ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 1, January 2014

Building an Efficient Database Driven Reverse Mapping Dictionary

Mrs. A. Mangalakshmi[1], P. Louise Pragasia Marie[2], M. Jayanthi[3]

Assistant Professor M.E, Department of Computer Science and Engineering, Christ College of Engineering and Technology, Puducherry, India[1]

UG Student, Department of Computer Science and Engineering, Christ College of Engineering and Technology, Puducherry, India[2]

UG Student, Department of Computer Science and Engineering, Christ College of Engineering and Technology, Puducherry, India[3]

Abstract: With the enormous availability of words in usage it is always being a challenge to find the meaning. Even the versatile speaker may thrash about finding a meaning for certain unheard words. In such cases they need some source for reference likeLexicon. In traditional model for using dictionary, forward concept is implemented where it result in set of definition and it may produce a comprehensive phases. This may even confuse the user with the different concept of understanding or sometimes user could not understand the detailed concept. To overcome this concept, we facilitate reverse dictionary in which for any phases or word, the appropriate single word meaning is given. This system also facilitates to provide the relevant meaning even if that word is not available in the database. It will also produce instant output for the user input.

Index -Term: Phrase, Lexicon, Database.

I. INTRODUCTION

It is important that the detailed study and concept is always understand. Datamining is a concept in which the detailed and in depth information is analyzed and extract the information from the data set. Data mining is a run-throughof spontaneouslysearching large stores of data to discover pattern and trends that go beyond simple analysis. Data mining uses sophisticated mathematical algorithm to segment the data evaluate the probability of future events. Data mining is alsocalled as knowledge discovery in data (KDD). The data is extracted from database. Data ware house support this concept by implementing multiple database. It is a central respotiary of data which is created by integrating data from one or more disparate sources. Data warehouse stores up-to-date data as well as ancient data and are used for generating a new trending reports for senior management report.

II. EXISTING SYSTEM

In the fact that it is more significant to make a reference for unheard word, user prefers a source like dictionary for better understanding. The performance allows online interaction with users Current semantic correspondence measurement schemes that are highly computationally rigorous. In this technique, concepts are represented as vectors in a keyword space. The two most common methods to accomplish this, latent semantic indexing (LSI) and principal component analysis (PCA), both analyze the keywords of documents in anaggregate to identify the dominant concepts in the document. Consequently these dominant concepts are represented as vectors in the keyword space and are used as the basis of similarity comparison for classification. In most implementations of Concept Similarity Problem (CSP) solutions, vectorization is done a priori, and at runtime, only vector distances are computed.

Drawbacks

- It requires the user's input phrase to contain words that exactly match a dictionary definition;
- It does not scale well—for a dictionary containing more than 100,000 defined words, where each word may have multiple definitions; it would require potentially hundreds of thousands of queries to return a result.

III. PROPOSED SYSTEM

Report the formation of the WordStar Reverse Dictionary (WRD), a scalable-driven RD system that attempts to address the core issues identified above. The WRD not only fulfills new functional ideassketched above, it does so at an order of magnitude performance and scale improvement over the greatest concept similarity measurement structures

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 1, January 2014

available without impacting key quality. We also reveal that the WRD is distantenhanced in solution quality than the two salable RDs available. Our reverse dictionary system is based on the concept that a phrase that conceptually defines a word should bear a resemblance to the word's genuine definition, if not matching the exact disputes, then at least conceptually similar. Consider, for example, the following concept phrase: "the tower is made of steel girders criss-crossed to make it stronger". Based on such a phrase, a reverse dictionary should return words such as "mercerize, iron work, and shove". In our RD, a user might input aphraserefer toan unknown term ofconcern. Since an input phrase might hypothetically satisfy the definition of manifoldwords, a RD should return a set of potential matches from which a user may select his/her optimal of terms. This isdense, however, because the user is suspectto enter a definition that exactly accordsone found in a lexicon.
The significance of the phrase the user entered should be conceptually similar enough to a genuine dictionary definition to generate a set of possible matches.

Advantages

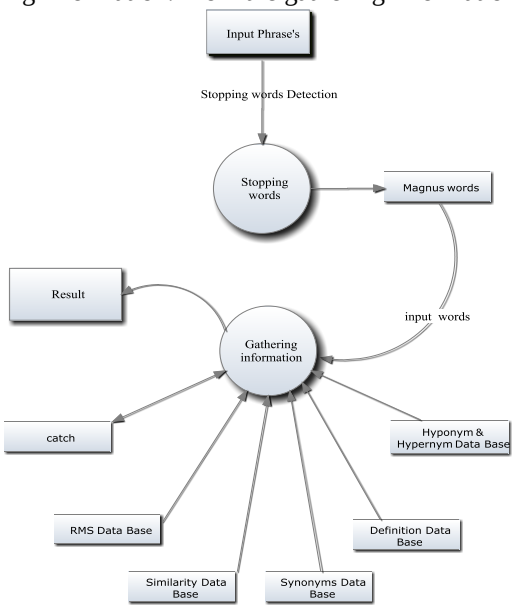
- It does so at an order of magnitude performance
- Scale improvement over the best concept similarity measurement schemes available without impacting solution quality

IV.FUTURE ENHANCEMENT

K-means clustering additionally Lamstar Network:
The course of action follows a simple way to organize a given dataset through a definite number of clusters (assume k clusters) fixed a priority. The main thought is to define k centroids, each cluster. Centroids should have a ability to achieve a result from different location. So, centroids are placed far away from one after another. The next step is to take each point belonging to a given data set and connect it to the nearest centroid. When no point is right through, the first step is fulfilled and an early groupage is done. The centroid's position is recalculated consistently until cluster have to add all components this want to be continues till all the components are cluster into the final necessary number of clusters. After this we have k new centroids, a new fastening has to be done among the same data set and the closest centroids. A loop has been produced. Result of this loop want to be notice that the k centroids change their position step by step until no more changes are done.

V.DATAFLOW DIAGRAM

The input is given by the user in the form of phrase in search box. Stopping words are those that are pre-defined by the developers. The input phrase id entered into stopping phrase in this conjunction, preposition, negative word are removed. The remaining phrase is moved to the mangus word. The mangus word stores the important words and pass it to the gathering information. From the gathering information



the words are sent to the five databases simultaneously. Five database such as synonym db which give the relevant meaning for that important word, rms db creates the parse tree for that dictionary definition , hyponym db a word that is more specific than a given word / hypernym db a word that is more generic than a given input word , antonym db which gives opposite answer to given word, definitions db is describing the word briefly .the output from five database is again send to gathering information it follow some raking algorithm in which the words search first in database is displayed first.

VI.CONCLUSION

In this paper, we describe the significant challenges inherent in building a reverse dictionary, and map the problem to the well-known conceptual similarity problem. We propose a set of methods for building and querying a reverse dictionary, and describe a set of experiments that show the quality of our results, as well as the runtime performance under load. Our experimental results show that our approach can provide significant improvements in performance scale without sacrificing solution quality. Our experiments comparing the quality of our approach to that of Dictionary.com and OneLook.com reverse dictionaries show that the Wordster approach can provide significantly higher quality over either of the other currently available implementations

REFERENCES

[1] R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval. ACM Press, 2011.
[2] D.M. Blei, A.Y. Ng, and M.I. Jordan, "Latent Dirichlet Allocation," J. Machine Learning Research, vol. 3, pp. 993-1022, Mar. 2003.
[3] J. Carlberger, H. Dalianis, M. Hassel, and O. Knutsson, "Improving Precision in Information Retrieval for Swedish Using Stemming,"

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

Technical Report IPLab-194, TRITA-NA-P0116, Interaction and Presentation Laboratory, Royal Inst. of Technology and Stockholm Univ., Aug. 2001.

[4] H. Cui, R. Sun, K. Li, M.-Y.Kan, and T.-S. Chua, "Question Answering Passage Retrieval Using Dependency Relations," Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 400-407, 2005.

[5] T. Dao and T. Simpson, "Measuring Similarity between Sentences," http://opensvn.csie.org/WordNetDotNet/trunk/Projects/Thanh/Paper/WordNetDotNet_Semantic_Similarity.pdf (last accessed 16 Oct. 2009), 2009.

[6]Dictionary.com, LLC, "Reverse Dictionary,"http://dictionary.reference.com/reverse, 2009.

[7] J. Earley, "An Efficient Context-Free Parsing Algorithm," Comm. ACM, vol. 13, no. 2, pp. 94-102, 1970.

[8] Forrester Consulting, "Ecommerce Web Site Performance Today," http://www.akamai.com/2seconds, Aug. 2009.

[9] E. Gabrilovich and S. Markovitch, "Wikipedia-Based Semantic Interpretation for Natural Language Processing," J. Artificial Intelligence Research, vol. 34, no. 1, pp. 443-498, 2009.

[10] V. Hatzivassiloglou, J. Klavans, and E. Eskin, "Detecting Text Similarity over Short Passages: Exploring Linguistic Feature Combinations Via Machine Learning," Proc. Joint SIGDAT Conf. Empirical Methods in Natural Language Processing and Very Large Corpora, pp. 203-212, June 1999.