

Database Driven Reverse Mapping:-Dictionary

Vaishali Chande, Snehal Pawar, Akanksha Goel

Department of Information Technology, ISB&M School of Technology, Pune 412115, Maharashtra

vaisuchande12@gmail.com, 9175825816

Abstract— In this paper, the design and put into effect of a reverse dictionary is being described. Uncharacteristic of a traditional forward dictionary, which maps from single meaningful unit of language to their definitions, a reverse dictionary perform an action a user input a small group of words forming a unit describing the desired concept, and go back a set of candidate single meaningful unit of language that satisfy the input phrase. This work has significant application not only for the general public, particularly those who work closely with words, but also in the general field of conceptual search. This present a set of algorithms and the thing caused or produced by something else of a set of experiments make with no errors the retrieval accuracy of the methods on the runtime response time performance of its implementation. The experimental results show that the approach of the system can provide significant improvements in carry out scale without sacrificing the quality of the result. The experiments comparing the quality of its approach to that of currently available reverse dictionaries show that of it approach can provide significantly higher quality over either of the other currently available implementations. We facilitate reverse dictionary in which for any phrase or word the appropriate single word meaning is given. This system also facilitates to provide the relevant meaning even if that word not able to be used in the database. It will also produce instant output for the user inputs.

Keywords— Dictionaries, Thesauruses, Search Process, Web based Services, Clustering.

INTRODUCTION

"This dictionary lists occurring words and phrases that can be found in reverse speech. It is not complete and never will be. Updated regularly, it is constantly develop gradually a new words that are found and further research existing entries allows definition to be sharpened and refined". Data mining uses having experience and taste in matters of culture mathematical algorithm to segment the data evaluate the probability of future events. Data mining is also called as 'knowledge discovery in Data (KDD)'. Data is extracted from RD database. Data warehouse support this concept by implementing multiple database i.e. Antonym database, Hyponym database etc. It is a central repository of data which is created by integrating database from one or more disparate sources.

With the very large availability of words in usage it is always being a challenge to find the meaning. Even the able to do speaker may thrash about finding a meaning for certain unheard words. In such cases they need some source for reference like Lexicon. In traditional model for using dictionary, forward concept is implemented where 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 phrases or word, the appropriate single word meaning is given. This system also facilitates to provide the relevant meaning even if that word not available in the database. It will also produce instant output for the user input.

REMAINING CONTENTS

Data mining is the process of discovering actionable information from large sets of data. Data mining uses mathematical analysis to derive patterns and trends that exist in data.

Basic Terminology:

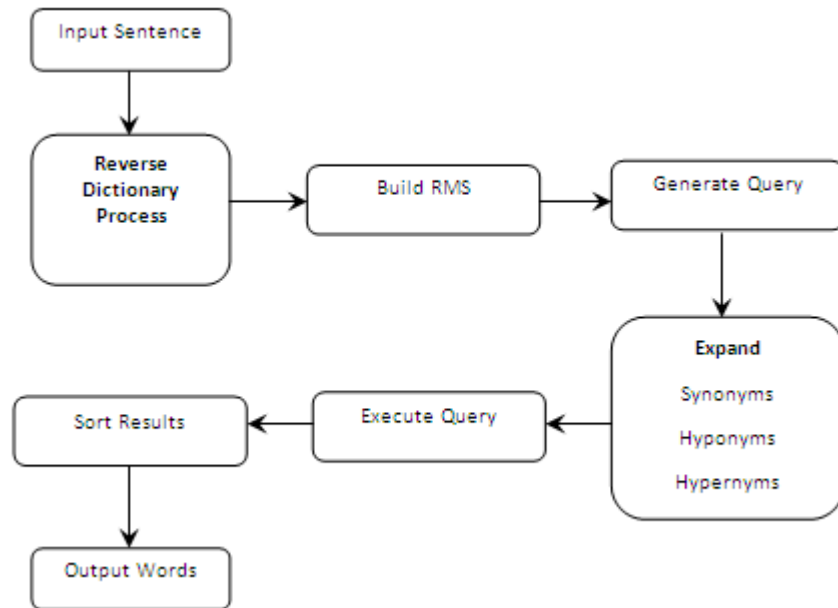
Forecasting: Estimating sales, predicting server loads or server downtime.

Risk and Probability: Choosing the best customers for targeted mailings, determining the probable break-even point for risk scenarios, assigning probabilities to diagnoses or other outcomes.

Finding Sequences: Analyzing customer selections in a shopping cart, predicting next likely events.

Grouping: Separating customers or events into cluster of related items, analyzing and predicting affinities.

Recommendations: Determining which products are likely to be sold together, generating recommendations.



System Architecture

Applications

Education: Student will get exact or relevant meaning of phrase or set of words which will help to improve knowledge.

Business: There is meeting conducted in every business and everyone write the notes sometime phrase or some set of words are use so the exact meaning is known.

Puzzling: Answer basic identification questions to solve crossword puzzle clues, or find words if you only know few letters.

Modules and algorithm

K-Means Algorithm:

K-means clustering tends to find clusters of comparable spatial extent, while the expectation maximization mechanism allows clusters to have different shapes.

Description:

Given a set of observations $(X_1, X_2, X_3 \dots X_n)$

where each observation is d-dimensional real vector

K-mean clustering partition the no. of observation into K cluster ($k < n$) sets

Where $S = S_1, S_2, S_3 \dots S_k$ so as to minimize the within cluster sum of square

Regarding computational complexity, finding the optimal solution of k-means clustering problem for observation in d-dimensions:

1. NP hard in general Euclidean d even for two Clusters.
2. NP hard for a general no. of cluster k even in plane.

Algorithm Build RMS:

RMS stands for Reverse Mapping Set. It is a mapping algorithm designed to map the word to words of similar meaning. It improves the quality of word mapped i.e. not vulnerable to the input phrase. For an input dictionary D a mapping R is created for all term appearing in the sense phrase. The RMS algorithm describe this reverse mapping pattern.

Algorithm GenerateQuery:

Here we generate a query for all the Set Type that are mean to be used for mapping and retrieval of reversed term for the given input phrase. Here Query Q is generated for all other algorithm that are returned to get the meaning for the given set of terms in the phrase. The is the building algorithm for the Set Type and Sorted queries.

Algorithm Execute Query:

For a given query Q if u have phrase that contain terms $T_1, T_2, T_3, \dots, T_k$, it performs AND/OR operations in query. If it performs OR operation then the terms of the phrase are union with reverse term and if it performs AND operation then the term of the phrase intersect with the reverse term and we returned the union or intersection of the reverse term.

Algorithm Expand Antonyms:

Given: A query Q of the form $T_1, T_2, T_3, \dots, T_k$, it creates a copy of the query and perform negation to create a sub query to replace all the terms and negated terms. If copy of the query is not equals to the copy of the original query the return copied query or else return its negated terms.

Algorithm Expand Query:

Given: A query Q of the form $T_1, T_2, T_3, \dots, T_k$, we perform AND/OR operation for all t_i in the query If AND is perform in Set Type as synonyms, antonyms and hyponyms, hyponyms to create a subquery q for the above Set Type respectively. For OR the term are replaced in query q from Q and at last it return Execute Query.

Algorithm Sort Results:

Create an empty list K and all the term are arranged in order of its retrieval priority for ease mapping. The sorted term are arrange according to it searched priority i.e. term importance, semantic and weighted similarity factor to generate a candidate set that must be ranked using mathematical computation.

Mathematical models

NP: - NP means we can solve it in polynomial time if we can break the normal rules of step-by-step computing.

NP-Hard:-NP-hard (Non-deterministic Polynomial-time hard), in computational complexity theory, is a class of problems that are, informally, "at least as hard as the hardest problems in NP".

NP-Complete:-In computational complexity theory, a decision problem is NP-complete when it is both in NP and NP-hard. The set of NP-complete problems is often denoted by NP-C or NPC. The abbreviation NP refers to "nondeterministic polynomial time".

Future Scope

The reverse dictionary works in such a way that, given an input phrase, a word related to that phrase is given as output. For the inflected forms, an algorithm is used. According to which, the inflected forms will be converted into a root form. Example: Sleeplessness during night time, in this example the verb Sleeplessness plays a very important role. But it is an inflected form of Sleep. So we have to use the Algorithm to convert the inflected form Sleeplessness to its root form sleep. For which, an algorithm called Porter-stemming algorithm is used. According to which a word which is in its inflected form is converted to its root form

CONCLUSION

Thus, the system work in developing a meaning-to-word dictionary. Depending on the phrase input there may be variation of the results shown. With our feature of finding words from an abstract string given as an input empowers the concept of reverse dictionary Parts of Speech classification feature also add quality. Improving the efficiency by holding the results makes the access fast. Adding new features like words searching enhances our work from previously available.

REFERENCES

[1] Ryan Shaw, Member, IEEE, Anindya Datta, Member, IEEE, Debra VanderMeer, Member, IEEE, and Kaushik Dutta, Member,

- IEEE, Building a Scalable Database-Driven Reverse Dictionary IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 3, MARCH 2013.
- [2] Balachandra Reddy Kandukuri, Ramakrishna Paturi V, Dr. Atanu Rakshit, "Data mining and clustering", IEEE International Conference on Services Computing, pp. 517- 520, September 2009.
- [3] T. Hofmann, Probabilistic Latent Semantic Indexing, SIGIR 99: Proc. 22nd Ann. Intl ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 50-57, 1999.
- [4] T. Korneius, J. Laurikkala, and M. Juhola, On Principal Component Analysis, Cosine and Euclidean Measures in Information Retrieval, Information Sciences, vol. 177, pp. 4893-4905, 2007.
- [5] D.M. Blei, A.Y. Ng, and M.I. Jordan, Latent Dirichlet Allocation, J. Machine Learning Research, vol. 3, pp. 993-1022, Mar. 2003
- [6] R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval. ACM Press, 2011.
- [7] T. Korneius, J. Laurikkala, and M. Juhola, "On Principal Component Analysis, Cosine and Euclidean Measures in Information Retrieval," Information Sciences, vol. 177, pp. 4893-4905, 2007.
- [8] J. Kim and K. Candan, "Cp/cv: Concept Similarity Mining without Frequency Information from Domain Describing Taxonomies," Proc. ACM Conf. Information and Knowledge Management, 2006.
- [9] R. Mihalcea, C. Corley, and C. Strapparava, "Corpus-Based and Knowledge-Based Measures of Text Semantic Similarity," Proc. Nat'l Conf. Artificial Intelligence, 2006.
- [10] D. Milne and I. Witten, "Learning to Link with Wikipedia," Proc. 17th ACM Conf. Information and Knowledge Management, pp. 509-518, 2008.
- [11] U. of Pennsylvania, "The Penn Treebank Project," <http://www.cis.upenn.edu/treebank/>, 2009.
- [12] C. Manning, P. Raghavan, and H. Schutze, Introduction to Information Retrieval. Cambridge Univ. Press, 2008.