



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 5, May 2014

Intelligent Database Driven Reverse Dictionary

P. Arivazhagan, S.Thairiyalakshmi,

P.G Scholar, Department of Computer Science & Engineering, Karpaga Vinayaga College of Engineering & Technology,
Madhuranthagam (T.K), India.

Assistant Professor, Department of Computer Science & Engineering, Karpaga Vinayaga College of Engineering &
Technology, Madhuranthagam (T.K), India.

ABSTRACT— The reverse dictionary identifies a concept/idea/definition to words and phrases used to describe that concept. You can enter a single word, phrase, or a few words to get the correct meanings for that sentence. The proposed concepts similarity is applied using similarity for computing the relevancy between concepts verses and the documents or verses in English collections.

Reverse Dictionary (RD) is a non alphabetic order dictionary. We present a set of algorithms and the results of a set of experiments showing the retrieval accuracy of our methods and the runtime response time performance of our implementation. Our experimental results show that our approach can provide significant improvements in performance scale without sacrificing the Quality of the result. Our experiments comparing the quality of our approach to that of currently available reverse dictionaries show that of our approach can provide significantly higher quality over either of the other currently available implementations.

INDEX TERMS— RD,RMS,DICTIONARIES,FILTERING

I. INTRODUCTION

A reverse dictionary is a dictionary organized in a non-standard order that provides the user with information that would be difficult to obtain from a traditionally alphabetized dictionary. For example, A Reverse Dictionary of the Spanish Language and Walker's Rhyming Dictionary are reverse dictionaries, the organization of which is based upon sorting each entry word based upon its last letter and the subsequent letters proceeding toward the beginning of that word. Consequently, in these reverse dictionaries all words that have the same suffix appear in order in the dictionary.

Such a reverse dictionary would be useful for linguists and poets who might be looking for words ending with a particular suffix, or by an anthropologist or forensics specialist examining a damaged text (e.g. a stone inscription, or a burned document) that had only the final portion of a particular word preserved. Reverse dictionaries of this type have been published for most major alphabetical languages (see numerous examples listed below). By way of contrast, in a standard dictionary words are organized such that words with the same prefix appear in order, since the sorting order is starting with the first letter of the entry word and subsequent letters proceeding toward the end of that word.

In section II we discussed the evolution of work in brief. In section III discusses about the overview of Reverse Dictionary. In section IV Problem description is discussed in brief. In section V the proposed methodologies are described. In section VI application of the project in various areas are discussed.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 5, May 2014

II. RELATED WORKS

T.Dao[5], described An approach for capturing similarity between Words that was concerned with the syntactic similarity of two strings. Today they are Back to discuss another approach that is more concerned with the meaning of words.

F. Sebastiani[7], described Research Community the dominant approach to this problem is based on machine learning Techniques: a general inductive process automatically builds a classifier by learning, From a set of reclassified documents, the characteristics of the categories.

W. Croft[6], described Information retrieval researchers have studied passage retrieval extensively, yet there is no consensus within the community about how to evaluate the results of passage retrieval Experiments. This paper describes five character-level passage Evaluation measures and tasks for which they may be Appropriate. In the second half of the paper they compare several passage retrieval models, including a new generative Mixture model that outperforms strong baselines on many of the evaluation measures discussed in part one.

PRELIMNARIES

In this section we describe about what is Reverse Dictionary and How it works .

A. *Reverse Dictionary in Brief*

A reverse dictionary is a dictionary organized in a non-standard order that provides the user with information that would be difficult to obtain from a traditionally alphabetized dictionary. For example, A Reverse Dictionary of the Spanish Language and Walker's Rhyming Dictionary are reverse dictionaries, the organization of which is based upon sorting each entry word based upon its last letter and the subsequent letters proceeding toward the beginning of that word. Consequently, in these reverse dictionaries all words that have the same suffix appear in order in the dictionary. Such a reverse dictionary would be useful for linguists and poets who might be looking for words ending with a particular suffix, or by an anthropologist or forensics specialist examining a damaged text (e.g. a stone inscription, or a burned document) that had only the final portion of a particular word preserved. Reverse dictionaries of this type have been published for most major alphabetical languages (see numerous examples listed below). By way of contrast, in a standard dictionary words are organized such that words with the same prefix appear in order, since the sorting order is starting with the first letter of the entry word and subsequent letters proceeding toward the end of that word.

Reverse dictionaries of this type were historically difficult to produce before the advent of the electronic computer and have become more common since the first computer sorted one appeared in 1974. Another use of the term "reverse dictionary" is for a reference work that is organized by concepts, phrases, or the definitions of words. This is in contrast to a standard dictionary, in which words are indexed by the headwords, but similar in function to a thesaurus, where one can look up a concept by some common, general word, and then find a list of near-synonyms of that word. (For example, in a thesaurus one could look up "doctor" and be presented with such words as healer, physician, surgeon, M.D., medical man, medicine man, academician, professor, scholar, sage, master, expert.) In theory, a reverse dictionary might go further than this, allowing you to find a word by its definition only (for example, to find the word "doctor" knowing only that he is a "person who cures disease"). Such dictionaries have become more practical with the advent of computerized information-storage and retrieval systems.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 5, May 2014

The report work on creating an online reverse dictionary (RD). As opposed to a regular (forward) dictionary that maps words to their definitions, a RD performs the converse mapping, i.e., given a phrase describing the desired concept, it provides words whose definitions match the entered definition phrase. For example, suppose a forward dictionary informs the user that the meaning of the word “spelunking” is “exploring caves.” A reverse dictionary, on the other hand, offers the user an opportunity to enter the phrase “check out natural caves” as input, and expects to receive the word “spelunking” (and possibly other words with similar meanings) as output.

III. PROBLEM DESCRIPTION

The two most common methods to achieve latent semantic indexing (LSI) and principal component analysis (PCA), both analyze the keywords of documents in a corpus to identify the dominant concepts in the document. In most implementations of CSP (concept similarity problem) solutions, vectorization is done a priori, and at runtime, only vector distances are computed. Concepts are represented as vectors in a feature (or keyword) space.

Our experimental results show that our approach can provide significant improvements in performance scale without sacrificing the quality of the result. Our experiments comparing the quality of our approach to that of currently available reverse dictionaries show that of our approach can provide significantly higher quality over either of the other currently available implementations.

The user input is unlikely to exactly match the definition of a word in the forward dictionary. The response efficiency needs to be similar to that of forward dictionary online lookups. Vectorization is only the concept for keyword search which is not suitable in online speed.

IV. PROPOSED METHOD

Similarity is an important and fundamental concept many other fields. In existing for similarity measures are heuristic in nature and tied to a particular domain or form of knowledge representation. In this project, it describes a reverse mapping technique of similarity in terms of information theory. The universality of the definition is demonstrated by its applications in different domains where different similarity measures have been employed. Then we provide the synonyms, antonyms, hyponyms and hypernyms for each input phrase which are provided by users.

The mappings for the RMS, synonyms, hyponyms, hypernyms, and antonyms are stored as integer mappings, where each word in the Wordnet dictionary is represented by a unique integer. This both condenses the size of the mapping sets, and allows for very fast processing of similarity comparisons, as compared to string processing. This work can improve the efficiency of the simple method proposed above by reducing the set of definitions to compare to the user input phrase by knowing which definitions a given word appears in.

We create a scalable driven database reverse dictionary. We propose a set of methods for building and querying a reverse dictionary. Reverse dictionary system is based on the notion that a phrase that conceptually describes a word should resemble the word’s actual definition, if not matching the exact words, then at least conceptually similar. Consider, for example, the following concept phrase: “talks a lot, but without much substance.” Based on such a phrase, a reverse dictionary should return words such as “gabby,” “chatty,” and “garrulous.” This approach consists of two sequential steps. Upon receipt of a user input phrase, we first find candidate words from a forward dictionary data source, where the

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 5, May 2014

definitions of these candidate words have some similarity to the user input. We then rank the candidate words in order of quality of match.

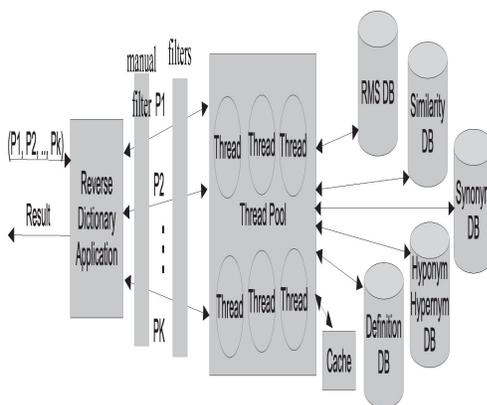


Fig 1 : Architecture Diagram

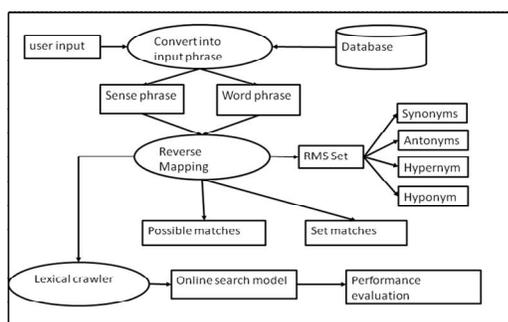


Fig 2: Block Diagram

V. APPLICATIONS

Effectively, the RD addresses the “word is on the tip of my tongue, but I can’t quite remember it” problem. A particular category of people afflicted heavily by this problem are writers, including students, professional writers, scientists, marketing and advertisement professionals, teachers, the list goes on. In fact, for most people with a certain level of education, the problem is often not lacking knowledge of the meaning of a word, but, rather, being unable to recall the appropriate word on demand. The RD addresses this widespread problem.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 5, May 2014

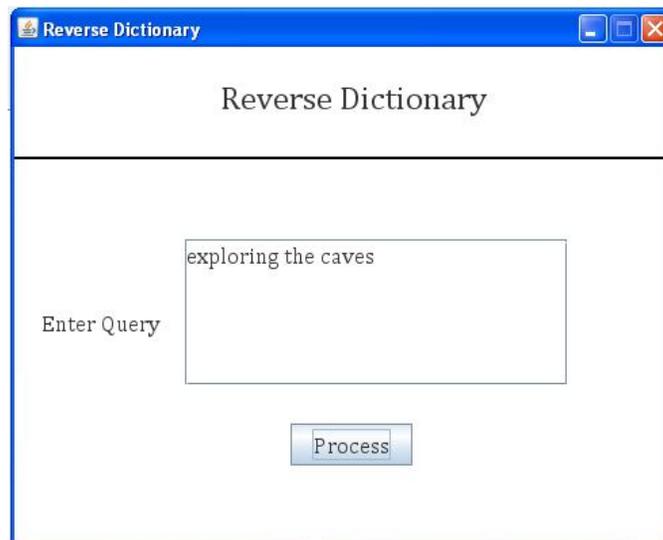
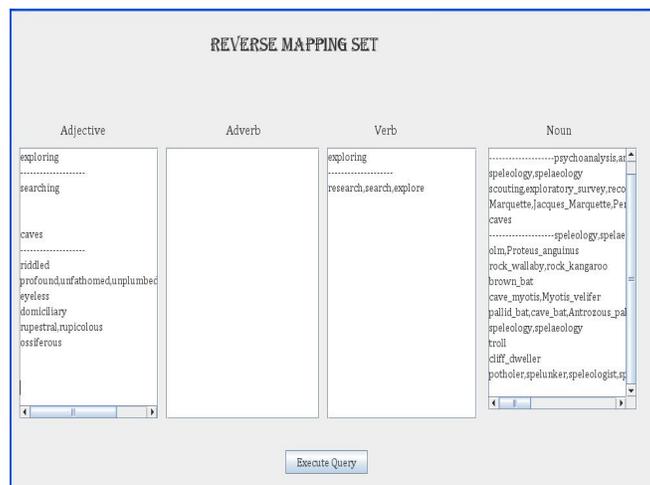


Fig 3: user input phrase



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 5, May 2014

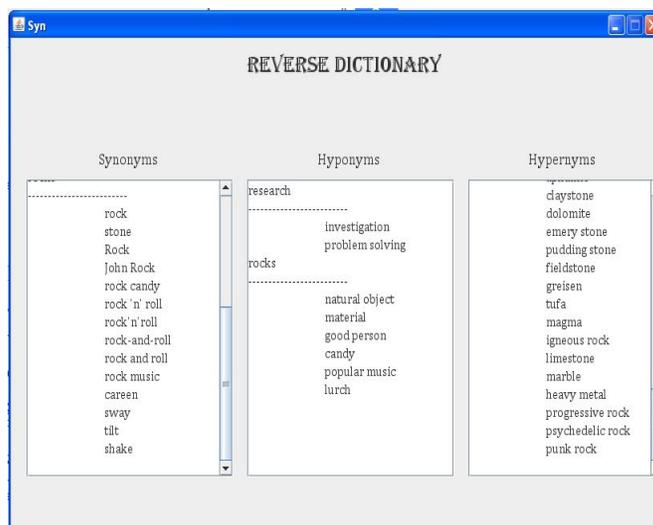


Fig 4: RMS Set



Fig 5: Candidate words

VI. CONCLUSION

The proposed approaches to create a reverse dictionary from an existing forward using Word net. It shows that a high precision reverse dictionary can be created without using any other intermediate dictionaries or languages. Using the Word net hierarchy increases the number of entries in the created dictionaries. It conclude that our reverse dictionary to



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 5, May 2014

analyze the sentences and identify the phrases. The dictionary have several number of Reverse map set includes synonyms, antonyms, hypernyms and hyponyms. Then describe a set of experiments that show the quality of our results, as well as the runtime. It can overcome the runtime load based lexical crawler to extend this works.

Crawlers are basic entity that makes reverse dictionary to work efficiently in World Wide Web. Semantic concept is implied into the search engine to provide precise and constricted search results which is required by end users of Internet. Search engine could be enhanced in searching mechanism through semantic lexical database such as WordNet, ConceptNet, Search results would be retrieved from lexical and semantic Knowledge Base (KB) by applying word sense and metadata technique based on the user query. The Uniform Resource Locator (URL) could be added and updated by the user to semantic Knowledge Base.

REFERENCES

- [1] Anindya Datta, Ryan Shaw, Debra VanderMeer and Kaushik Dutta (2013) 'Building a Scalable Database-Driven Reverse Dictionary'-VOL. 25, NO. 3, pp.528-540
- [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," 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] X. Liu and W. Croft, "Passage Retrieval Based on Language Models," Proc. 11th Int'l Conf. Information and Knowledge Management, pp. 375-382, 2002.
- [7] F. Sebastiani, "Machine Learning in Automated Text Categorization,"(2002) ACM Computing Surveys, vol. 34, no. 1, pp. 1-47.

BIOGRAPHY

P.Arivazhagan is an P.G Scholar in the Department of Computer Science & Engineering, Karpaga Vinayaga College of Engineering and Technology, Chennai -603 308 (INDIA). (Mobile: +91-9944118838; e-mail: arivazhagan.palani@gmail.com).

S.Thairivalakshmi, is an Assistant Professor in the Department of Computer Science & Engineering, Karpaga Vinayaga College of Engineering and Technology, Chennai -603 308 (INDIA). (Mobile: +91-9751602782; e-mail: thairiya@gmail.com).