Advances In Multidisciplinary & Scientific Research
A Multidisciplinary & Interdisciplinary Journal

Vol. 2 No. 4, December , 2016

Effect of Similarity Measures of Terms in Information Retrieval System

Abdullah, K.K.A.
Department of Computer Science
Olabisi Onabanjo University
Ago-Iwoye, Nigeria
E-mail: uwaizabdullah9@gmail.com
Phone: +2348060046592,

ABSTRACT

Similarity measures represent the degree of closeness or separation of the target objects. The most difficult part of retrieval system is when the closeness or separation is analyse in two different ways. However, most approaches to enhance similarity in retrieval system are based on single information source. In each of the approaches, there are shortcomings and the extents to which the degree of similarity can be determined. Therefore, different similarity measures are considered with their effect in retrieval system taking into consideration the semantic level of each of the measures.

Keywords: Information source, Similarity Measures, Query expansion, Retrieval

Aims Research Journal Reference Format:
Abdullah, K.K.A. (2016): Effect of Similarity Measures of Terms in Information Retrieval System.
Advances in Multidisciplinary Research Journal. Vol. 2. No. 4, Pp 219-228

1. INTRODUCTION

The similarity between terms or concepts (use interchangeably) can be measured by quantifying the relatedness between the words utilised in knowledge obtained from certain information sources. Different information sources are used to determine the similarity of terms in retrieval system. Zhang [1] measured the similarity between words in information retrieval using web documents. Web resources provide an important source of knowledge background for similarity measures. Text representation, categorisation, clustering and other applications are at the crossroads of information retrieval and machine learning. Similarity denotes how much this term contributes to the classification in the retrieval system. In {2] used the weighting measures for the information retrieval and text analysis. Query expansion is applied in information retrieval to solve the problem of word mismatch and ambiguity of terms that arose from differences in the words used by search engines [3]. However, users found it difficult to formulate query in search engines. Conesa [4] tried to reformulate web queries based on semantic knowledge about different application domains to expand the query. The query terms would be disambiguated so that it matched to a unique concept.

Many researchers use web search engines results as a resource and provide an efficient interface to the vast information. In [5] Google is used to determine relationship between pairs of concepts using Hearst pattern-based techniques. The strength is that it reduced the high cost of establishing adequate background knowledge. Indeed, the background knowledge sources are dynamically discovered and relied on combination of online available textual sources and thesauri [6]. Subsequently, this is a large collection of text documents that is used for language research which is also used for semantic similarity measure. A corpus-based determines the similarity between words according to information gained from large corpora. The measure provides better recall but suffer from lower precision since most of methods rely on a simple representation. Word sense is used and lexical concepts for indexing and retrieval [7] however, [8] used Single Value Decomposition (SVD) for representation of words that occurred in similar contexts. This did not solve the problem of co-occurrence of words.

This semantic-based information retrieval system utilised LSI techniques. But the approach was limited by employing analysis of semantics rather than by taking different measures or inherent semantics from texts. However, concepts or terms extracted can be used to disambiguate regarding to the context of the document [9]. Semantic similarity or distance on the basis of WordNet to explain human similarity judgments independently of associative strength, lexical co-occurrence or feature similarity [10]. Moreso, various approaches have been used to quantify the similarity between concepts while still maintain information contain in the structure [11]. Therefore, existing systems [12] cannot resolve the semantic issues of polysemy or synonyms because it requires identification of the context of terms to comprehend its actual semantics. Moreover, the existing systems also ignore other important relationships such as semantic neighbourhoods [13] that can also contribute to useful search results.

The rest of the paper is organised as follows: section 2 considered the information sources used in determining the similarity of term. Section 3 outlines the related work and their limitation in each of the similarity measures considered with their defect in relation to the information sources used. Finally, section 4 concludes the work and future work.

2. THEORETICAL FRAMEWORK ON INFORMATION SOURCES

The similarity between terms or concepts can be measured by quantifying the relatedness between the words utilised in knowledge obtained from certain information sources. In [1], measured the similarity between words in information retrieval using web documents. Semantic similarity used word sense to disambiguate between words in WordNet [14] while [15] improved the accuracy of semantic concepts.

These information sources can be categorised into :

ix. Multiple Document Source from web
x. Corpus-Based Resources
xi. Thesauri and Semantic Networks
xii. Domain Ontology Knowledge

This section explores the determination of similarity of terms by a number of information sources.

2.1 Information Source: Multiple Document Sources from Web

Web resources provide an important source of knowledge background for similarity measures. Many researchers use web search engines results as a resource and provide an efficient interface to the vast information) In [5] used Google to determine relationship between pairs of concepts using Hearst pattern-based techniques. The strength is that it reduced the high cost of establishing adequate background knowledge. Indeed, the background knowledge sources are dynamically discovered and [6] relied on combination of online available textual sources and thesauri. The problem similarity is addressed but introduced a novel method for measuring the similarity between short text snippets by leveraging web search results to provide greater context for the short texts.

2.2 Information Source: Corpus-Based Resources

This is a large collection of text documents that is used for language research and it is also used for semantic similarity measure. Furthermore, corpus-based determines the similarity between words according to information gained from large corpora. The measure provides better recall but suffer from lower precision since most of methods rely on a simple representation (depicted in Figure 1). LSA assumes words that are close in meaning that occurred in similar pieces of text where a matrix containing word counts is constructed from a large piece of text. LSA used a mathematical technique called Singular Value Decomposition (SVD) which is used to reduce the number of columns while preserving the similarity structure among rows. In [8] used SVD for representation of words that occurred in similar contexts. This did not solve the problem of co-occurrence of words. However, LSI techniques is used to enhance searches but the approach was limited by employing analysis of semantics rather than by taking different measures or inherent semantics from texts.

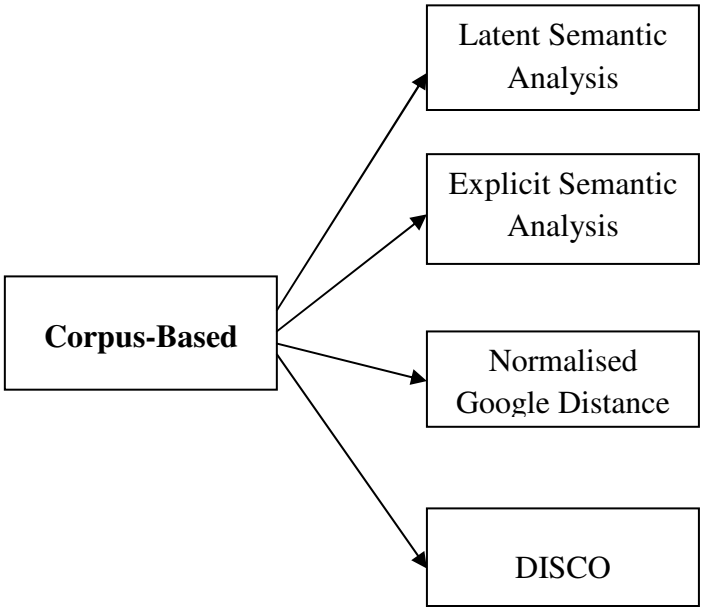

Figure 1: Corpus-Based Similarity Measure

To overcome the issues of corpus-based and lexical-based techniques while maintaining the precision or enhance precision, a semantic network-based approach to semantic similarity is used. The methods are based on linguistic knowledge and thus provide a more precise representation than co-occurrences or bag-of-word models.

2.3 Information Source: Semantic Network

According to Quillian, (1968) defined semantic network as

> "*Semantic network is broadly described as any representation interlinking nodes with arcs, where the nodes are concepts and the links are various kinds of relationships between concepts*".

The concepts extracted are used to disambiguate regarding to the context of the document) In [9]. In [18] obtained semantic similarity or distance on the basis of WordNet to explain human similarity judgments independently of associative strength, lexical co-occurrence or feature similarity. Ozcan and Aslangdogan (2005) extended each concept with similar words using a combination of Latent Semantic Analysis (LSA) and WordNet (Fellbaum, 1998) but the test performance showed a promising result.

2.4 Information Source: Domain Ontology Knowledge

WordNet has many synsets and a particular synset. A hierarchical structure can represent the context that is, circumstances in which something happens or should be considered. Therefore, the existing systems cannot resolve the semantic issues of polysemy or synonyms because it requires identification of the context of keywords to comprehend its actual semantics [17]. Moreover, the existing systems also ignore other important relationships such as semantic neighbourhoods and can also contribute to useful search results.

To overcome the limitations of existing similarity in retrieval system systems, one need to represent the context of terms through *IS-A* hierarchy for effective searching using domain knowledge. With domain ontology, a particular sense are chosen base on IS-A hierarchy concept by relating it to the actual domain concepts. The system concentrate on searching terms using IS-A hierarchy and not on the individual keywords.

3. THEORETICAL FRAMEWORK ON TERM-BASED SIMILARITY

In [2] used the weighting measures for the information retrieval and text analysis but documents are presented in high dimensional space. However, not every similarity measure is a metric, however similarity measure must satisfy the following four conditions:

Let a and b be any two objects in a set and $S(a,b)$ would be the similarity or distance between a and b

i. The similarity between any two points would be be non-negative: $S(a,b) \geq 0$

ii. The similarity between two objects would be one (1) if and only if the two objects are identical, that is, $S(a,b) = 1$ if and only if $a = b$

iii. Similarity would be symmetric, that is, distance from a to b would be the same as the distance from b to a , i.e. $S(a,b) = S(b,a)$

iv. The measure must satisfy the triangle inequality, which would be $S(a,c) \leq S(a,b) + S(b,c)$

The similarity between two object A and B can be easily computed. A variety of similarity or distance measures has been proposed and widely applied using term similarity measures in Figure 2. As shown, some of such measures are Jaccard, Correlation Coefficient, Euclidean Distance, Block Distance, Matching Coefficient, Dice Coefficient, Cosine Similarity and Radial Basis Function (non-linear) etc.

[1] Jaccard Coefficient

The Jaccard Coefficient (Tanimoto Coefficient) is a statistical measure of the extent of overlap between two vectors. It measures similarity as the intersection divided by the size of the union of the vector dimension sets. For text documents, the Jaccard coefficient compares sum weight of shared terms to the sum weight of terms that are present in the two objects. Term similarity analyses the simplicity and retrieval effectiveness but does not consider term frequency and rare term in a document collection. The definition is as follows:

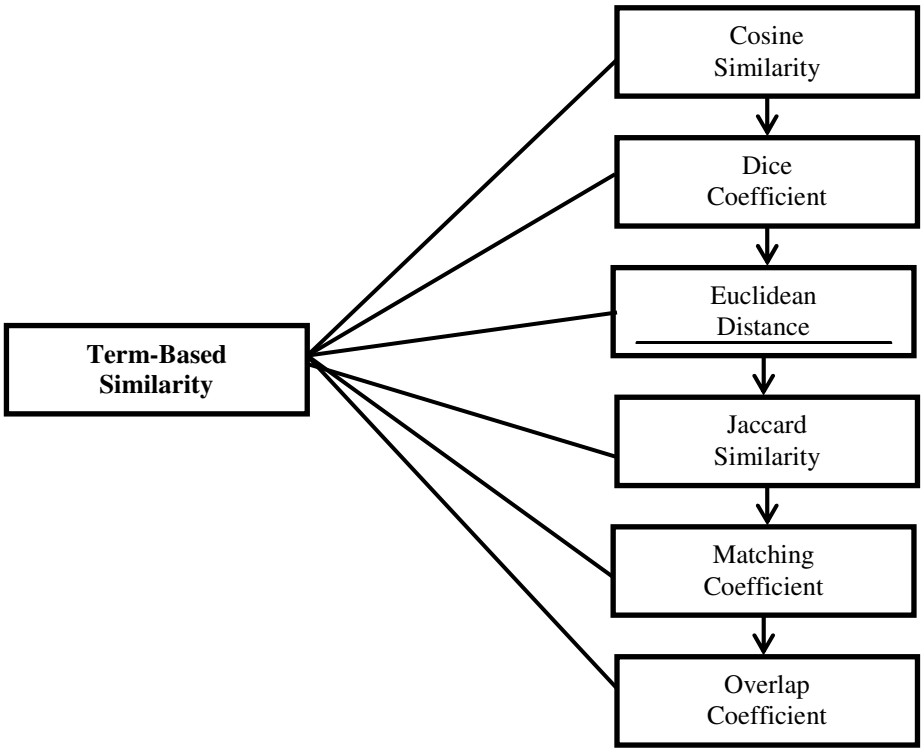

Figure 2. Term-Based Similarity Measures

$$SIM_J = (\overrightarrow{d_a}, \overrightarrow{d_b}) = \frac{\overrightarrow{d_a} \bullet \overrightarrow{d_b}}{\left|\overrightarrow{d_a}\right|^2 + \left|\overrightarrow{d_b}\right|^2 - \overrightarrow{d_a} \bullet \overrightarrow{d_b}} \quad (1.1)$$

where *d* is the document set a and b, SIM is similarity

The Jaccard coefficient (J) is a similarity measure values ranges between 0 and 1. If it is 1 then the $\overrightarrow{d_a} = \vec{d}_b$ and 0 when $\overrightarrow{d_a}$ and $\vec{d}_b$ are disjointed, where 1 means the two objects are the same and 0 means it is completely different. The corresponding distance measure is $D_J = 1 - SIM_J$.

[2] Overlap Coefficient

The overlap coefficient (Szymkiewicz-Simpson coefficient) is a similarity measure related to the Jaccard coefficient that measures the overlap between two sets.
It is defined as the size of the intersection divided by the smaller of the size of the two sets but considers two strings a full match if one is a subset of the others.

$$\text{Overlap}(a,b) = \frac{|a \cap b|}{\min|a| \cdot |b|} \quad (1.2)$$

If the set a is a subset of b or the converse, then the overlap coefficient is equal to 1.

[3] Manhattan (Block) Distance

A distance measure between two points along axes at right angle in a plane with p_1 at (a_1, b_1) and p_2 at (a_2, b_2) . Manhattan distance returns the maximum absolute difference in coordinates which corresponds to D = 1.

$$manh(a,b) = |a_2 - a_1| + |b_2 - b_1|$$

Therefore, it can be represented in form of weight (w) as:

$$manh(\vec{w}_a - \vec{w}_b) = \sum_{j-1}^{d} |w_{a,j} - w_{b,j}| \quad (1.3)$$

[4] The Dice Coefficient

Dice Coefficient measured intersection between two sets scaled by size giving a value in the range 0 to 1.

$$Dice(a,b) = \frac{2|a \cap b|}{|a| + |b|} \quad (1,4)$$

[5] Euclidean Distance

Euclidean distance is a standard metric for geometrical problems. It is the distance between two points and can be easily measured with a ruler in two or three-dimensional space. Euclidean distance is used in clustering problems. For example, K-means algorithm measured distance between text documents but large for vectors of different lengths. The Euclidean distance between $\overrightarrow{d_a}$ and $\overrightarrow{d_b}$ is large even though the distribution is very similar.

Given two documents d_a and d_b represented by term vectors $\vec{t_a}$ and $\vec{t_b}$ respectively, and weight (w), then Euclidean distance (*DE*) of the two documents is defined as:

$$D_E(\vec{d_a},\vec{d_b}) = (\sum_{t=1}^{m} \left| w_{d_a} - w_{d_b} \right|^2)^{1/2} \quad (1,5)$$

Where the term set is $T = \{t_1,..t_m\}$ as mentioned above. The $tfidf$ feature selection can be used in Euclidean term weights.

[6] Linear Kernel Function Similarity

Linear Kernel's measures of similarity is such that it calculates the dot product of two vectors $s(a,b) \succ s(a,c)$ if objects *a* and *b* are more similar than object a and c, then a kernel is positive. The function linear kernel is a polynomial kernel with a degree =1 and coefficient =0 (homogeneous). If a and b are column vectors, weight (w) and document (d) then linear kernel (k) is define as:

$$k(d_a, d_b) = \sum (w_a)^T w_b \quad (1.6)$$

It does not consider the optimisation problem and the computation becomes increasingly expensive with increasing simple size.

[7] Radial Basis Function (RBF)

The RBF is a non linear measure and it is used to map the data onto infinite dimensions. It computes the vector between two vectors. The minus sign in 2.17 inverts the distance measure into a similarity score due to its exponential. The similarity ranges from 1 to 0. RBF is applied in many science and engineering fields.

For document (*d*) a and b, γ is gama and weight (*w*). The kernel (k) is defined as:

$$k(d_a, d_b) = \exp(-\gamma \left\| w_a - w_b \right\|^2) \quad (1.7)$$

where

$\left\| w_a - w_b \right\|^2$ is the square of the Euclidean distance $\sum_{t=1}^{i} \left| w_{d_a} - w_{d_b} \right|^2$) between two a and b vectors.

RBF has few basic functions that cannot fit the training data adequately due to limited flexibility. On the other hand, those with too many basic functions yield poor generalisation abilities because of the limited flexibility of the RBF and its ability to erroneously fit the noise in the training data.

[7] Cosine Similarity

Documents are represented as term vectors. The similarity of two documents corresponds to the correlation between the vectors. This is quantified as the cosine of the angle between vectors, that is, cosine similarity. Cosine similarity is one of the most popular similarity measures applied to text documents in information retrieval applications and clustering. Similarity between a and b and Weight (w) is defined as:

Cosine Similarity (A,B) = $$\frac{|a \cap b|}{\sqrt{|a|} \bullet \sqrt{|b|}}$$

$$\text{Cosine } (W_a, W_b) = \frac{\sum_{j=1}^{n} (W_{aj} * W_{bj})}{\sqrt{\sum_{j=1}^{n} W_{aj}^2} \sqrt{\sum_{aj=1}^{n} W_{bj}^n}} \quad (1.8)$$

When two same copies of document d for example, are combined to get a new pseudo-document d', the cosine similarity between d and d' is 1. This means that these two documents are regarded to be the same. The purpose of normalisation is to make similarity of each element in a vector to be in the same range so that individual element gets the same weight when measures are applied. Vectors are normalised by sizes.

$$\left\|\vec{x}\right\|_2 = \sqrt{\sum_i x_i^2} \qquad (1.9)$$

where $1 \le i \le n$

Given two documents d_a and d_b, the cosine similarity is:

$$SIM_C = (\overrightarrow{d_a}, \overrightarrow{d_b}) == \frac{\overrightarrow{d_a} \bullet \overrightarrow{d_b}}{\left|\overrightarrow{d_a}\right| * \left|\overrightarrow{d_b}\right|} \qquad (1.10)$$

Where $\vec{d}_a$ and $\vec{d}_a$ are n-dimensional vectors over the term set $T = \{t_{1,\ldots,} t_n\}$, each dimension represents a term with its weight in the document. Cosine similarity is non-negative and bounded between [0, 1].

3.1 Theoretical Framework on Knowledge-Based Similarity

Finding similarity plays an important stage of text similarity. Similarity can be in two ways. These are lexical and semantic similarities. Lexical similarity can be done by different string term-based similarities while semantic similarity is done by corpus-based and knowledge-based algorithms. Pedersen [18] developed software called "WordNet::Similarity" that measures the similarity of concepts using different measures that used dictionary definition. This programme is used to compute conceptual similarity of words. Turney (2006) measured semantic similarity between words or concepts based on features of concepts and this plays an important role in many research areas such as Artificial Intelligence (AI), Natural Language Processing (NLP), cognitive science and knowledge engineering.

However, some of the most popular semantic similarity methods in Figure 2. are implemented and evaluated using WordNet as the underlying reference ontology. Semantic similarity measures with WordNet to enrich ontology with information about its leaf-nodes for disambiguation. However, disambiguation provides a small ranked list of WordNet-senses for each leaf node in the ontology hierarchies. These WordNet-senses are good candidates for the description of node as a whole or in parts.

Based on the WordNet utilisation, semantic similarity or distance measures between two concepts or words in any application can be classified into four categories:

1. Path length based measures
2. Information Content based measures
3. Feature based measures and
4. Hybrid measures.

3.1.1 Path Length Based Measures

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. It uses link or edge as parameter to refer to the relationships between concept nodes. The path length can be categorised into:

i. *The Shortest Path Based Measure:* The measure only takes $len(c_1, c_2)$ into consideration. The $sim(c_1, c_2)$ depends on how close the two concepts are in the taxonomy and measures variant on the distance method. The conceptual distance between two nodes is proportional using the number of edges separating the two nodes in the hierarchy.

For concept A and B in WordNetSimilarity, the following similarities are:

$$sim_{path}(c_A, c_B) = 2 * depth\_\max - len(c_A, c_B) \quad (1.11)$$

From equation 1.11, the similarity between two concepts (c_A, c_B) is the function of the shortest path $len(c_A, c_B)$ from c_A to c_B.

ii. *Wu & Palmer's Measure:* This similarity measure takes the position of concepts c_A and c_B in the taxonomy

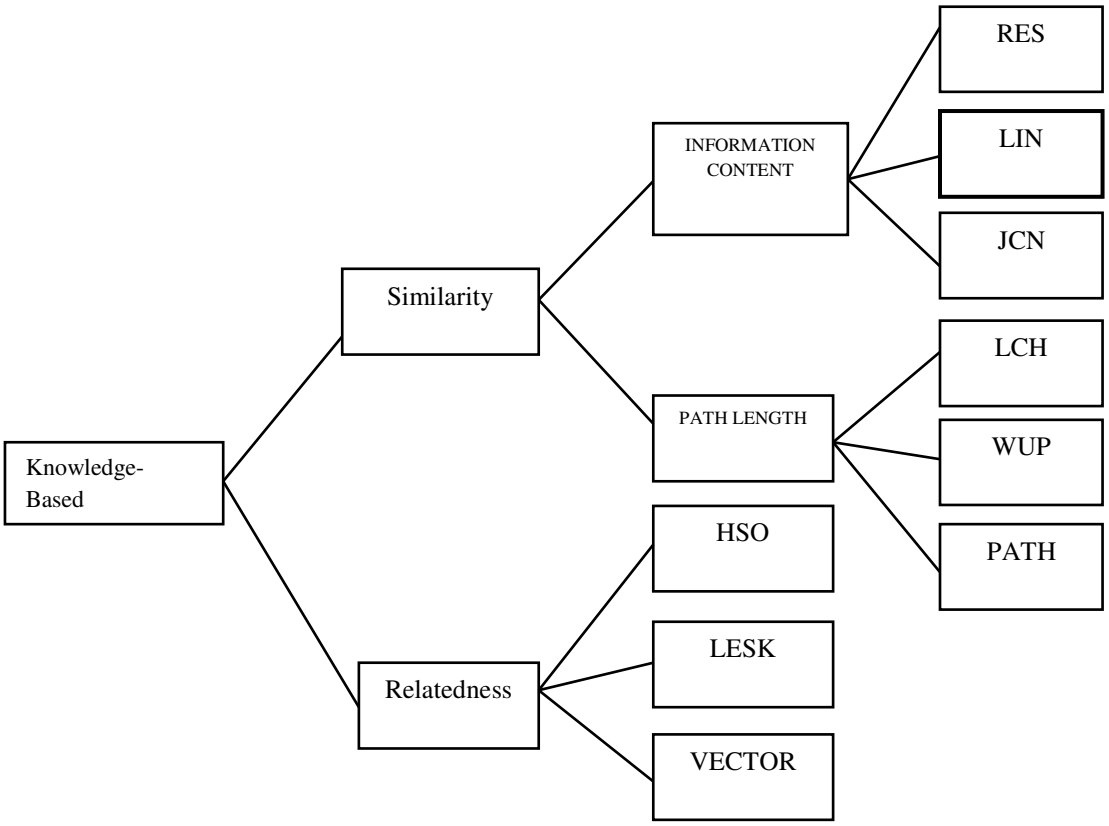

Figure 3: Knowledge-Based Similarity (adapted by Gomaa and Fahmy, 2013)

relatively to the position of the least common subsumer concept ($lcs(c_A, c_B)$) into account.

It assumes the similarity between two concepts as the function of path length and depth in path-based measures.

$$sim_{wp}(c_A, c_B) = \frac{2 * depth(lcs(c_A, c_B))}{len(c_A, c_B) + 2 * depth(lcs(c_A, c_B))} \quad (1.12)$$

From equation 1.12, the similarity between two concepts (c_A, c_B) is the function of the distance and the least common subsumer $lcs(c_A, c_B)$. It is not a similarity measure but a distance measure.

iii. *Leakcock & Chodorow's Measure:* Leakcock and Chodorow (1998) proposed the maximum depth of taxonomy and it has the following measure:

$$sim_{LC}(c_A, c_B) = -\log \frac{len(c_A, c_B)}{2 * deep\_\max} \quad (1.13)$$

From equation 1.13, the similarity between two concepts (c_A, c_B) is the function of the shortest $len(c_A, c_B)$ from c_A to c_B. The measure is based only on the positions of the concepts in the taxonomy but it assumes the links between concepts and represents its distances. All the paths have the same weight. However, it notes that the density of concepts throughout the taxonomy is not constant.

3.1.2 Information Content-Based Measure

Information Content (IC) assumes that each concept is associated with much information in WordNet. An information-based statistic method is based on the Information Content (IC) of each concept. The more common information two concepts share, the more similar the concepts are. This solved the problem to find a uniform link distance in path length based methods. Information content to determine the common concepts and presents the common information content by finding the common features of the compared entity classes [19]. This attempt to exploit the information contained to evaluate the similarity between the pairs of concepts. However, matching (term) similarity based on linguistics is considered as analysing entities in isolation while ignoring the relationships with other entities.

It was defined as:

i. *Resnik's Measure:* It assumes two concepts where the similarity depends on the information content that is subsumed in the taxonomy. In Resnik's measure, taxonomy of noun concepts in information content is calculated using the noun frequencies of each concept.

$$sim_{\mathrm{Re}snik}(c_A, c_B) = -\log P^n(lcs(c_A, c_B)) = IC(lcs](c_A, c_B)) \quad (1.14)$$

From equation 1.14, the values only rely on concept pair's lowest subsumer in the taxonomy. Resnik similarity has the problem of concept pair with the same lcs resulting in the same similarity values.

ii. *Lin's Measure:* Similarity measure based on information content and used both the amount of information needed to state the commonality between two concepts and the information needed to fully describe these terms/concepts.

$$sim_{Lin}(c_A, c_B) = \frac{2 * IC(lcs(c_A, c_B))}{IC(c_A) + IC(c_B)} \quad (1.15)$$

From equation 1.15, the measure has taken the information content of compared concepts into account and the values of this measure vary between 1 and 0. The length or distance between each concept in taxonomy is not considered.

iii. *Jiang's Measure:* This calculates 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 Jiang and Conrath method which provided the best results when measuring semantic relatedness.

$$dis_{J\&C}(c_A, c_B) = IC(c_A) + IC(c_B) - 2IC(lcs(c_A, c_B) \quad (1.16)$$

From equation 1.16, the measure has taken the IC of compared concepts into account and the value is semantic distance between two concepts not semantic similarity.

3.3 Feature-Based Measure

The feature-based measure is independent on the taxonomy and the subsumer of the concepts, although it attempts to exploit the properties of the ontology concepts to obtain the similarity values. This was based on the assumption that each concept is described by a set of words indicating its properties or features.

3.4 Hybrid Measure

In [13] presented hybrid measures that combined both the ideas of the methods and the relationship such as IS-A, part-of etc. in the taxonomy. Information content based measures and path based measures as parameter are commonly used. The measure is semantic relatedness not semantic similarity between concepts.

4. CONCLUSION

This paper presented different similarity measures with effect of information sources to adjust the weight of terms in order to generate an optimal result. Each of the measure is considered in relation to its semantic similarity level.

REFERENCES

[1] Zhang, F., Srihari, R. K. Z. and Rao, A. B 2000. Intelligent indexing and semantic retrieval of multimodal documents, Information Retrieval, 2:245-275.
[2] Newman, M. E. J. and Girvan, M. 2004. Finding and evaluating community structure in networks. Phys. Rev. E., 69: 26-113.
[3] Xu, J and Croft, W. B. 2000. Improving the effectiveness of information retrieval with local context analysis. ACM Trans. Inf. Syst., ISSN 1046-8188. doi:http://doi.acm.org/10.1145/333135.333138. 18.1:79–112.
[4] Conesa, J., Storey, V.C. and Sugumaran, V. 2006. Using Semantic Knowledge to Improve Web Query Processing, In: NLDB 2006, Springer-Verlag Berlin, 106 – 117.
[5] Cimiano, P.and Staab, S. 2004. Learning by Googling. SIGKDD Explor. Newsl., 6.2:24–33,
[6] van Hage, W. R. Katrenko, S. and Schreiber, G. 2005. A method to combine linguistic ontology-mapping techniques. In Proceeding 4th International Semantic Web Conference (ISWC), volume 3729 of Lecture notes in computer science, Galway (IE), 732–744.
[7] Mahesh, K., Kud J. and Dixon, P 1999. Oracle at TREC8: A Lexical Approach, In proceeding of the 8th Text Retrieval Conference (TREC-8), NIST special publication 500.
[8] Kwantes, P. J. 2005. Using context to build semantics. Psychological Bulletin and Review, 12: 703-710.
[9] 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.
[10] Yang, C. and Wu, S. 2011. A WordNet based information retrieval on the semantic web. In: Networked Computing and Advanced Information Management (NCM), 7th International Conference IEEE. 324–328.
[11] 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:
[12] Navigli, R. 2009. Word Sense Disambiguation: A Survey, ACM Comput. Surv., 41.2:1-69.
[13] Sahami, M. and Heilman, T.D. 2006. A web-based kernel function for measuring the similarity of short text snippets. In Proceedings of the 15th World Wide Web Conference, ACM, 377-386.
[14] 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.
[15] 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.
[16] Pedersen, T., Patwardhan, S. and Jason, M. 2004. WordNet:: Similarity - Measuring the Relatedness of Concepts. in the Proceedings of the Nineteenth National Conference on Artificial Intelligence (AAAI-04), San Jose, CA
[17] 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.