A Study of Building an Reverse Dictionary

D.Gaya

Research Scholar, Department of Computer Science and Engineering, Pondicherry University
Puducherry
ganyaramesh@gmail.com

Abstract— A traditional forward dictionary is where the user inputs a word and gets one or more definitions as output. Whereas, in a reverse dictionary the user enters the desired phrase with logic and gets a certain number of words as outputs similar to the concept of the user input phrase. The reverse dictionary addresses the wide spread problem of knowing the meaning of a word, but unable to recall the appropriate word on demand .In this paper introduce a broad overview of the reverse dictionary and what are the techniques is used in reverse dictionary, some limitation of the reverse dictionary.

Keywords—Reverse Dictionary, Pos Tagging, Concept mining, stemming

INTRODUCTION

A reverse dictionary can be viewed as an organized collection of concepts, word meanings or phrases and this is in contrast to a forward dictionary. So its function is similar to that of a thesaurus where one can look up a phrase by a general word or find a similar word or synonyms to the input word. The user enters the desired phrase with logic and gets a certain number of words as outputs similar to the concept of the user input phrase. The user input may be matching the definition that is already present or not.

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 dictionary of this type have been published for most major alphabetical languages. By way of contras, in a standard dictionary words are organized such that word and subsequent letters proceeding toward the end of that word.

A reverse dictionary performs a reverse mapping i.e., given a phrase describing a desired concept. It provides words whose definitions match the entered definition phrase, as opposed to a regular dictionary that maps words to their definitions. For example, a forward dictionary informs the user that the meaning of the word "regret" is "to feel sorry." Whereas reverse dictionary, offers the user an opportunity to enter the phrase "feeling of loss or longing for someone" as input, and can be expected to receive the word "regret" and possibly other words with similar meaning as output.

Most of the techniques for the creation of reverse dictionary is based on the creation of multiple databases for synonyms, hyponym, antonym etc.

This paper is organized as follows. In Section II, we present the system study of the Reverse Dictionary (RD). In section III, the previous works related to this research are reported. In section IV, we present the problem description. Conclusion is presented in the final section V.

SYSTEM STUDY

In this section we describe about what is reverse dictionary and structure of the reverse dictionary and how it is works.

A. History of Reverse Dictionary

A **reverse dictionary** is a dictionary organized in a non-standard order (usually referring to being in a so called "reverse" order) that provides the user with information that would be difficult to obtain from a traditionally alphabetized dictionary. Reverse dictionaries were historically difficult to produce before the advent of the electronic computer and have become more common since 1974 when Stahl and Scavnicky's Reverse Spanish Dictionary appeared (the first computer sorted reverse dictionary ever published).

The reverse word dictionary refers to a dictionary where the word entries in the dictionary are not alphabetized in the same manner as traditional 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 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. Some sort utilities can sort on a reversed version of the sort key, making it easy to generate reverse dictionaries on the fly if the dictionary data is available in electronic form.

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 (i.e. computer databases).

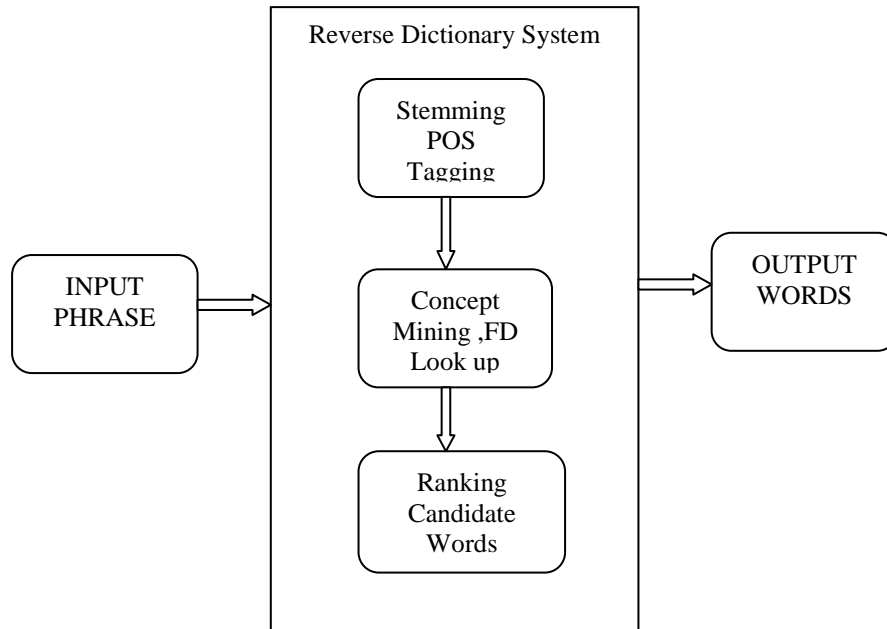*B. Architecture for a Reverse Dictionary*

Fig., 1: Architecture for a Reverse Dictionary

The general steps involved in most of the reverse dictionary systems available today. The following steps as follows

a) Stemming

        Steaming is the process of finding the root word of a word contained in a given phrase. Normally stemming is accomplished by means of porter stemmer.

b) POS Tagging

        Part-of –Speech Tagging is the process of each word in a sentence is partitioned into the each word (and other token), such as noun, verb, adjective, etc. pos tagger is piece of software which performs pos tagging. English taggers us the Penn Tree tag set.

c) Concept Mining

         Concept mining is the process of extracting meaning from a sentence or a set of words. Normally the conversion of words to concepts has been performed using a thesaurus. A number of methods are available now days to extract meaning from a given phrase or sentence.

d) Forward Dictionary Look Up

        The meaning extracted after the concept mining phase can be used and consults a forward dictionary to selects those whose definitions are similar to this concept.

e) Ranking Candidate Words

        After the forward dictionary look up phase a set of words whose meaning similar to the input phrase can be identified and this words has to be ordered by means of some probabilistic methods.

PREVIOUS WORKS

In a reverse dictionary a user phrase is given and we receive certain number of words as output ranked with an algorithm. The related works to this reverse dictionary are the reverse dictionaries [1] [2].

Reference[3] Deals with the design and implementation of a reverse dictionary It described finding candidate words from a forward dictionary data source for user phrase. Candidate words are ranked by quality of match. It consists of two key sub steps build the RMS and Query the RMS. Some Limitation of this paper as follows

- The response efficiency is not to be similar to forward dictionary online lookups.
- There is no enough words or less than the target number of words, we can extend the search using Synonyms, hypernyms, hyponyms.
- Database is not updated dynamically.
- The lack of contextual information.

Reference [4] described Input phrase is analysed using semantic similarity measures and SVM assisted by Word Net. Applying algebraic analysis to select candidate words and rank them. The similarity and relatedness measures are Jiang and Conrath (JCN) and Lin. Semantic similarity measure between each member of the input is ranked and highest values words are chosen. Some Limitation of this paper as follows

- Context Based target words are not achieved properly.
- Only nouns as word members of the semantic space.
- There is no Dynamic updation of reverse dictionary database.

Kesavaram.P.H [5] described the search is enhanced by considering the synonyms, hypernyms and hyponyms of that particular word. The final step is to sort out the results based on its rank. Some Limitation of this paper as follows

- Inflected forms of the words are not considered. For example: Root word "sleep" is considered from "sleeplessness".
- Spell checker is not provided.

PROBLEM DESCRIPTION

In Reverse Dictionary, similarity is calculated based on sentence phrase which is effective as compared with similarity based on each word. Context Based target words are not achieved properly. And only nouns as word members of the semantic space, There is no Dynamic updation of reverse dictionary database. The drawback of [4] is overcome from this proposed system. This paper is in proceeding

- Each word in a sentence is partitioned into a list of tokens words using Part-Of-Speech Tree tagger.
- Here Porter stemming algorithm is used. Porter stemming is a process of removing the common morphological and inflexional endings of words.
- The similarity of the sentences based on the similarity of the pairs of words is computed.
- Most appropriate sense for every word in a sentence can be found using Word Sense Disambiguation.
- List of possible target words are given as suggestion in addition to the required word.

V   CONCLUSION

In this paper, we discuss the steps and different methods involved in the creation and implementation of the reverse dictionary has been analyzed. It is found that the various techniques that are used in the steps involved in the creation of reverse dictionary can be replaced with more accurate methods. The main challenges that are identified in the creation of reverse dictionary are, first a user input is unlikely to exactly match the definition of a word in the forward dictionary and second the response efficiency needs to be similar to that of forward dictionary look up. The most effective methods for overcoming these challenges.

REFERENCES

[1]Dictionary.com,LLC,"Reverse Dictionary, " http://dictionary.reference.com/reverse/,2009
[2] OneLook.com, "Onelook.com Reverse Dictionary," http://onelook.com/,2009
 [3] Ryan Shaw., Anindya Datta., Debra VanderMeer and Kaushik Dutta "**Building a Scalable Database-Driven Reverse Dictionary**", *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, VOL. 25, NO. 3, MARCH 2013
[4] Oscar Méndez, Hiram Calvo, and Marco A. Moreno-Armendáriz "**A Reverse Dictionary Based on Semantic   Analysis Using WordNet**" *MICAI 2013*, Part I, LNAI 8265, pp. 275–285, Springer-Verlag Berlin Heidelberg 2013.
[5] Kesavaram.P.H, T.Muthamilselvan,"**Building a ranking based reverse dictionary"** *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 2, Issue 4, April 2014.