A Survey and Comparison of WordNet Based Semantic Similarity Measures

[1]Ayesha Banu, [2]Syeda Sameen Fatima, [3]Khaleel Ur Rahman Khan
[1]Dept. of CS, Alluri Institute of Management Sciences, KU, Warangal, India
[2]Dept. of CSE, Univ. College of Engineering, OU, Hyderabad, AP, India
[3]Dept. of CSE, ACE Engineering College, JNTU, Hyderabad, AP, India

Abstract
Semantic Similarity relates to computing the similarity between concepts within ontology. We explore the different categories of approaches to compute semantic similarity and the most popular measures are evaluated using WordNet as the source ontology. We compare the measures using the benchmark dataset of Miller & Charles with WordNet to rank the measures category wise and overall.

Keywords
Semantic Similarity, WordNet, Least Common Subsumer, Synsets, Is-A Hierarchy

I. Introduction

Semantic Similarity is a measure used to compute the similarity between concepts within ontology. This is highly investigated research subject in the fields of data processing, Artificial Intelligence, linguistics and in particular, the field of the information retrieval which is largely based on the similarity identification measures between documents [1].

Ontology is an explicit formal specification of the world that we wish to represent for some purpose. Ontology specifies terms in the domain and relations among them (Gruber 1993) [2]. Ontologies can be distinguished into Domain Ontologies, representing knowledge of a particular domain, and General Purpose Ontologies representing common sense knowledge about the world [3].

WordNet is one of the general purpose ontology which attempts to model the lexical knowledge of a native speaker of English. It can be used as both a thesaurus and a dictionary. English nouns, verbs, adjectives, and adverbs are organized into synonym sets, called synsets, each representing a concept [4].

Finding similar concepts is a core task in the area of ontology alignment/merging. Remaining paper is organized as follows: Section II presents an overview of WordNet and available semantic

Similarity measures. Section III gives a comparison of the measures and discussions of the results. Section IV concludes the paper and References are included in Section V.

II. WordNet and Semantic Similarity Measures

WordNet1 is an on-line lexical reference system developed at Princeton University. WordNet attempts to model the lexical knowledge of a native speaker of English. WordNet can also be seen as ontology for natural language terms. It contains more than 100,000 terms, organized into taxonomic hierarchies. Nouns, verbs, adjectives and adverbs are grouped into synonym sets (synsets). As per WordNet statistics2 there are 117659 synsets in WordNet 3.0.

The synsets (or concepts) are related to other synsets higher or lower in the hierarchy by different types of relationships.

The most common relationships are the Hyponym/Hypernym (i.e., Is-A relationships), and the Meronym/Holonym (i.e., Part-of relationships). Fig. 1 shows a fragment of WordNet Is-A hierarchy.

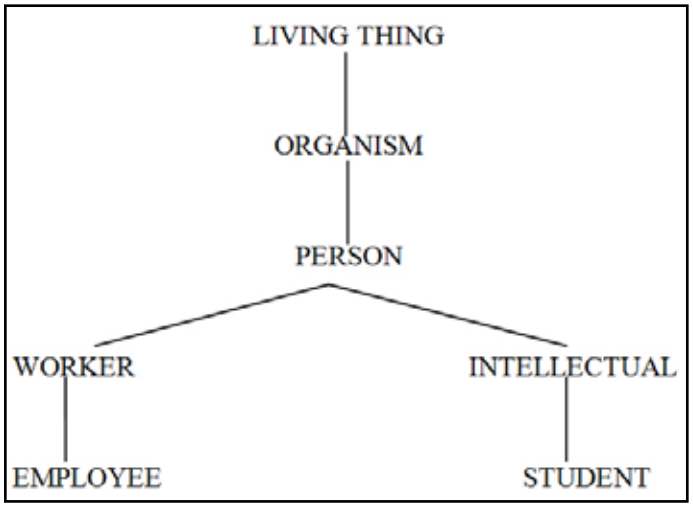


Fig. 1: Fragment of WordNet Is-A Hierarchy

Semantic Similarity measures are classified in to four main categories.

A. Edge Counting Measures

Let O be an ontology and c1,c2 ∈ C.These methods measure the similarity between two concepts c1, c2 by determining the path linking the terms in the taxonomy and the position of the terms in the taxonomy.

(i) Path Length:(1989) Rada et al. [5] proposed this measure.

DistPath(c1, c2)= d(c1, c2) (1)

Where d(c1, c2) is the shortest path between the concepts c1, c2.

(ii) Leacock & Chodorow:(1998) This measure [6] also uses the path length value along with the depth of the taxonomy given as

$$\text{SimLC}(c1, c2) = -\log\left(\frac{d(c1, c2)}{2D}\right) \quad (2)$$

Where d(c1, c2) is the shortest path between the concepts c1, c2 and D is the depth of the taxonomy.

(iii) Wu and Palmer: (1994) This measure [7] Considers the depth of the Least Common Subsumer or the Closet Common Parent Cp for the concepts c1, c2. The measure is given as

$$\text{SimWP}(c1, c2) = \frac{2Np}{N_1 + N_2 + 2Np} \quad (3)$$

is the depth of Cp from root, N_1 is depth of c1 from Cp and N_2 is depth of c2 from Cp. N_1+N_2 will result in shortest path between c1 and c2. Depth here is the number of Is-A links.

(iv) Mao et al. Measure:(2002) This Measure [8]defines a similarity measure using both shortest path information and number of descendents of compared concepts. The measure is given as

$$\text{SimMao}(c1,c2)= \frac{\delta}{d(c1,c2)\log_2(1+d(c1)+d(c2)} \quad (4)$$

Where d(c1,c2) is the number of edges between c1 and c2, d(c1),d(c2) the number of descendants of c1,c2. δ is a constant whose value is set to 0.9.

(v). Li et al. Measure :(2003) This measure [9] combines the shortest path length(number of edges) between the concepts c1, c2 (L) and the depth of the closest common parent(Np). The measure is given as

$$\text{SimLi}(c1,c2) = e^{-\alpha L}\left[\frac{e^{\beta Np}-e^{-\beta Np}}{e^{\beta Np}+e^{-\beta Np}}\right] \quad (5)$$

α ≥0 and β≥0 are parameters scaling the contribution of shortest path length and depth respectively. As per Li et al. the optimal parameters are α=0.2 and β=0.6.

(vi). Concept Specificity Measure (2006) Al-Mubaid & Nguyen [10] propose a similarity measure where they assume every branch of the ontology at root node as one cluster.

$$\text{SimCS}(c1,c2)= \log\left[(path-1)^{\alpha}(CSpec)^{\beta}+k\right] \quad (6)$$

Where CSpec(c1,c2) = D – Np , Path is the shortest path between c1,c2. α>0, β>0 are contribution factors of two features; k is a constant. The values of α,β,k are set to 1 experimentally.

(vii) Super Concept based Similarity:(2011) This measure proposed by M.Batet et al.[11]for c1, c2 C. define a set T(Ci)={Ci} {CjC | Cj is the super concept of Ci}The similarity between 2 concepts c1,c2 is given as

$$\text{Sim}_{\log}= -\log_2\left[\frac{|T(c1)\cup T(c2)|-|T(c1)\cap T(c2)|}{|T(c1)\cup T(c2)|}\right] \quad (7)$$

The Edge Counting Methods consider only the shortest path between the concept pairs. However, wide and detailed ontologies such as:

WordNet, MeSH incorporate multiple taxonomical inheritance, resulting in several taxonomical paths. By taking only the minimum path between concepts, many of the taxonomical knowledge explicitly modelled in the ontology is omitted. Another problem of path-based measures typically admitted is that they rely on the notion that all links in the taxonomy represent a uniform distance.

B. Information Content-Based Measures

In view of the limitations of edge-counting approaches, Resnik proposed to complement the taxonomical knowledge provided by an ontology with a measure of the information content of concepts computed from corpora like WordNet. The idea behind semantic similarity information content methods is that the similarity of two concepts is related to information they share in common.

(i). Resnik's Measure: (1995) As per resnik [12] for any concept C the information content is given as:

$$\text{IC}(C) = -\log P(C) \quad (8)$$

Where P(C) is the probability of the concept C in the corpora. The probability is computed as

$$P(C) = \text{freq}(C) / N \quad (9)$$

N is the total number of terms in the taxonomy which in WordNet is the number of noun synsets. As per WordNet statistics2 the value is 82115. The frequency of any concept is given in WordNet as noun frequency. The frequency of concept Employee of figure 1 given by WordNet in brackets is

Noun

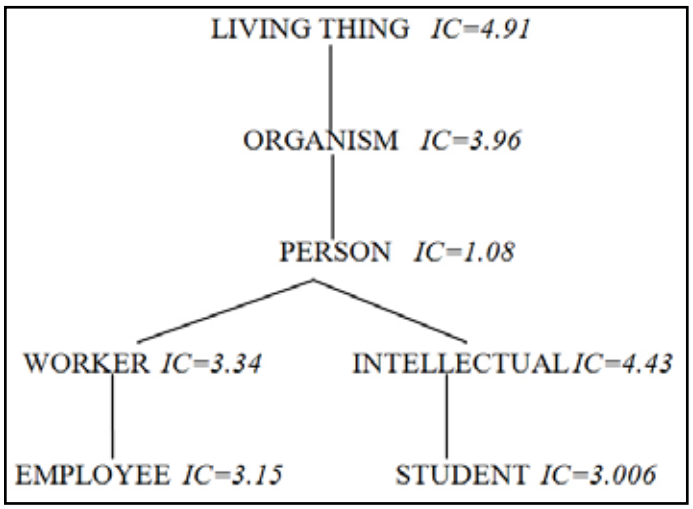
(57)S: (n) employee (a worker who is hired to perform a job)

The IC values computed for the concepts of fig. 1 are shown in fig. 2.

The resnik's semantic similarity measure can now be given as

$$\text{SimRES}(c1,c2) = \text{IC}(\text{LCS}(c1,c2)) \quad (10)$$

One of the problems of Resnik's proposal is that any pair of terms having the same LCS results in exactly the same semantic similarity.



(ii) Jiang and Conrath measure:(1997) This measure[13] is proposed using resniks measure

$$\text{Sim}_{JC}(c1,c2) = (\text{IC}(c1) + \text{IC}(c2) - 2* \text{Sim}_{RES}(c1,c2) \quad (11)$$

This based on quantifying the length of the taxonomical links as the difference between the IC of a concept and its subsumer. When comparing term pairs, they compute their distance by substracting the sum of the IC of each term alone from the IC of their LCS

(iii). Lin's Measure: (1998)As per Lin[14] the similarity between two terms should be measured as the ratio between the amount of information needed to state their commonality and the information needed to fully describe them. His measure considers commonality in the same manner as Resnik's approach on one hand and the IC of each concept alone on the other hand.

$$\text{Sim}_{LIN}(c1,c2) = 2 * \text{Sim}_{RES}(c1,c2) / \text{IC}(c1)+\text{IC}(c2) \quad (12)$$

(iv). Lord et al. Measure:(2003) This measure[15] is given as

$$\text{Sim}_{Lord}(c1,c2)=1- \text{Sim}_{RES}(c1,c2) \quad (13)$$

(v). Seco et al. Measure: (2004) This measure [16] considers the hyponyms of the WordNet to calculate the Information Content value. The similarity function is obtained by normalizing and appliyng a linear transformation to the Jiang and Conrath formula. They argue that the more hyponyms a concept has the less information it expresses and concepts that are leaf nodes are the most specified in the taxonomy so the information they express is maximal. The information content value is computed as

$$\text{IC}_{WN}(c) = 1-[\log(\text{hypo}(c)+1) / \log(\max_{wn})] \quad (14)$$

where the function hypo returns the number of hyponyms of a given concept and maxwn is a constant that is set to the maximum number of concepts that exist in the taxonomy. The similarity value using (14) can be given as

$$Sim(c1,c2)= 1-\frac{ic_{wn}(c1)+ic_{wn}(c2)-2*sim_{res}(c1,c2)}{2} \quad (15)$$

Simres corresponds to Resnik's similarity function but now accommodating ICWN values.

C. Feature Based Measures

In above category of measures the features of the terms in the ontology are not taken into account.

The features of a term contain valuable information concerning knowledge about the term. The information provided by the input ontology can be exploited as features. For WordNet, concept synonyms (i.e., synsets, which are sets of linguistically equivalent words), definitions (i.e., glosses, containing textual descriptions of word senses) and different kinds of semantic relationships can be considered. The measure proposed by Tversky 1997 [17] considers also the features of terms in order to compute similarity between different concepts.

$$SimTversky(c1,c2)= \frac{|C1\cap C2|}{|C1\cap C2|+k|C1/C2|+(k-1)|C2/C1|} \quad (16)$$

where C1, C2 correspond to feature sets of terms c1 and c2 respectively and $k \in [0,1]$ defines the relative importance of the non-common characteristics. This measure scores between 1 (for similar concepts) and 0, it increases with commonality and decreases with the difference between the two concepts. In reverse to all the above presented measures, it has nothing to do with the taxonomy and the subsumers of the terms.

D. Hybrid Measures

This category of similarity measures, combine ideas from the above presented approaches, considering the path connecting the two terms in the taxonomy, the IS-A links of the terms with their parents in the graph and as well as features of the terms.

The measure proposed by Rodriguez and Egenhofer 2003[18] can be used both for single or cross ontology similarities. A similarity function determines similar entity classes by using matching methods over synonym sets, semantic neighborhoods and distinguishing features. The similarity function is a weighted sum of the similarity values for synonys sets, neighborhoods and features.

$$Sim_{R\&E}(c1^p,c2^q)=W_w*S_w(c1^p,c2^q)+W_u*S_u(c1^p,c2^q) + W_n*S_n(c1^p, c2^q) \quad (17)$$

The functions Sw , Su , and Sn are the similarity between synonym sets, features, and semantic neighborhoods between entity classes c1 of ontology p and c2 of ontology q . Weights Ww, Wu, and Wn are the respective weights of the similarity of each specification component.

$$Sim (c1,c2)= \frac{|C1\cap C2|}{|C1\cap C2|+\alpha|C1/C2|+(\alpha-1)|C2/C1|} \quad (18)$$

The difference between the above Equation(18) and the Tverksy function(16) is in the way α(or k) is computed.

In Tversky function k defines the relative importance of the non-common characteristics, but here α is computed as a factor of the depth where the two compared concepts are in each taxonomy.

$$\alpha(c1^p, c2^q) = \frac{depth(c1p)}{depth(c1p)+depth(c2q)}, \quad depth(c1^p)\leq depth(c1^p) \quad (19)$$

$$\alpha(c1^p,c2^q) = 1- \frac{depth(c1p)}{depth(c1p)+depth(c2q)}, \quad depth(c1^p)> depth(c1^p) \quad (20)$$

III. Comparison of Semantic Similarity Measures

In previous section we have presented a complete survey on almost all the semantic similarity measures available used to find the similarity between any 2 concepts c1 & c2 of an ontology O. In this section we made an attempt to compare these measures over 2 different benchmark datasets one of English terms(we call it is Dataset 1) and the other of Medical terms(we call it as Dataset 2). This comparison can rank the measures as per their efficiency to calculate the similarity value correlated with Human & Experts Judgments.

A. Dataset 1

As part of an investigation into "the relationship between similarity of context and similarity of meaning (synonymy)", Rubenstein and Goodenough [19] obtained "synonymy judgements"from 51 human subjects on 65 pairs of words. The pairs ranged from "highly synonymous" to "semantically unrelated", and the subjects were asked to rate them, on the scale of 0.0 to 4.0, according to their "similarity of meaning" For a similar study, Miller and Charles[20] chose 30 pairs from the original 65, taking 10 from the "high level (between 3 and 4. . .), 10 from the intermediate level (between 1 and 3), and 10 from the low level (0 to 1) of semantic similarity", and then obtained similarity judgments from 38 subjects, given the same instructions as above, on those 30 pairs. We use the miller & Charles dataset as our Dataset 1 in the comparison.

B. Dataset 2

To make our comparison more effective we take Dataset2 which is pairs of medical terms with Physicians & Experts Score ranging from 1 to 4. This is a dataset of 30 concept pairs from Pedersen et al. [21] which was annotated by 3 physicians and 9 medical index experts. Each pair was annotated on a 4-point scale. We take only the Physician's judgment for our comparison.

C. Comparisons

We compared totally 8 measures from first 3 categories. The measures which we compared are of the equations (1) (2) (3) (5) (10) (11) (12) (13) on both Dataset 1 and Dataset 2. To rank the measures by their performance, each measure is evaluated again as the correlation of these results with the human judgment [in Dataset 1] and Experts Judgment[in Dataset 2] to find which measure produces a value closer to the Judgment value.

Table 1 shows the comparison of the 9 measures over Miller & Charles dataset and Table 2 shows the comparison over part of Pedersen's dataset.

We randomly selected 12 pairs of terms from the original dataset such that physician's judgment ranges from 4 to 1. We used the semantic similarity system developed by Technical University of Crete available at [22] for computation purpose.

D. Results & Discussions

As the result of comparing Dataset 1 over the 8 measures and computing the correlation of these values with Human Judgment (HJ) values we can observe that Leacock & Chodorow (LCH) measure (2) shows a closer correlation value of 0.828 in edge counting measures and the measure proposed by Jiang & Conrath(J&C) measure (12) shows a closer correlation value of 0.890 in information content measures. The overall performance

when considered J&C measure ranks first among all 8 measures compared.

Table 1: Comparison of Semantic Similarity Measures across Human Judgment on Dataset1

#	WORD PAIR	HJ	SP	LCH	WP	Li	RES	LIN	JIANG	LORD
1	car,automobile	3.92	1	3.58	1	0.99	0.67	1	1	0.49
2	gem , jewel	3.84	1	3.5	1	0.99	1	1	1	0.63
3	journey,voyage	3.84	0.97	2.89	0.92	0.81	0.65	0.84	0.87	0.48
4	boy,lad	3.76	0.97	2.89	0.93	0.81	0.76	0.86	0.88	0.53
5	coast , shore	3.7	0.97	2.89	0.9	0.81	0.77	0.98	0.98	0.54
6	asylum,madhouse	3.61	0.97	2.89	0.94	0.81	0.93	0.96	0.96	0.6
7	magician , wizard	3.5	1	3.58	1	0.99	0.79	1	1	0.54
8	midday , noon	3.42	1	3.5	1	0.99	1	1	1	0.63
9	furnace - stove	3.11	0.8	1.5	0.46	0.23	0.18	0.22	0.38	0.16
10	food - fruit	3.08	0.8	1.5	0.22	0.13	0.05	0.12	0.62	0.05
11	bird , cock	3.05	0.97	2.8	0.94	0.81	0.4	0.59	0.73	0.32
12	bird - crane	2.97	0.91	2.19	0.84	0.54	0.4	0.59	0.73	0.32
13	tool , implement	2.95	0.97	2.89	0.9	0.81	0.41	0.92	0.96	0.34
14	brother - monk	2.82	0.97	2.8	0.94	0.81	0.82	0.9	0.91	0.56
15	brother - lad	1.66	0.88	1.97	0.71	0.44	0.18	0.2	0.27	0.16
16	crane ,implement	1.68	0.88	1.97	0.66	0.44	0.23	0.36	0.59	0.21
17	journey , car	1.16	0	0.87	0	0	0	0	0.33	0
18	monk , oracle	1.1	0.8	1.5	0.58	0.24	0.18	0.21	0.34	0.16
19	cemetery , woodland	0.95	0.75	1.28	0.18	0.08	0.05	0.06	0.19	0.05
20	food , rooster	0.89	0.63	0.94	0.13	0.03	0.05	0.08	0.4	0.05
21	coast - hill	0.87	0.88	1.97	0.66	0.44	0.49	0.63	0.71	0.39
22	forest , graveyard	0.84	0.75	1.28	0.18	0.08	0.05	0.06	0.19	0.05
23	shore , woodland	0.63	0.86	1.79	0.44	0.3	0.08	0.1	0.3	0.07
24	monk , slave	0.55	0.88	1.97	0.71	0.44	0.18	0.23	0.38	0.16
25	coast , forest	0.42	0.83	1.63	0.4	0.25	0.08	0.1	0.28	0.07
26	lad , wizard	0.42	0.88	1.97	0.71	0.44	0.18	0.21	0.31	0.16
27	cord - smile	0.13	0.66	1.01	0.14	0.04	0.17	0.18	0.2	0.16
28	glass , magician	0.11	0.75	1.28	0.3	0.13	0.08	0.1	0.3	0.07
29	rooster , voyage	0.08	0	0.58	0	0	0	0	0.07	0
30	noon , string	0.08	0	1.09	0	0	0	0	0.17	0
CORRELATION		1	0.585	0.828	0.763	0.822	0.803	0.844	0.890	0.814

The comparison of Table 2 on pairs of medical terms also shows that LCH measure (2) has a closer correlation value of 0.733 in edge counting measures and the measure proposed by J&C (12) shows a closer correlation value of 0.406 in information content measures.

The overall performance when considered unlike the comparison of table 1, LCH measure ranks first among all 8 measures.

Table 2: Comparison of Semantic Similarity Measures across Physician's Judgment on part of Dataset2

#	WORD PAIR	PJ	SP	LCH	WP	Li	RES	LIN	JIANG	LORD
1	renal failure , kidney failure	4	1	1.3	1	0.99	0.84	1	1	0.57
2	heart , myocardium	3.33	0.95	1.04	0.85	0.77	0.59	0.83	0.88	0.44
3	stroke , infarction	3	0.8	0.6	0.6	0.42	0.37	0.57	0.72	0.31
4	delusion , schizophrenia	3	0.8	0.65	0.22	0.13	0	0	0	0
5	congestive heart failure , pulmonary edema	3	0.7	0.45	0.25	0.16	0	0	0	0
6	metastasis , adenocarcinoma	2.66	0.7	0.45	0.4	0.25	0.24	0.38	0.59	0.21

7	mitral stenosis , atrial fibrillation	2.33	0.8	0.6	0.6	0.42	0.45	0.45	0.45	0.36
8	rheumatoid arthritis , lupus	2	0.6	0.34	0.33	0.16	1	1.29	1.23	0.63
9	diabetes mellitus , hypertension	2	0.75	0.52	0.28	0.19	0	0	0	0
10	pulmonary fibrosis , lung cancer	1.66	0.85	0.69	0.66	0.51	0.51	0.64	0.71	0.4
11	appendicitis , osteoporosis	1	0.6	0.34	0.2	0.1	0	0	0	0
12	hyperlipidemia , metastasis	1	0.65	0.39	0.22	0.13	0	0	0	0
CORRELATION		1	0.725	0.733	0.62	0.645	0.364	0.376	0.406	0.3876

IV. Conclusions and Future Work

In this paper we explained almost all the measures available to compute semantic similarity between pairs of terms which we call as concepts. We experimented with Semantic Similarity Measurement Methods and evaluated their performance on 2 different datasets. One related to pairs of English terms with human judgment values and the other related to pairs of medical terms with expert's and physician's judgment. Our comparison results conclude that in both the Datasets LCH performs well among edge counting methods and J&C gives good results among Information content measures. But the overall comparison ranks J&C as top on DataSet1 whereas LCH as rank 1 on DataSet2.

All the methods discussed in this survey are applicable to single ontology only. The single ontologies to which our work was restricted were the WordNet ontology for English terms and Mesh ontology for medical terms.The concepts whose similarity was computed also belong to one single ontology.

These measures can be further used in cross ontology approach also which means the concepts for which the similarity value is computed do not belong to same single ontology. If there are 2 different ontologies O1, O2 then concept 1 belongs to O1 and concept 2 belongs to O2. Finding similar concepts is a core task in the area of ontology alignment/merging

References

[1] R. Baeza-Yates, B. Ribeiro-Neto,"Modern Information Retrieval", ACM Press; Addison-Wesley: New York; Harlow, England; Reading, Mass.,1999.

[2] Thomas R. Gruber,"Toward Principles for the Design of Ontologies Used for Knowledge Sharing", International Journal Human-Computer Studies 43, pp. 907-928, March 1993.

[3] R. Studer,Stefan Decker, Dieter Fensel, Steffen Staab, "Situation and Perspective of Knowledge Engineering", In J. Cuena and et al., editors, Knowledge Engineering and Agent Technology. IOS Press,Amsterdam, 2000.

[4] G. A. Miller, R. Bechwith, C. Felbaum, D. Gross, K. Miller,"Introduction to WordNet: An on-line lexical database", International Journal of Lexicography, 3(4), pp. 235-244, 1990.

[5] R. Rada, H. Mili, E. Bichnell, M. Blettner,"Development and application of a metric on semantic nets", IEEE Transaction on Systems, Man, and Cybernetics, pp. 17-30. 1989.

[6] Claudia Leacock, Martin Chodorow,"Combining local context and WordNet similarity for word sense identification", In Christianne Fellbaum. WordNet: An Electronic Lexical Database. The MIT press, 1998.

[7] Z. Wu, M. Palmer,"Verb semantics and lexical selection", In Proceedings of the 32nd Annual Meeting of the Associations for Computational Linguistics, pp. 133-138. 1994.

[8] Mao, W., Chu, W. W.,"Free text medical document retrieval via phrased-based vector space model", in Proc. of AMIA'02, San Antonio,TX, 2002.

[9] Yuhua Li, Zuhair A. Bandar, David McLean,"An Approach for Measuring Semantic Similarity between Words Using Multiple Information Sources", IEEE Transactions on Knowledge and Data Engineering, 15(4), pp. 871-882, 2003.

[10] Al-Mubaid, H., Nguyen, H.A.,"A Cluster-Based Approach for Semantic Similarity in the Biomedical Domain, In Proc. The 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBS, New York, USA, September 2006.

[11] Batet, M., Sánchez, D., Valls, A.,"An ontology- based measure to compute semantic similarity in biomedicine", Journal of Biomedical Informatics, 2011.

[12] Resnik P.,"Using Information Content to Evaluate Semantic Similarity in a Taxonomy",14th International Joint Conference on Artificial Intelligence, IJCAI 1995, Montreal, Quebec, Canada, Morgan Kaufmann Publishers Inc 448-453.

[13] Jiang, J. J., D. W. Conrath,"Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy", International Conference on Research in Computational Linguistics, ROCLING X, Taipei, Taiwan, 19-33.

[14] Lin, D,"An Information-Theoretic Definition of Similarity", 15th International Conference on Machine Learning (ICML98), Madison, Wisconsin, USA, Morgan Kaufmann, pp. 296-304.

[15] P.W. Lord, R.D. Stevens, A. Brass, C.A. Goble, "Investigating Semantic Simiilarity Measures across the Gene Ontology: The Relationship between Sequence and Annotation", Bioinformatics, 19(10), pp. 1275-83, 2003.

[16] Seco, N., T. Veale, J. Hayes,"An Intrinsic Information Content Metric for Semantic Similarity in WordNet", 16th Eureopean Conference on Artificial Intelligence, ECAI 2004.

[17] A. Tversky,"Features of Similarity", Psycological Review, 84(4), pp. 327-352, 1977.

[18] M.A. Rodriguez, M.J. Egenhofer,"Determining Semantic Similarity Among Entity Classes from Different Ontologies", IEEE Tramsactions on Knowledge and Data Engineering, 15(2), pp. 442-456, March/April 2003.

[19] Rubenstein, H., Goodenough, J.B.,"Contextual Correlates of Synonymy; Computational Linguistics; 8, pp. 627-633, 1965.

[20] [Online] Available: http://www.cs.technion.ac.il/~gabr/resources/data/wordsim353/
[21] Pedersen,T., Pakhomov, S., Patwardhan, S.,"Measures of Semantic Similarity and Relatedness in the Medical Domain", University of Minnesota Digital Technology Center Research Report DTC 2005/12.
[22] [Online] Available: http://www.intelligence.tuc.gr/similarity/index.php

Ayesha Banu is a Post Graduate in M.Sc (CS) from Kakatiya University in 2002 and M.Tech (CSE) from JNTUH in 2009 and she is pursuing her PhD from JNTUH in the Area of "Semantic Web Mining". She has 10 years of teaching experience and is presently working as Associate Professor in Department of Computer Science at Alluri Institute of Management Sciences. She has published papers in International Journals and Conferences.

Dr. Syeda Sameen Fatima obtained B.Tech. Electronics and Communication Engineering from JNTU in 1982,M.Phil., Computer Methods, University of Hyderabad, India in 1983 M.S., Computer Science, University of Massachusetts, Amherst, USA in 1993 and Ph.D. in Computer Science and Engineering, from Osmania University, India in 2004. Her Research interests Include Information Retrieval Systems, Data Mining, Artificial Intelligence, and Machine Learning. She has more than 25 years of teaching experience and is presently working as Professor and Head , Dept of CSE , Univ. College of Engineering , OU , Hyderabad. She has published papers in various National and International Journals and Conferences.

Khaleel Ur Rahman Khan obtained B.E. (CSE) from Osmania University in 1993 and M.Tech (CS) from JNTU in 1998. PhD in Computer Science from Osmania University in the area of Wireless Mobile Ad Hoc Networks. He is presently working as Professor and Dean at ACE Engineering College. His research interests include Heterogeneous Networks, Opportunistic Networks, Transaction Management in Ad Hoc and Sensor Networks, Data and Web Mining. He has published over 30 papers in various Peer Reviewed International Journals and conferences. He has teaching experience of over 18 years