Transactions of the Nigerian Association of Mathematical Physics
***Volume 4,* (*July, 2017*), pp217 – 224**
© Trans. of NAMP

Mitigating Word Semantic Similarity In Information Retrieval Using Concept Information Content And Its Hierarchy

Abdullah K-K. A[1] and Robert A.B.C [2]

[1]Department of Mathematical Sciences, Olabisi Onabanjo University, Ago Iwoye, Ogun State, Nigeria.
[2] Department of Computer Science, University of Ibadan, Ibadan. Oyo State, Nigeria.

Abstract

Word and concept similarity are used interchangeably in natural language processing and information retrieval. Word similarity in retrieval systems is based on lexical matching, which determines if query terms are useful and reflect the users' information need. Previous work on semantic similarity used one similarity measure but still pose problem since concepts are organised according to hierarchies. In the paper, the propose work use different semantic similarity network (WordNet) by amplifying structured hierarchy on Information Content (IC) to optimise the mismatch between the words. However, semantic similarity from text documents extracts lexicosyntactic patterns in the WordNet hierarchy to determine the similarity of words. Comparative analysis of semantic similarity methods are considered on the basis of performance and their limitations. From the study, experimental result on Miller-Charles benchmark dataset show that the proposed method performs better than using single existing similarity measure in WordNet.

Keywords: Semantic similarity, WordNet Similarity, Ontology, Semantic network, WordNet

1.0 Introduction

Similarity is an aspect of information retrieval of web documents and filtering. It is an important component of various tasks on the web such as information extraction, text mining, word sense disambiguation, and natural language language. The available web documents are in different forms and the information they contain is difficult to access. This is due to the growth of web information that is so enormous. The search engine plays a more critical role in finding relation among input keywords but fails in retrieving semantically related documents. As a result, it retrieves more irrelevant documents than needed. The attempt is to match the user's query to the source documents and present it to the user documents that match the user keyword. The retrieval system depends on the similarity between indexer and the queries, which is measured by comparing the values of certain attributes to indexer and user requests. But indexers and user do not always use the same terms according to [1] therefore, synonym terms fail to retrieve relevant documents with a decrease in recall. Subsequently, polysemy causes retrieval of irrelevant documents, which implies a decrease in precision retrieval.

Most approaches developed to enhance word similarity based on a single similarity measure but similarity between entities changes over time and across domains. In such approaches, there is representation of documents in a linear feature vector in which similarity or relation among features is considered with context or structure only. Therefore, traditional retrieval systems have limited abilities to exploit the conceptualisations involved in user needs and content meanings due to inability to describe the relation among search terms. This is because search engines are keyword based which have not bridged the gap of vocabulary mismatch problem in retrieval system. The word mismatch is a problem in the usage of natural language [2]. Language mismatch and ambiguity of words in documents' repository on web content causes difficulties in retrieving relevant documents in related domains [3-5].

The use of semantic indexing is based on the hypothesis that a document is viewed as a set of concept. The importance of a concept depends on the number of links with other concepts that share the same document.

Correspondence Author: AbdullahK-K. A[1]., Email: [1]uwaizabdullah9@gmail.com, Tel: +2348060046592

The concept of Semantic Web [6] manually or automatically constructs taxonomy of semantic concepts and its relations. This concept can be used to map documents in the retrieval and queries. The concept of Semantic Web depends on ontology as a searching tool for information and related resources. Therefore, ontology captures the semantic relationship between concepts or vocabulary used in a particular domain. This discovers relationships between descriptions of entities, used in information retrieval [7] query expansion, indexing and retrieval [8].

To overcome the limitations of existing web search systems and difficulty of keywords search engines, queries need to be represented with context through ontology structure for effective search [9]. Subsequently, semantic network (WordNet) involving information content with its hierarchies computes semantic similarity of terms in retrieval system. This controls the vocabulary in the web by taking into consideration the semantic meaning of two concepts in related domains using different terms in WordNet. But most existing semantic networks use a single similarity method with one weakness or the other [10, 11]. Existing semantic network similarity on Information Content (IC) does not consider the path length of taxonomy in WordNetSimilarity. However, similarity methods that consider the position of concepts in the taxonomy perform better than only path length [12].

The remaining part of this work is as follows: Section 2 describes the related work and information sources of different similarity measure of words. Section 3 describes the methodologies on how to reduce or overcomes the shortcoming of different similarity measures in related retrieval domains. The section 4 analyse the result generated from the section 3 and discussion of the result. Finally, section 5 describes conclusion and future work.

2.0 Related Work

The documents retrieved from the retrieval system are in response to a query arranged according to the relevance of the query. Although users are still often faced with the daunting task of sifting through multiple pages of results, many of which are irrelevant. Roush [13] indicated that almost 25% of web searchers are unable to find useful results in the first set of Uniform Resources Locator (URL) that are returned. This is due to the keywords based searches which have a tough time distinguishing between words that are spelled the same way but have different meanings. This often results in hits that are completely irrelevant to the query. Also, search engines cannot return hits keywords that are the same but entered different words in the query. With the conceptual knowledge, search engines based on concepts can effectively handle the problems where domain specific ontology based semantic search is used.

Most similarity techniques that have been used in information retrieval systems do not consider the semantic of terms in retrieval of web document in related domains [14, 15]. Nirgude [16] proposed Page Count and Snippets Method (PCSM) to estimate semantic similarity between any two words or entities based on page counts and text snippets retrieved from a web search engine. However, context provides extra information to improve search result’s relevance. Finding similarity in related domains cannot be based on single information source as in some of the existing work such as in [17]. Furthermore, exploiting the information content with structure of the taxonomy (hierarchy) also performs better [18]. It used path and information content that is inversely proportional to length but does not consider the position of the hierarchies of concepts. Although, semantic network source has the advantage to be fast and makes it possible to have a reusable resource even though the corpus changes. Its drawback is the possibility to omit some concepts with different forms that appeared in the source text and in the ontology.

2.1 Theoretical Framework on Semantic Network

This section explores the determination of semantic similarity measures on Thesauri or Semantic Networks. The similarity between terms or concepts can be measured by quantifying the relatedness between the words utilised in knowledge obtained from certain information sources. Zhang [19] measured the similarity between words in information retrieval using web documents. Also, Navigli [20] used semantic similarity to disambiguate word sense between words in WordNet while Kaza and Chenn [21] improved the accuracy of semantic concepts. The methods are based on linguistic knowledge and thus provide a more precise representation than co-occurrences or bag-of-word models. Maki [22] obtained semantic similarity or distance on the basis of WordNet. Meanwhile, WordNet senses are in form of nouns, verbs, adverbs and adjectives which were organised by a variety of semantic relations into synsets. The concepts extracted from WordNet are used to disambiguate words regarding the context of the document [23]. However, ontology in these cases typically served as thesauri contained synonyms, hypernym /hyponyms but did not consider the context of each term relations.

WordNet has many synsets and a particular synset may have more than one sense. But word sense disambiguation results in a single decision. A hierarchical structure (Domain Ontology) can represent the context taking circumstances in which an event happens or considered.

Various approaches have been used to quantify the similarity between concepts in ontology while still maintaining information contained in the hierarchical structure [24] therefore, the existing systems [3, 4] cannot resolve the semantic issues of polysemy or synonyms because they require identification of the context of keywords to comprehend their actual semantics. Moreover, the existing systems ignore other important relationships such as semantic neighbourhoods that can contribute to useful search results [25].
To overcome the limitations of existing semantic searching systems, one needs to represent the context of terms through *IS-A* hierarchy for effective searching using domain knowledge [9]. With domain ontology, a particular sense is based on IS-A hierarchy concept by relating it to the actual domain concepts. The system concentrates on searching terms using IS-A hierarchy and not on the individual keywords. Paralic and Kostial [26] proposed an ontology-based approach to information retrieval where document resources are associated with concepts in ontology. They focused on query processing where concepts were matched to corresponding concepts in the ontology. The query concepts were matched with the document concepts and matched documents would be retrieved.

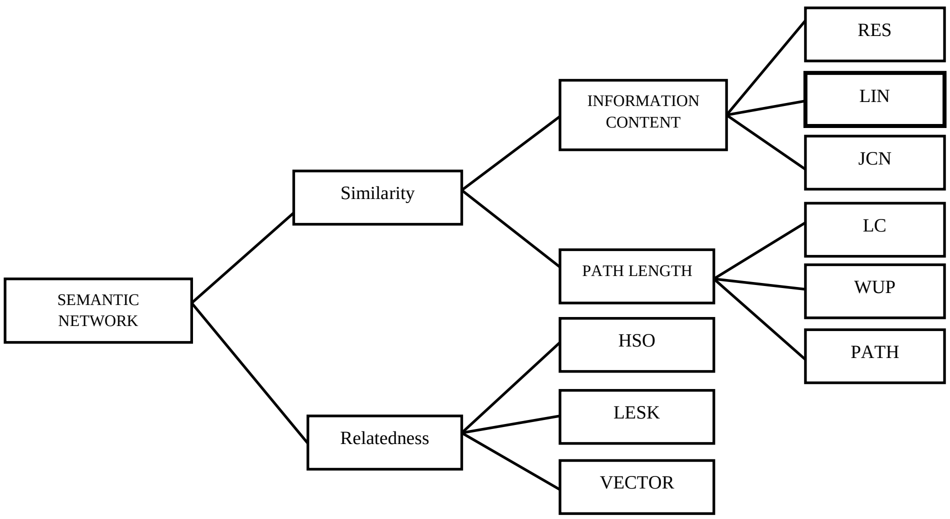

Figure 1: Knowledge-Based Similarity (adapted from [27])

1. Path Length Measures

Matching similarity based on linguistics (i and ii) is considered as analysing entities in isolation while ignoring the relationships with other entities. The path length measures the similarity between two concepts as a function of the length of the path linking the concepts and the position of the concepts in the taxonomy in WordNet. This use link or edge as parameter to refer to the relationships between concept nodes.

i. *The Shortest Path Based Measure:* The measure only takes $len(c_1, c_2)$ into consideration. Knappe [28] assumed that the $sim(c_A, c_B)$ depends on how close the two concepts are in the taxonomy and measures variant on the distance method.

$$sim_{path}(c_A, c_B) = 2 * depth\_\max - len(c_A, c_B) \quad (1)$$

ii. *Wu and Palmer similarity* introduced a scale that measures the position of concepts c_a and c_b in the taxonomy. Varelas [29] described the conceptual distance between two nodes and proportional to the number of edges at:

$$sim_{wp}(c_A, c_B) = \frac{2 * depth(lcs(c_A, c_B))}{len(c_A, c_B) + 2 * depth(lcs(c_A, c_B))} \quad (2)$$

iii. *Leakcock & Chodorow's Measure* proposed the maximum depth of taxonomy, However, they limit their attention to IS-A links and scale the path length by the overall depth of the taxonomy [30] and has the following measure:

$$sim_{LC}(c_A, c_B) = -\log \frac{len(c_A, c_B)}{2 * deep\_\max} \quad (3)$$

2.0 Information Content-Based Measure

Information Content (IC) assumes that each concept is associated with much information in WordNet.

i. Resnik proposed a similarity measure an information-based statistic method which defined the Information Content (IC) of each concept. The similarity depends on the information content that is subsumed in the taxonomy, this bring together lexical database and corpora. Therefore, calculates similarity by considering the information content (IC) of the LCS of two concepts [31].

$$sim_{\mathrm{Res}}(c_A, c_B) = -\log P^n(lcs(c_A, c_B)) = IC(lcs](c_A, c_B)) \quad (4)$$

ii. Lin proposed similarity measure based on information content and used both the amount of information needed to state the commonality between two concepts and information needed to describe them. Saruladha *et al.* [32] used information content to determine the common concepts and presented the common information content by finding the common features of the compared entity classes.

$$sim_{Lin}(c_A, c_B) = \frac{2 * IC(lcs(c_A, c_B))}{IC(c_A) + IC(c_B)} \quad (5)$$

3. Hybrid Measure

i. *Jiang and* Conrath *Measure:* Calculated semantic distance derived from the edge-based notion of distance with the addition of the information content as a decision factor to obtain semantic similarity. According to Pirrò [10] and Hirst and Budanitsky [33], the measure provided the best results when measuring semantic relatedness. The measure is semantic relatedness not semantic similarity between concepts.

$$dis_{J\&C}(c_A, c_B) = IC(c_A) + IC(c_B) - 2IC(lcs(c_A, c_B)) \quad (6)$$

3.0 Methodology

The Conceptual Knowledge for alleviating word semantic similarity in retrieval system consists of two phases. The phases are describes as follows: the extraction of two related domains for querying the search engines and how these queries are searched in the corpora information source respectively. If these two related terms *A* and *B* are used as query, would the similarity values be the same? The last phase improves the first phase by using additional knowledge information source for searching the queries. Finally, integration of the phases from the sources is combined to determine the similarity of words.

3.1 Model User' Query

The two related domains describe the set of collection of entities (classes) as an instance derived from concepts. The classes consist of concepts and sub-concepts where the property indicates the relationship of each concept. Tomassen [34] presented method of indexing documents with ontology vocabulary of the index terms derived from the ontology to the domain terminology. This method is adapted with each concept from the domain used as a search term. As the queries vocabulary are controlled, the web resource is also structured to suit the controlled queries. This will enable conceptual match between extracted concepts that are relevant to the document in the knowledge repository as an alternative to keyword match as depicted in figure 2. This requires additional source (WordNet) to adjust the similarity score to semantic level.

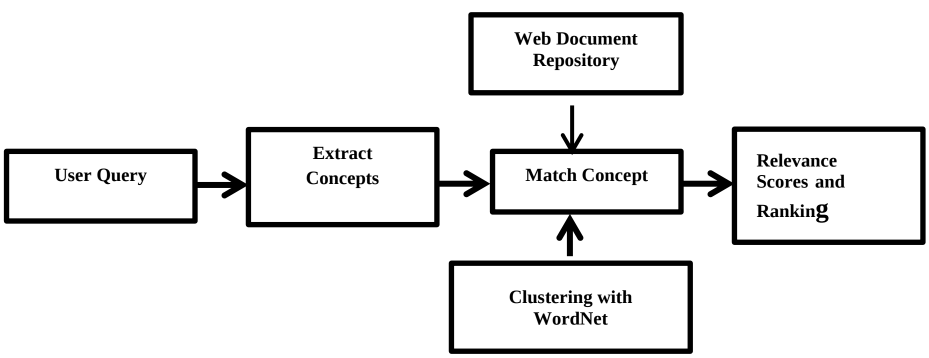

Figure 2: Conceptual Match in Document Source

3. 2 Adjust Similarity Value with Semantic Network

The similarities in Document Source failed to deal with words not covered by synonym dictionaries or are not able to cope with acronyms, abbreviations, buzzwords etc. But conceptual knowledge in semantic network source uses some kind of web intelligence to determine the degree of similarity between text expressions. The same concepts query *A* and *B* used in Document are also used in semantic similarity in WordNet lexicon to adjust the word similarity. The WordNet relational dictionary (WordNetSimilarity) calculates the semantic similarity of different measures in the knowledge source. This WordNetSimilarity accessed information contained in concepts of a query and determined the similarity between query *A* and *B* but this required human intervention. Meng [18] used path and information content that is inversely proportional to length (CA, CB) but does not consider the position of the hierarchies of concepts A and B.

Also, the Information Content (IC) similarity in WordNetSimilarity lacks the path from hierarchical structure of concepts. Hierarchy senses that relates to the documents chosen were also used in WordnetSimilarity. This describes the Information Content (IC) with extension of Path Length which can be named as LIPA as shown below:

$$sim_{LIPA}(c_a, c_b) = \frac{2 * IC(lcs(c_a, c_b))}{IC(c_a) + IC(c_b)} * (1 - k)^l \qquad (5)$$

where

sim_{LIPA} = *Similarity between concept A and B*

lcs = least common subsume

$\begin{cases} if\ l = 0, then\ C_a = C_b\ \ else \\ \quad l = \max\ C_a\ or\ C_b \end{cases}$

K is a parameter and $0 \le k \le 0$, which can be adapted manually to make the metrics to get the best performance, however, the threshold k values of 0.5 is assumed for the two words or terms to be similar or related in WordNet and the maximum length *l* of taxonomy between the two concepts. An Independent software package developed by Ted Pederson [35] is used to calculate the similarity based on WordNet2.1. This involves semantic similarity measures described by Leacock and Chodorow, Jiang and Conrath, Resnik, Lin, Wu and Palmer and the developed LIPA.

3.4 *The Conceptual Knowledge is presented below with following Algorithm:*

1. *get dataset of the ontology A and B*
2. *Ontology A ,B ∈ concept hierarchy (* $C_1, C_2, C_3 \cdots C_n$ *)*
3. *for each* $C_1, C_2, C_3 \cdots C_n$
4. *each* $C_1, C_2, C_3 \cdots C_n$ *as search term*
5. *retrieve textual data for the two related concepts A and B*
   Select snippet relate to entities
6. *Search Ontology a and b in WordNetSimilarity*
   Let constant K = 0.5 for minimum Similarity of a and b
   for each t_i *in Ontology_a*
   for each t_i *in Ontology_b:*
   loop all t_i *in ontology_b*
   get Hypernyms of t1 from WordNet = HYPERNYMS_a
   get Hypernyms of t2 from WordNet = HYPERNYMS_b
   select Best-fit hypernym from HYPERNYMS _a = HYP _BEST_FIT a
   select Best-fit hypernym from HYPERNYMS _b = HYP_BEST_FIT_b

   LCS_t1_t2 = Least Common Subsumer of (c_a, c_b*) of (HYP_BEST_FIT_a, HYP_BEST_FIT_b)*

   L1 = Longest path length from c_a *to LCS_* c_a*_*c_b *in WordNet*
   L2 = Longest path length from c_b *to LCS_* c_a*_*c_b *in WordNet*
   Length = IF t1 == t2 THEN 0 ELSE L1 > L2 ? L1 : L2
   IC = IC(LCS_t1_t2)
7. *Compute SIM_*c_a*_*$c_{b(LIPA)}$ *= ((2 * IC) / (IC(t1) + IC(t2))) * (I - K) ^ Length*

4.0 Evaluation and Discussion of Semantic Similarity Measures

In this section, the exiting methods listed for each of the similarity measures in section II would be compared with proposed method to evaluate the performance of the word semantic similarity in retrieval system. An experiment was set up to compute the similarity of a set of word pairs and examine the correlation between human judgement and WordNet. Miller and Charles [36] had presented a sample of 30 noun pairs and these were used as our experimental dataset.

From the results, it shows that Resnik (Information Content) and Leakcock & Chodorow's (Path Length) generated results that is above similarity level $0.5 \le wt \le 1$. The result generated for the proposed semantic similarity is better than when individual methods are used.

TABLE I: WORD SEMANTIC SIMILARITY MEASUREMENT

WORD PAIR		KNOWLEDGE SIMILARITY					
		$Dist_{JCN}$ (Hybrid)	Sim_{LIPA} (New Hybrid)	Sim_{Path}	Sim_{LIN}	Sim_{RES}	Sim_{LC}
automobile	Car	0.000	**1.000**	1.000	1.000	6.026	3.637
gem	Jewel	0.000	**1.000**	1.000	1.000	10.435	3.637
journey	Voyage	0.164	**0.903**	0.500	0.653	5.735	2.944
boy	Lad	0.292	**1.064**	0.500	0.814	7.491	2.944
coast	Shore	1.615	**1.216**	0.500	0.966	8.8515	2.944
asylum	Madhouse	2.466	**1.232**	0.500	0.982	11.416	2.944
magician	Wizard	0000	**1.000**	1.000	1.000	11.821	3.637
midday	Noon	0000	**1.000**	1.000	1.000	10.318	3.637
furnace	Stove	0.055	**0.218**	0.0909	0.214	2.476	1.239
Food	Fruit	0.088	**0.142**	0.100	0.138	0.909	1.335
bird	Cock	0.206	**1.010**	0.500	0.759	7.655	1.335
bird	Crane	0.000	**0.000**	0.250	0.000	7.655	2.251
implement	Tool	0.817	**1.170**	0.500	0.920	7.014	2.251
brother	Monk	0.060	**0.255**	0.125	0.224	2.419	2.251
crane	Implement	0.000	**0.000**	0.071	0.000	1.436	0.995
brother	Lad	0.080	**0.310**	0.143	0.278	2.418	2.419
car	Journey	0.066	**0.001**	0.044	0.000	0.000	1.692
monk	Oracle	0.057	**0.248**	0.125	0.217	2.419	1.692
cemetery	Woodland	0.054	**0.142**	0.111	0.135	1.436	1.440
food	Rooster	0.067	**0.108**	0.071	0.109	0.909	0.998
coast	Hill	0.188	**0.835**	0.200	0.710	6.516	2.028
forest	Graveyard	0.054	**0.143**	0.111	0.135	1.436	1.440
shore	Woodland	0.060	**0.272**	0.200	0.147	1.436	2.028
monk	Slave	0.064	**0.299**	0.200	0.237	2.419	2.028
coast	Forest	0.057	**0.008**	0.083	0.000	0.000	1.152
lad	Wizard	0.073	**0.386**	0.200	0.261	2.418	2.028
chord	Smile	0.068	**0.282**	0.091	0.274	2.784	1.239
glass	Magician	0.056	**0.147**	0.083	0.139	1.436	1.153
rooster	Voyage	0.047	**0.001**	0.041	0.000	0.000	0.459
noon	String	0.047	**0.005**	0.055	0.000	0.000	0.747

5.0 Conclusion and Future Work

In this paper, we present a concept semantic similarity matching based on information content with hierarchies of the two words from WordNet. The semantic matching approach is found by computing semantic similarity among different ontologies. It was found that if a query is replaced with synonymous words, this will improve the information retrieval system. Most users find it difficult to describe the information they want to retrieve in the search query which lead to poor retrieval.

In future work, the semantic matching approach can further be extended by computing semantic similarity among different ontologies. The algorithm presented can also further be enhanced by incorporating multiple document sources with WordNet.

References

[1]. Landauer, T. K. and Dumais, S. T. 1997. A solution to plato's problem: The latent semantic analysis theory of the aquisition, induction, and representation of knowledge. Psychological Review, 104, 211-240.

[2]. Croft, W. B., Bendersky, M.; and Metzler, D. 2010. Learning concept importance using a weighted dependence model. In WSDM, 31–40.

[3]. Alipanah. N., Parveen, P., Menezes, S., Kha, L., Seida, S.B., and Thuraisingham, B. 2010. Ontology-driven query expansion methods to facilitate federated queries. In: Proceedings of 2010 IEEE International Conference on Service-Oriented Computing Applications (SOCA). IEEE, 1–8.

[4]. Rinaldi, A.M. 2009. An ontology-driven approach for semantic information retrieval on the web. ACM Transactions on Internet Technology (TOIT) 9.3: 10-14.

[5]. Lee, C.Y. and Soo, V.W. 2005. Ontology-based information retrieval and extraction. In: Proceedings of 3rd International Conference on Information Technology: Research and Education (ITRE). IEEE, 265–269.

[6]. Berners-Lee, T., Hendler, J., Lassila, O. 2001. The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. In Scientific American, Mai http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html. 285.5:34-43

[7]. Egozi, O., Markovitch, S., and Gabrilovich, E. 2011. Concept-Based Information Retrieval Using Explicit Semantic Analysis‖. ACM Transactions on Information Systems. April 2.8: 1-34.

[8]. Carpineto C.and Romano. G. 2012. A Survey of Automatic Query Expansion in Information Retrieval‖. ACM Comput. Surv.. DOI 10.1145/2071389.2071390 http://doi.acm.org/10.1145/2071389.2071390. January 44.1:1-50.

[9]. Khan, S. and Marvon, F. 2006. Identifying relevant sources in query reformulation. In: Proceedings of the 8th International Conference on Information Integration and Web-based Applications & Services (IIWAS). Yogyakarta, Indonesia, 99–130.

[10]. Pirrò, G. 2009. A semantic similarity metric combining features and intrinsic information content. Data and Knowledge Engineering, doi.10.1016j.datak. 68.11: 1289-1308.

[11]. Mihalcea, R., Corley, C. and Strapparava, C. 2006. Corpus-based and knowledge based measures of text semantic similarity. In Proceedings of the 21st National Conference on Artificial Intelligence Menlo Park: AAAI Press, 775-780.

[12]. Li, Y., Bandar, Z. A. and McLean, D. 2003. An approach for measuring semantic similarity between words using multiple information sources, IEEE Transactions on Knowledge and Data Engineering, July-August, 15. 4: 871 – 882.

[13].Roush, W. 2004. Search beyond Google. Technology Review. http://www.technologyreview.com/articles/print_version/roush0304.asp., March 2004.

[14]. Prathvi, K. and Ravishankar, K. 2013. Measuring Semantic Similarity between Words using Page-Count and Pattern Clustering Methods" International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, 3.2: 31-34.

[15]. Bollegala D, Matsuo Y. and Ishizuka M. 2007. Measuring semantic similarity between words using web search engines. Proceeding of International Conference on World Wide Web, 757-766.

[16]. Nirgude, V., Sharma, R. and Sedamkar, R. 2013. A Web Search Engine based approach to Measure the Semantic Similarity between Words using Page Count and Snippets Method (PCSM) International journal of advanced research in computer engineering & technology (IJARCET) ISSN: 2278 – 1323 July 2.7: 2252-2257.

[17]. Bollegala, D., Matsuo, Y. and Ishizuka, M. 2011. A web search engine-based approach to measure semantic similarity between words", IEEE Transactions on Knowledge and Data Eng., 23.7: 977-982.

[18]. Meng, L. Huang, R. Gu, J. 2014. Measuring Semantic Similarity of Word Pairs Using Path and Information Content. International Journal of Future Generation Communication and Networking 7.3:183-194

[19]. Zhang, F., Srihari, R. K. Z. and Rao, A. B 2000. Intelligent indexing and semantic retrieval of multimodal documents, Information Retrieval, 2:245-275.

[20]. Navigli, R. 2009. Word Sense Disambiguation: A Survey, ACM Comput. Surv., 41.2:1-69.

[21]. Kaza, S. and Chen, H. 2008. Evaluating ontology mapping techniques: An experiment in public safety information sharing, Journal of Decision Support System., doi:10.1016/j.dss.2007.12.007. 45.4:714-728.

[22]. Maki, W. S., McKinley, L. N. and Thompson, A. G. 2004. Semantic distance norms computed from an electronic dictionary (WordNet). Behavior research methods, instruments, and computers, 36: 421-431.

[23]. Baziz, M., Boughanem, M., Aussenac-Gilles, N. 2004. The Use of Ontology for Semantic Representation of Documents. In The 2nd Semantic Web and Information Retrieval Workshop (SWIR), SIGIR 2004, Sheffield UK, 29. Yin Ding, Keith van Rijsbergen, Iad Ounis, Joemon Jose (Eds) July 38-45.

[24]. Schickel-Zuber, V. and Falting, B. 2007. OSS: A semantic similarity function based on hierarchical ontologies. In Proceeding of the 20th international joint conference on artificial intelligence, Morgan Kaufmann Publisher Inc. 551-556.

[25]. Rodriguez M. A. and Egenhofer M. J. 2003. Determining Semantic Similarity among Entity Classes from Different Ontologies, IEEE Trans. on Knowledge and Data Engineering, 15.2:

[26]. Paralic, J. and Kostial, I. 2003. Ontology-based Information Retrieval, Information and Intelligent Systems, Croatia, 23-28.

[27]. Gomaa, W H. and Fahmy, A. A. 2013. A survey of text similarity approaches. International Journal of Computer Applications (0975 – 8887) April 68.13: 13-18.

[28]. Knappe, R., Bulsko H. and Andreasen T. 2002. On Measuring Similarity for Conceptual Querying. Proceedings of the 5th International Conference on Flexible Query Answering Systems,October Copenhagen, Denmark, 27-29.

[29]. Varelas, G., Voutsak,i E., Raftopoulou, P., Petrakis, E. G. and Milios, E. E. 2005. Semantic similarity methods in WordNet and their application to information retrieval on the web. Proceedings of the 7th annual ACM international workshop on Web information and data management, November, Bremen, Germany.

[30]. Leacock, C. and Chodorow, M. 1998. Combining Local Context and WordNet Similarity for Word Sense Identification. In WordNet": An Electronic Lexical Database, C. Fellbaum, MIT Press.

[31]. Resnik,P 1999. Semantic similarity in a taxonomy: An information based measure and its application to problems of ambiguity in natural language". Journal of Artificial Intelligence Research, 11. 95- 130.

[32]. Saruladha, K Aghila, G. and Sathiya, B. 2011. A comparative analysis of ontology and schema matching system. International Journal of Computer Applications 34.8: 14-21.

[33]. Hirst, G. and Budanitsky, A. 2006. Evaluating WordNet-based measures of semantic distance," Comput. Linguist., 32.1: 13-47.

[34]. Tomassen, S. L. and Strasunskas, D. 2009. Construction of Ontology Based Semantic-Linguistic Feature Vectors for Searching: The Process and Effect. In Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - IEEE Computer Society, Washington, 3: 133-138.

[35]. Ted Pedersen and Jason Michelizzi. 2009. http://marimba.d.umn.edu/cgibin/ similarity/similarity.cgi.

[36]. Miller, G. A. and Walter G. Charles. 1991. Contextual correlates of semantic similarity", Language and Cognitive Process, 6(1):1-28.