ISSN:0975-9646

A Survey of Building a Reverse Dictionary

Jincy A K, Sindhu L

Department of Computer Science& Engg,College Of Engineering,
Cochin university of science and Technology,
Poonjar,Kottayam,kerala,India

***Abstract*- A reverse dictionary takes as input a phrase or a sentence describing a concept, and returns a set of candidate words that satisfy the meaning of the input phrase. This paper presents a literature survey regarding the design and implementation of a reverse dictionary. The reverse dictionary identifies a concept/idea/definition to words and phrases used to describe that concept. It has significant application for those who work closely with words and also in the general field of conceptual search. Mainly for linguists, poets, anthropologists and forensics specialist examining a damaged text that had only the final portion of a particular word preserved.**

***Keywords*- Thesaurus, Stemming, Pos Tagging, Concept Mining**

I. INTRODUCTION

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 (forward) 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 meanings as output.

Most of the techniques for the creation of reverse dictionary is based on the creation of multiple databases for synonyms, hyponym, antonyms etc . The basic steps including the processing of a document includes lexical analysis of the text, elimination of stop words, stemming, index term selection etc. In addition to these the implementation of reverse dictionary includes other steps like concept mining and a forward dictionary lookup. The basic steps included in the creation of a reverse dictionary are same for all languages, whereas the change relies on the technique used to extract meaning from the given phrase.

A reverse dictionary can be viewed as a 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.

II. GENERAL TECHNIQUE

The different steps for the implementation of reverse dictionary includes lexical analysis, elimination of stop words, stemming, pos tagging, concept mapping, building the reverse mapping set, querying the reverse mapping set, ranking candidate words, term similarity...

A. Stemming

Stemming 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 but studies proved that it has certain drawbacks

B. POS Tagging

Part-Of-Speech Tagging is the process of assigning parts of speech to each word (and other token), such as noun, verb, adjective, etc. pos tagger is a piece of software which performs pos tagging. English taggers use the Penn Treebank tag set.

C. Concept Mining

Concept mining refers to 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 having high accuracy are available now a days to extract meaning from a given phrase or sentence.

D. Forward Dictionary Look Up

The meaning extracted after the concept mining phase can used and consults a forward dictionary to
selects those words 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

This is the general steps involved in most of the reverse dictionary systems available today. The differences is on the methods which is used for the different processes in these steps

Most of the existing techniques follows an approach called wordster approach which relies on the creation of multiple databases for synonyms, antonyms, hyponyms and hypernyms for analyzing each input phrase which are provided by users. The following architecture shows a general overview of the steps included in the creation and implementation of reverse dictionary.

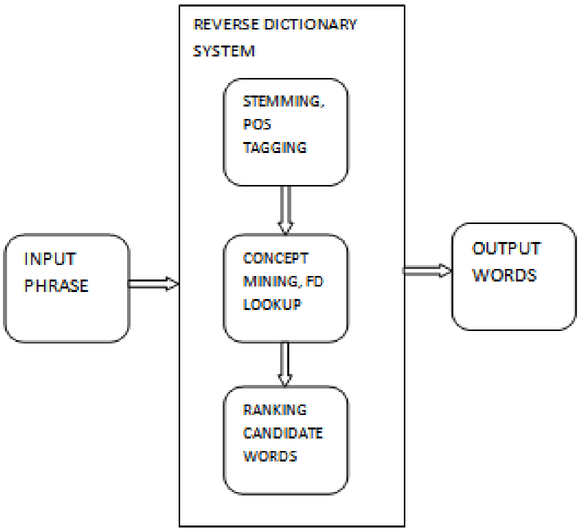

Fig 1: General architecture for a Reverse Dictionary

III. Related Works

Reference[1] Deals with the design and implementation of a reverse dictionary It describes the significant challenges inherent in building a reverse dictionary, and map the problem to the well-known conceptual similarity problem. A set of methods for building and querying a reverse dictionary has been proposed.

An approach called wordster approach has been proposed in this paper for the implementation of reverse dictionary but it includes the creation and maintenance of multiple databases which is considered to be space and time consuming

Reference [2] Covers several areas, but relevant for the creation of reverse dictionary includes text processing. The steps for document processing include lexical analysis, elimination of stop words, stemming, index item selection etc. . . . In the concept of reverse dictionary all these are to be carried out and this also describes various methods for performing these steps.

Reference [3] Proposes an improved version of the original Porter stemming algorithm for the English language. The proposed stemmer is evaluated using the error counting method. The New Porter stemmer will be used in order to improve text summarization, clustering of documents, information extraction, indexing documents , Question-Answering systems, etc. The new algorithm can be useful for words 'normalization as well as reducing the space representation.

Reference [4] reveals the inaccuracies encountered during the stemming process and proposes the corresponding solutions. There exist a number of errors associated with porter stemming and this paper proposes certain methods for correcting those errors.

Reference [5] proposes a method called Explicit Semantic Analysis(ESA), for semantic interpretation of unrestricted natural language texts. This method uses knowledge concepts explicitly defined and manipulated by humans. This approach is applicable to many NLP tasks whose input is a document. The disadvantage of these method is that it expresses meaning of texts only in terms of Wikipedia based concepts.

Reference [6] Proposes algorithms for creation of new reverse bilingual dictionaries from existing bilingual dictionaries in which English is one of the two languages. The algorithms exploit the similarity between word-concept pairs using the English Word Net to produce reverse dictionary entries. Since the algorithms rely on available bilingual dictionaries, they are applicable to any dictionary of more than one language as long as one of the two languages has Word Net type lexical ontology

Reference [7] Proposes some dictionary-based algorithms to capture the semantic similarity between two sentences, which is based on the WordNet semantic dictionary. The Steps for computing semantic similarity between two sentences includes tokenization, pos tagging, stemming words, finding the most appropriate sense for every word in a sentence and Finally, compute the similarity of the sentences based on the similarity of the pairs of words

IV. Conclusion

In this paper 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 wordster approach is the widely used method for the creation of reverse dictionary.

The main challenges that are identified in the creation of a 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 overcomojng these challenges includes latent semantic indexing (LSI) and principal component analysis (PCA).

Acknowledgement

I am using this opportunity to express my gratitude to everyone who supported me to publish this paper . I am thankful for their aspiring guidance, invaluably constructive criticism and friendly advice during the work. I am sincerely grateful to them for sharing their truthful and illuminating views for preparing this work.

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] R. Baeza-Yates and B. Ribeiro-Neto, Modern Information Retrieval. ACM Press, 2011

[3] Wahiba Ben Abdessalem Karaa, 'A New Stemmer To Improve Information Retrieval'-(IJNSA), Vol.5, No.4, July 2013

[4] Fadi Yamout, Rana Demachkieh, Ghalia Hamdan, Reem Sabra,' Further Enhancement to the Porter's Stemming Algorithm', C&E American University I., Beirut, Lebanon

[5] Evgeniy Gabrilovich ,Shaul Markovitch 'Wikipedia-based Semantic Interpretation for Natural Language Processing', Journal of Artificial Intelligence Research 34 (2009) 443-498

[6] Khang Nhut Lam , Jugal Kalita 'Creating Reverse Bilingual Dictionaries', Department of Computer Science University of Colorado, USA

[7] Thanh Ngoc Dao , Troy Simpson 'Measuring Similarity between sentences'