Scientific Journal of Impact Factor (SJIF): 4.14
e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 3, Issue 11, November -2016

"Development of term similarity measurement using Semantic network approach"

[1] Miss.Swapnaja suhas kulkarni, [2] Prof. K.S.Kadam

[1,2] Department of Computer science Engineering, D. K.T.E. Society's Textile and Engineering Institute, Ichalkaranji.

Abstract — *Traditional internet search engines are keyword-based. Such a mechanism is effective once the user is aware of precisely the right words within the websites they are trying to find. In many web applications semantic similarity between two terms can be computed. For computing semantic similarity initially there are two approaches namely knowledge based and corpus based approaches. Existing system is able to find similarity between words but, this approach can compute similarity between multi-word expressions. A semantic network is an efficient and effective approach for measuring semantic similarity between two terms. A new clustering approach introduced to improve the accuracy of computing similarity of two terms. This approach is more effective because, it computes semantic similarity between multi-word expressions with ambiguity.*

I. INTRODUCTION

Keyword based mostly search works well if the users understand specifically what they need and formulate queries with the "correct" keywords. It is typically hopeless if the users may have rough ideas concerning what they are asking.

In lexical semantics [1] basic problem is to measure semantic similarity between terms. But existing systems can only find semantic similarity between words instead of multi-word expression. The followings square measure some samples of such "conceptual queries":

Q1. Info conferences in asian cities

Q2. School firms saying

Although the intentions of those queries square measure quite clear, they're not "good" keyword queries by ancient common place. Within the 1st question, the user desires to understand concerning the database conferences settled in Asian cities, while not knowing the names of the conferences or cities. Within the second question, the user desires to seek out the varied slogans of firms. Most of applications find in web and document search [2], and text understanding scenarios. A single term can be considered as either single words or multi-word expressions (MWEs). If two terms are sharing same attributes or their meaning is close then they are semantically similar. For example, consider "emerging markets" and "developing countries" are two terms and they are similar because their semantic contents are close. Consider another example, "Google" and "Microsoft" are both software companies so that they are semantically similar. However, "car" and "journey" are not semantically similar but they both are related with each other because "car" is a transport means for the activity "journey". On an isA taxonomy semantic similarity can be defined as measuring distance between two terms. So, that the distance between "car" and "journey" is quite too far away in an isA taxonomy from WordNet [3].Semantic similarity is a more specific and is hard to distinguish from relatedness.

Initially there are two main approaches for computing term similarity: knowledge based and corpus based. First, Knowledge based approaches depends on handcrafted resources such as thesauri, taxonomies or encyclopedias, as the context of comparison. . Most add this work [4] depends on the linguistics isA relations in Word internet that could be a manually curate lexicon and taxonomy. Corpus primarily based approaches work by extracting the contexts of the terms from giant corpora so causation the spacing properties of words or n-grams. Corpuses are often something from web pages, net search snippets to different text repositories.

II. LITERATURE REVIEW

This system A new semantic relatedness measurement using wordnet features [5] by M. A. H. Taieb, M. B. Aouicha, and A. B. Hamadou which allow us to compute semantic similarity between words is a vital issue. It introduces a new Information Content (IC) method for computing semantic similarity. Information content also uses some taxonomical features which are extracted from ontology. This approach has two parts. In first half subgraph formed the thought subsumers exploitation the depth and also the descendents count as compartmentalization parameters. In a very second half, we tend to integrate this IC metric in a very new parameterized multistrategy approach for measurement word linguistics connection. This approach has been compared with other information content using a wide set of benchmarks. Results show that our IC technique and also the new connection live related to higher with human judgments than connected works.

This system Concept-based web search [6] by Wang, H. Li, H. Wang, and K. Q. Zhu describes a framework that compute the similarity through search engines. It also improves the net search experience through the knowledge domain. In traditional system net search engines are keyword-based. Similarity can be computed between those words which user is searching for. This type of system is efficient only when the user is aware of precisely the right words within the sites. But it has one drawback that it doesn't produces sensible results if a topic has an ambiguous meaning. This framework firstly classifies the net queries into different patterns. Here queries are nothing but the keywords that are searched by the user on search engine. After making the classification it performs decoding of queries using knowledge domain and produces the results.

Development and application of a metric on semantic nets [7] by R. Rada, H. Mili, E. Bichnell, and M. Blettner presents the method to find distance between two terms. Here the distance is nothing but average minimum path length between two subset of node. The distance can be used to find the conceptual distance between set of concepts. For finding the semantic similarity over semantic nets two tasks are performed. Firstly, the conceptual distance can be found. In second step distance judgment can be calculated. That determines whether semantic net s1 is better or worse than semantic net s2. People help to perform distance judgment over the semantic nets. The result shows that s1 is better than s2, if distance on s1 more like people than s2.

Using information content to evaluate semantic similarity in a taxonomy [8] by P. Resnik, describes an is-a taxonomy to measure the semantic similarity between words. The similarity can be found based on the notion of information content (IC). The information content is same like edge counting method. The results show that measuring semantic similarity using information content produces sensible results. This method is significantly better than the traditional method of measuring semantic similarity.

Exploring knowledge bases for similarity [9] by E. Agirre, M. Cuadros, G. Rigau, and A. Soroa presents graph-based algorithms for finding similarity between words. For that it uses completely different relations and versions of WordNet

along with graph based algorithms. Here the WordSim353 dataset is used. The results show that victimisation the adequate relations the performance improves over antecedently printed WordNet-based results.

III. MOTIVATION

Existing approaches are suitable for computing semantic similarity between words instead of multi-word expressions (MWEs), they do not scale very well and are not time efficient. Therefore there is need to improve the quality of measuring semantic similarity. The effective refined approach is introduced for semantic similarity between multi-word expressins.

IV. SYSTEM ARCHITECTURE

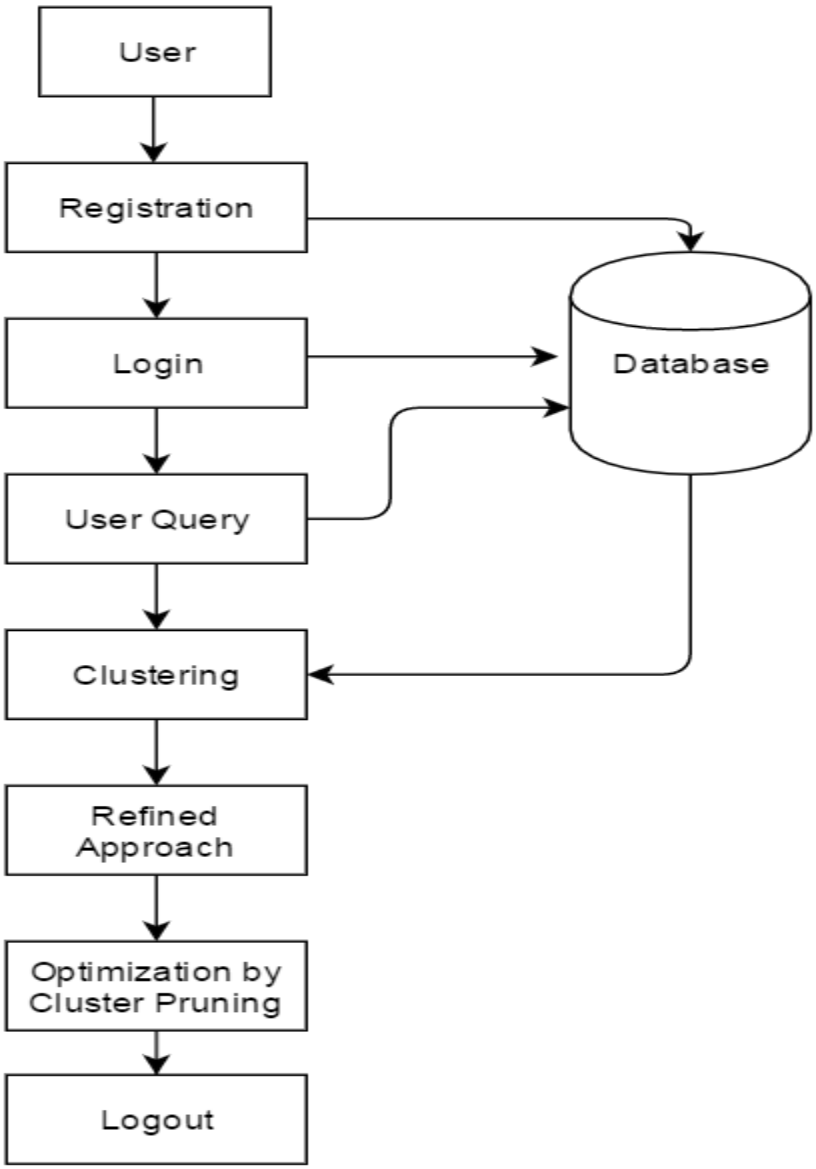

Fig1. Architecture of semantic similarity

The above diagram illustrates architecture of semantic similarity. Where a user may perform login and may give a query to the database. Here admin is responsible to maintain a dataset. After generating a dataset it may upload that into database. From a database terms are extracted and given as an input to type checking method. Then context of a term can be extracted from its type. By using a clustering algorithm according to its contexts clusters are formed. Finally by using similarity functions semantic similarity can be measured and output is generated.

REFERENCES

[1] A. Budanitsky and G. Hirst, "Evaluating wordnet-based measures of lexical semantic relatedness," Comput.Linguistics, vol. 32, pp. 13–47, 2006.

[2] M. G€artner, A. Rauber, and H. Berger, "Bridging structured andunstructured data via hybrid semantic search and interactiveontology-enhanced query formulation," Knowl. Inf. Syst.,vol. 41,pp. 761–792, 2014.

[3] G. A. Miller, "WordNet: A lexical database for english," Commun.ACM, vol. 38, no. 11, pp. 39–41, 1995.

[4] P. Resnik, "Using information content to evaluate semantic similarity in a taxonomy," in Proc. 14th Int. Joint Conf. Artif.Intell.,1995, pp. 448–453.

[5] A new semantic relatedness measurement using wordnet features, M. A. H. Taieb, M. B. Aouicha, and A. B. Hamadou,2013.

[6] Y. Wang, H. Li, H. Wang, and K. Q. Zhu, "Concept-based websearch," in Proc. 31st Int. Conf. Conceptual Model. ER, 2012,pp. 449–462.

[7] R. Rada, H. Mili, E. Bichnell, and M. Blettner, "Development and application of a metric onsemanticnets," IEEE Trans. Syst., ManCybern., vol. 9, no. 1, pp. 17–30, Jan./Feb. 1989.

[8] P. Resnik, "Using information content to evaluate semantic similarity in a taxonomy" in proc.14th Int. joint conf. Artif. Intell.,1995, pp.448-453.

[9] E. Agirre, M. Cuadros, G. Rigau, and A. Soroa, "Exploring knowledge bases for similarity," in Proc. 7th Int. Conf. Language ResourcesEval., 2010, pp. 373–377.

[10] M. A. Hearst, "Automatic acquisition of hyponyms from large text corpora," in Proc. 14th Conf. Compute. Linguistics, 1992, pp. 539–545.

[11] Y. Song, H. Wang, Z. Wang, H. Li, and W. Chen, "Short text conceptualization using a probabilistic knowledgebase," in Proc.22nd Int. Joint Conf. Artif.Intell., 2011, pp. 2330–2336.

[12] W. Wu, H. Li, H. Wang, and K. Q. Zhu, "Probase: A probabilistic taxonomy for text understanding," in Proc. ACM SIGMOD Int.Conf. Manag. Data, 2012, pp. 481–492.

[13] Ryan Shaw, Member, Anindya Datta, "Building a Scalable Database-Driven Reverse Dictionary",2013.

