknesses as follows -

1. The queries used predefined lexical mapping or approximations and would skip any direct or indirect references via semantic alternatives. Homonymous lexemes can give many meanings leading to ambiguous queries and failed processes or ambiguous results if user is using the hyper word query. No intelligent mechanism is present in the NLI by which it will interpret the query.
2. When we write the query, then each lexeme gives only individual meaning of the word but lexemes are related to each other and produce a collocated meaning of the entire sentence. The classical IR model does not consider this aspect of IR.

To get over with these inadequacies in the classical IR mode, the NLI has to be made smarter with adequate semantic capabilities. Therefore we will provide the inferential capability to the existing NLI by providing the knowledge base to the system. This knowledge base will consist of the facts, concepts, synonymy, homonymy, hypernymy, discourse, and the contextual information and will help in generating appropriate and accurate results.

Key words: NLI, IR, QLE, Polysemy, Homonymy, Hypernymy, Discourse.

INTRODUCTION

Classical NLI's suffering from the problem of overshooting means asking queries that the system is not programmed to understand and undershooting means not realizing that the system can answer questions about areas the system knows about. Although for making Natural Language Interfaces, smarter and semantics based effort have done by different researchers. In late 1970, PLANES (Waltz, 1975) and REL (Thompson, 1975) systems were developed [2,3]. Some of these systems used semantic grammars. This is an approach in which non-terminal symbols of the grammar reflect categories of world entities (e.g.student_name, Designation_of_employee) instead of purely syntactic categories like noun phrase, verb phrase etc.

The LADDER system was designed as a natural language interface to a database of information about US Navy ships LADDER system uses semantic grammar to parse questions to query of distributed database.

KUQA System presented in TREC-9 (2000) developed by Soo-Min Kim and his colleagues categorizes questions based on expected answer and then uses NLP techniques as well as Wordnet for finding candidate answers which suit in corresponding category. However they don't handle any linguistic phenomena. Indeed all these effort are not worthless although they are not complete, so next section of this paper is trying to give the method for overcoming the mentioned inadequacies.

Method for making NLI's smarter

Synonymy

In query formulation, if user is using the synonyms of lexicon then classical retrieval Information system is inadequate to accessing the database due to lack of database of synonyms. It can be seen from example below-

- Query = "Number of student in computer science and physics"
- Query Logical expression (QLE) will be-
- QLE = student Λ computer science Λ physics, where Λ is and operator.

So from this query user will get the output (black space) as shown in the diagram 1

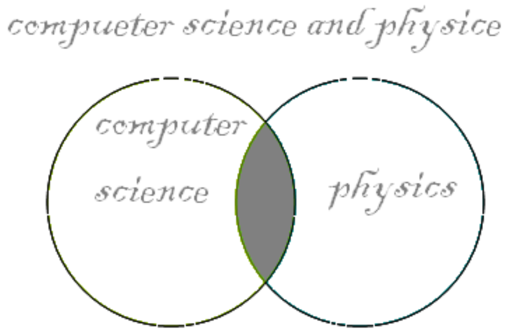

Diagram 1

But same query if user is formulating using the synonyms like as

- Query = "Show number of student in C.S. and Phy."
- Then query logical expression will be as
- QLE=Student LC.S L Phy. Where L is and operator

Then in this case Conventional NLI will be unable to give the output because keywords C.S and Phy. are not defined in the system although they are similar in meaning to Computer Science and Physics. So for making the search semantic based, Conventional NLI System will embed the database of synonyms. So if any keyword does not match, then system will not skip from the processes and system will check weather non matching keyword is synonym of any keyword by traversing the synonyms graph as shown in the diagram. C.S. is synonym of Computer Science so system will replace the keyword C.S. into Computer Science and Keywords Phy. into Physics,and thereafter will execute the query and get the desired output.

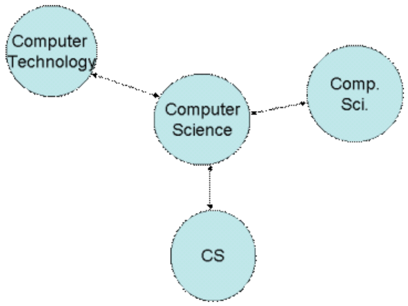

Diagram 2

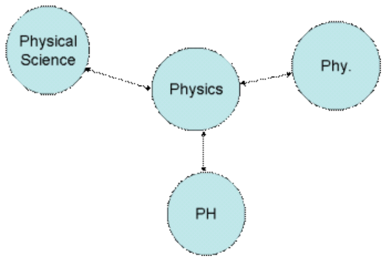

Diagram 3

Homonymy

Ambiguity arises in the sentence meaning if the user is using the homonymous keyword in the query formulation..So the possibility of producing the erroneous result is more in accessing the database because conventional NLIs are unable to distinguish actual meaning of homonyms. As we can see in the following query-

"How many swimmers are going to bank"- since here word bank which has two meanings

Bank

The margin of a watercourse; the rising ground bordering a lake, river, or sea, or forming the edge of acutting, or other hollow[4].

Bank

An establishment for the custody, loan, exchange, or issue, of money, and for facilitating the transmission of funds by drafts or bills of exchange, or a financial institution[4].

So overcoming these inadequacies of NLIs, system should embed the knowledgebase of homonyms. So from the below diagrams of bank it is clear that Bank is related to two different context and the query is related to bank that have water and further water is related to swimmer. Thereafter, system will deduct it when it will traverse the knowledge graph of homonymy.

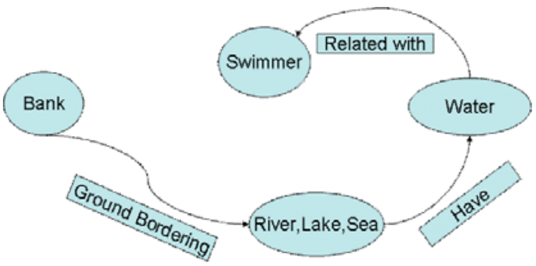

Diagram 4

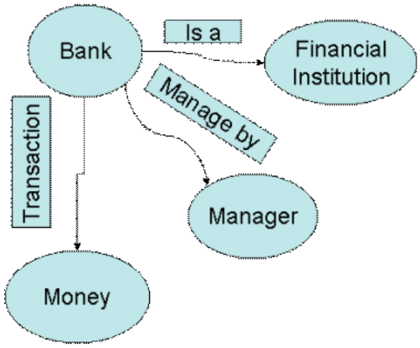

Diagram 5

Hypernym

In the query if user is using the broad category or general concept words or hypernyms, as in the query "Number of student of Science Course", the conventional NLI model will be inadequate because the system does not have database of hypernymys. 'Science' is the hypernym of Physics, Chemistry, Computer science, Mathematics etc. For making system smarter, we include the database of hypernyms.

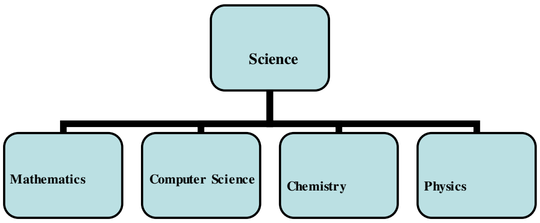

Diagram 6

Discourse

If a keyword is referring to another keyword in query as we can see in the following query Q = "Show john and his sister's marks?" Classical NLI will be unable to give the output due to lack of outer world knowledge. So if we provide the knowledge base of pronoun, anaphora and definite NP words to the system, we will produce intelligence in the system and output of this type of query will be achieve/improved.

Query = "Show john and his sister's marks?" Can be further broken in following relation

Name (john, X)
His(X)
Name (Mary, Y)
Sister(Y, X)
Marks(X) Λ Marks(Y)

So in the above query his is referring to male and john is the male, so his is referring to john and Y is the sister of X and Y is the Mary so after this deduction query will be interpreted as "his sister means Marry and give the marks of John and Mary."

CONCLUSION

Classical NLI model is based on restricted domain so only static format query will be executed by Classical NLIs. So due to this drawback, more cognitive burden will be put on the user for operating the system. So for making NLIs semantics based and user friendly, we make it smarter therefore we embedded the additional capability of knowledgebase in it. So these knowledge base and databases will overcome the lack of understanding of synonymy, hypernymy, homonymy, and discourse of conventional NLIs. After that its behavior will be intelligent and semantics based. Thereafter it will be able to execute all the open domain based queries.

ACKNOWLEDGEMENTS

Author thanks Dr.Girish Nath Jha, Jawaharlal Nehru University for reviewing and giving valuable comments that helped me revise the draft of this paper.

REFERENCES

1. Jurafsky, D., and Martin, J., "*Speech and Natural Language Processing*", Pearson Education Publication, 81-7808-594-1 (2002).
2. Mitkov, Ruslon, *The Oxford Handbook of Computational Linguistic*s, Oxford University Press (2003).
3. Jha, Girish Nath , "*A Natural Language Interface for Databases"* , Dept. of Linguistics, University of Illinois, Urbana-Champaign
4. http://ardictionary.com/Bank/742
5. Siddiqui, Tanveer and Tiwary, U.S., *Natural Language processing and Information Retrieval,* Oxford University press (2008).
6. Niranjan, Keshav, *Natural Language Interfaces for database:Using keyword approach*, Excel India Publishers, New Delhi
7. Handrix, Gary G., Chairperson, *Natural Language Interface,* SRI International, CA 94025
8. http://21cif.com/tutorials/micro/mmpdf/hypernyms.pdf