SEMANTIC SIMILARITY MEASURES BETWEEN WORDS: A BRIEF SURVEY

Ashraf Ali [1, 2 *], Fayez Alfayez [1] and Hani Alquhayz [1]
[1]Department of Computer Science and Information, College of Science Al Zulfi 11932, P.O 1221,
Majmaah University, Kingdom of Saudi Arabia
[2]International Center for Advanced Interdisciplinary Research (ICAIR), G-12/316, Ratiya Marg, Sangam Vihar,
New Delhi, India-110062
*Correspondence: a.haider@mu.edu.sa

ABSTRACT—*The semantic similarity measure is the ability to determine the similarity between various terms such as words, sentences, documents, concepts or instances. The aim of determining the semantic similarity measures between two sets of words is to find the degree of relevance by matching the words, which are conceptually similar but not necessarily lexicographically similar. Semantic similarity measure has great importance in many computer applications related field such as information retrieval, educational system, text summarization and natural language processing (NLP). There are several challenges to compute the semantic similarity between the words such as complexity of natural languages, the ambiguity of words and so on. One of the major challenges is the words are similar in meaning but they are not lexicographically similar. Traditional approaches for computing the semantic similarity is the major obstacle as they are not appropriate for many circumstances, many of the existing traditional approaches fail to deal with the term, which is not covered by synonyms and not able to handle with abbreviations, acronyms, brand names and so on. To overcome these problems, we present and evaluate the various promising methodologies that utilize several kinds of search engine based intelligence to determine the degree of similarity between the words. The objective of these types of methodologies is to utilize an assortment of paradigm including the study of text snippet comparison, frequent pattern finding, co-occurrence measures, trend analysis, and so on. The key objective is to replace the traditional methodologies where necessary.*

Keywords: Semantic Similarity Measure; NLP; Web Search Engine; Ontologies

1. INTRODUCTION

Computing the semantic similarity between words, terms, sentences, texts or statements which is same in meaning but not lexicographically similar is one of the critical tasks which have the major impact in many textual applications [9, 14]. In information retrieval, a similarity measure is used to assign a ranking score between a query and text in the corpus. Applications related to question-answer requires similarity identification between a question-answer. There are several types of ontologies used for computing the semantic similarity such as WordNet [9, 11], SENSUS [17], Cyc [27], UMLS [22], MeSH [24]. The diversity of natural language expressions makes it very difficult to compute the semantically equivalent terms. Whereas many applications have employed certain similarity functions to compute the semantic similarity between terms, most of the traditional approaches solving the problem by using manually compiled dictionaries such as WordNet [6]. The main problem is that a lot of terms (e.g. abbreviations, acronyms, brand names, buzzword etc.) that are not covered by these kinds of dictionaries. As a result, semantic similarity measures which are based on this type of resources cannot be used directly in these cases.

On the other hand, Web Search Engine (WSE) based approaches use some form of collective intelligence, which explores the potential and has promising collaborative work to solve the number of problems. We would like to utilize the benefit of the WSE collective intelligence for solving the problems related to the semantic similarity. To perform our experiments, we are going to utilize approaches that based on WSE (e.g. Google, Bing, Yandex, Ask etc).

This paper investigates and estimates the various promising approaches of semantic similarity to find out the degree of relevance between words using WSE based collective intelligence. We are mainly concerning those methods, which are able to intelligently measure the similarity between emerging terms and not frequently covered in dictionaries such as the method that consists of using the historical search patterns from WSE [15].

The remainder of this paper organized as follows: Section 2 reviews the various ontologies used for semantic similarity. Section 3 describes the related work. Section 4 describes the WSE based approaches for semantic similarity measure including the review of snippet comparison, page count based co-occurrence measure, frequent pattern finding, and trend analysis. Section 5 presents the statistical evaluation of the present methods using the benchmark of the dataset. Finally, we conclude the paper and presented future works direction of research in Section 6.

2. TYPES OF ONTOLOGIES USED FOR THE SEMANTIC SIMILARITY MEASURES

Over the years several types of ontologies available to use and utilized for computing the semantic similarity between the words including general purpose ontologies such as WordNet [9,11], SENSUS [17], Cyc [27] and domain-based ontologies such as UMLS [22] , MeSH [24], and STDS [1].

2.1 General-purpose Ontologies

General purpose ontologies are structured network of concepts that are interconnected by different types of assumption and semantic relations from multiple knowledge domains. These ontologies are developed to provide explicit specifications of general-purpose domains in a machine-readable and understandable format.

2.1.1. WordNet

WordNet [11] is a knowledge base in the form of the lexical database that stores the meaning of words and the relationship between them in a conceptually organized hierarchy. It is an online database which includes nouns, verbs, adjectives and adverbs grouped into a logical structure called synset. A synset represents a group of synonymous words that especially represents one underlying concept. A WordNet can be seen as the ontology for natural language terms and can be applied to compute the semantic similarity score. The latest version of WordNet is 3.1 announced in November 2012 and contains 155,287 words organized in

117,659 synsets for a total of 206,941 word-sense pairs, organized into taxonomic hierarchies [5]. Various kinds of relationships can be derived between the synsets or concepts. Synsets are organized into a conceptual hierarchy where synsets are linked together through various relations such as Hyponym/Hypernym relationship (i.e., Is-A relationship), and the Meronym/Holonym relationship (i.e., Part-Of relationship) are the most recognized relationships in WordNet. WordNet can be used as both a thesaurus and a dictionary. A portion of the WordNet Is-A hierarchy is demonstrated in Figure 1.

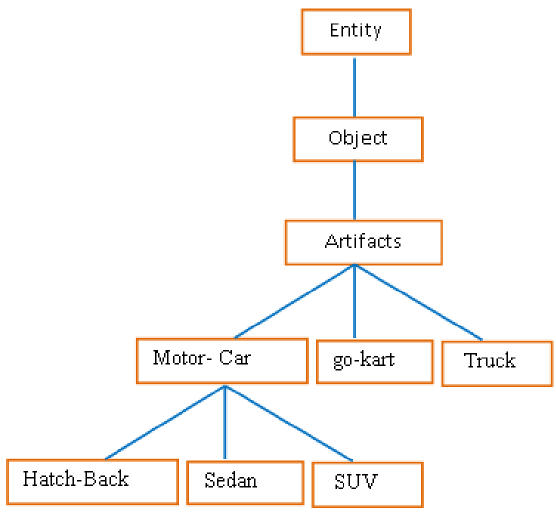

Figure 1. A fragment of the WordNet Is-A hierarchy.

2.1.2. SENSUS

Kevin Knight et al. [17] has constructed the ontology SENSUS which contains 90000 node concept thesaurus. SENSUS is an extension and reorganization of WordNet; at the top level, they have added nodes from the Penman Upper Model, and the major branches of WordNet been rearranged to fit. Each node is SENSUS represents one concept, i.e. each word has a unique specific sense and the concepts are linked in a straightforward IS-A hierarchy, becoming ever more general as you go upward toward the root of the ontology.

2.1.3. Cyc KB

Cyc is a large common-sense knowledge base, designed to serve as an encyclopedic repository of all human knowledge. Cyc consist of an extensive taxonomy of concepts, terms, and a large, self-reflective vocabulary for describing the most common definitional needs human have encountered over time. Cyc composed by the knowledge base of a general common-sense rule and assertion involving those terms. For example, the fundamental human knowledge that can be included in Cyc is the underlying set of facts, assumptions, and about the objects and events of everyday life. The ontology of Cyc grew to about 100,000 in 1994 and as of 2017 is about 2,450,000 terms, including over forty thousand types of relations, additionally to nine million assertions which relate these terms.

2.2. Domain-based Ontologies

The domain ontology (or domain-specific ontology) represents the concepts that belong to the specific field (or domain) such as education, medical, and so on. Different ontologies in the same domain could arise due to different languages, different intended uses of the ontologies, and different perceptions of the domain.

2.2.1. UMLS

The Unified Medical Language System (UMLS) is a very large, multi-purpose and multilingual Metathesaurus of biomedical controlled vocabularies developed by the US National Library of Medicine [2], intended to be used mainly by developers of systems in medical informatics. UMLS integrates over 2 million biomedical concepts and 9 million concepts name from more than 100 families of biomedical vocabularies, as well as over 12 million relations among these concepts. Each concept is assigned with at least one "Semantic type" in Metathesaurus and a certain semantic relationship may obtain between elements of various semantic types. The UMLS basically has three components

- Metathesaurus: It consists of terms and codes from various vocabularies integrated with UMLS such as MeSH, SNOMED, OMIM, GO, UWDA, NCBI, ICD-10-CM, and LOINC.
- Semantic Network: Broad categories (semantic types) and their relationships (semantic relations)
- SPECIALIST Lexicon and Lexical Tools: Natural language processing tools

2.2.2. MeSH

MeSH (Medical Subject Headings) is a comprehensive controlled vocabulary thesaurus created and updated by the United States National Library of Medicine [3]. It has organized as a set of terms naming descriptors in a hierarchical structure that allows searching at various levels of explicitness. For example, terms "Digestive sign and symptoms condition" is higher in taxonomy than most specific terms "Diarrhea". There are more than 28000 descriptors and over 90,000 entry terms in 2017 MeSH [3], that assist in finding the most appropriate MeSH Heading. There are more than 240,000 Supplementary Concept Records (SCRs) within a separate thesaurus in addition to these headings [3]. Figure 2 Shows the portion of the WordNet hierarchy.

2.2.3. STDS

STDS (Spatial Data Transfer Standard) is a standard, designed for transferring earth-referenced spatial data between disparate computer systems without any trouble. It is a transfer standard the describes the underlying conceptual model and detailed stipulations for the content, structure, spatial data, their associated attributes, features, data dictionary, and other supporting metadata all included in the transfer based on ontology. The commonly used concepts on topographic quadrangle maps and hydrographic charts are concepts in SDTS [34].

3. RELATED WORKS

Over the last few years, many researchers proposed various methods by proposing different ways of determining the semantic similarity between terms. Most of them were been tested on WordNet. According to the specific knowledge, information sources utilized and the way in which they used semantic similarity methods are classified.

Semantic similarity methods have mainly classified into four categories.

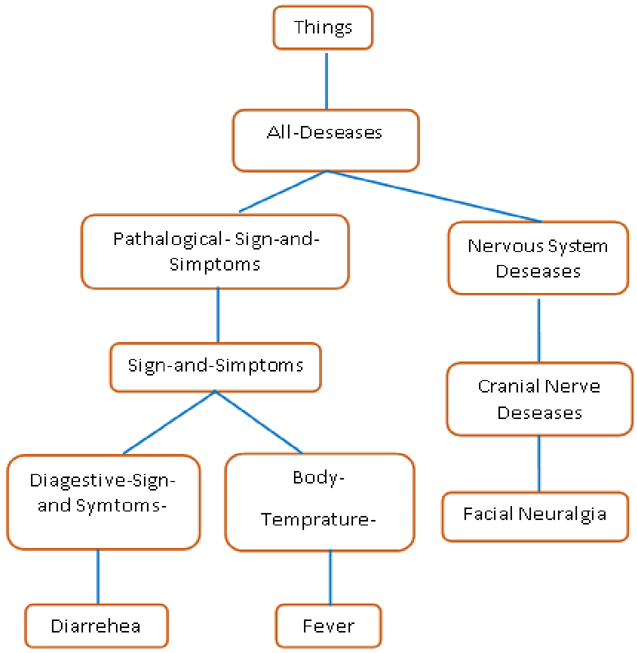

Figure 2. A fragment of the WordNet hypernym hierarchy

3.1. Edge Counting Methods

Edge counting measure was first introduced by Rada et al. [31], which applies to specific ontology with relations between two terms (concepts) of the taxonomic type (Is-A relationship) [18, 19, 30, 31, 36]. The main idea about these measurements is the fewer number of edges between two concepts, the more similar they are. In this case, the semantic similarity between two concepts C1 and C2 are given as:

$$dis(C1, C2) = \min(path(C1, C2)) \qquad (1)$$

Wu Z. Palmer [36] considered the depth of ontology in the measure, because the more specific two concepts are, the more similar they will be and vice versa. The conceptual similarity measure is given as:

$$ConSim(C1, C2) = \frac{2*N}{N1 + N2 + 2*N} \qquad (2)$$

Where N1 is the number of “Is-A relationship” edges between the concept C1 and C2 and least common subsumer (LCS) of (C1, C2), N2 is the number of “Is-A relations” edges between the concept C2 and the LCS of (C1, C2), and N is the number of edges “Is-A relations” between the LCS and ontology root.

If there is a need to calculate the conceptual similarity between “Fever” and “Diarrhea” in Figure 2, then the calculation will be doing as follows: firstly, determine the LCS of “Fever” and “Diarrhea” i.e. Signs-and-Symptoms. Next, determine that the length of the path from Fever to Signs-and-Symptoms is 2, which the length of the path from “Diarrhea” to Signs-and-Symptoms is 2, and the depth of the Signs-and-Symptoms is 3. It is now straightforward to determine that

$$ConSim(fever, Diarrhea) = \frac{2*3}{2+2+2*3} = 0.6$$

Several other measures were subsequently introduced by Li Y. Et al. [18], they attempted to make adjustments for particular aspects of Wu Z. Palmer [36] measures. This type of semantic measure is simple to implement but it is limited to ontology with taxonomic relations (Is-A relationship). Moreover, it does not allow for the context and can give possibly incorrect semantic similarity measures.

3.2. Information Contents Methods

Information contents method use the information content of concepts to measure the semantic similarity between two terms or concepts were first introduced by Rensik [22]. The information content value of a concept probability of occurring in a corpus such as WordNet [16, 20, 21, 28]: the higher the occurrence of the concept, the less the information content. The information content is given as:

$$IC(c) = -\log P(c) \qquad (3)$$

P(c) is the probability of encountering an instance of concept c in a large corpus.

Several other semantic similarity measures were subsequently proposed by Jiang et al. [11] which was inspired by Rensik [28].

Lin D. [21] proposed a measured based on an ontology which was restricted to hierarchic links and a corpus. This similarity measure takes the account of information shared by two concepts C1 and C2 like Rensik [28], but the difference between them is the definition. The definition holds the same factor as Rensik [28] but the combination is not a difference but a ration

$$Sim_{Rensik}(C1, C2) = \frac{(2*\ln((p_{mis}\ (C1, C2)))}{(\ln((c1)) + \ln(p(c2)))} \qquad (4)$$

Therefore, using this measure to compare the terms of ontology presents a better ranking of similarity than the Rensik [28] measure.

Jiang et al. [16] proposed a measure in a similar way as Rensik [28], they have used a corpus in addition to a hierarchic ontology (taxonomic relationship). The distance between two concepts C1 and C2, devised in this work is the difference between the sum of the information content of the two concepts and the information content of their most informative subsumer:

Among the limitations of these measures is their reliance on the corpus, as the concepts may be sometimes ambiguous or even not present. They also give the same result for any pair of concepts with the same LCS [32]. Their dependency on the design of the ontology and their lack of consideration for the context are also limitations.

3.3. Feature Based Methods

Measuring the similarity between two terms based on the features of the concepts in the taxonomy. A common characteristic of features tends to increase the similarity and a non-common characteristic tends to reduce the similarity of two concepts [35].

3.4. Hybrid Methods

Those methods consist of combining the ideas from the above three approaches in order to compute the semantic similarity between C1 and C2 [30]. Term similarity computed by matching synonyms, term neighborhoods, and term features. Term features further distinguished into parts, functions, and attributes matched to similarly [35].

4. WEB SEARCH ENGINE (WSE) BASED APPROACHES

The study of semantic similarity between words has become an essential part of many fields such as information retrieval, natural language processing and so on. The problem which we are addressing consists of trying to measure the semantic similarity between two given word w1 and w2. Similarity involves the measurement of inherent common characteristics between two or more concepts. Semantic similarity is a concept that extends beyond the synonymy and is often described as semantic relatedness in the literature [15].

Bollegala *et al.* [7] observed that; a certain degree of semantic similarity not only between synonyms (noon, midday) but also between meronyms (book, page) or hyponyms (rose, flower).

Explicit Semantic Analysis (ESA) [12] presented a unique approach for information retrieval studies and other related research. This method measures the semantic relatedness of two words in concepts space rather than a terms space, i.e. the relationship is not limited to the lexical form of text but expanded to include the meaning of words.

In this paper, we utilize the concepts space instead of terms space to compute the semantic similarity between words, i.e. comparing the meaning of terms instead of comparing their related lexicography. For example, the terms hat and rat are relatively similar to the lexicographical point of view but do not express the same meaning at all. We are only interested in the real-world concept that they represent, taking into consideration that a similarity score of 0 stands for complete inequality and 1 stands for complete equality of concepts being compared.

Over the years plenty of work have been carried out on measuring the semantic similarity using Web content. Approaches for semantic similarity measures using WSE based methods can be categorized as follows:

- Snippet based methods
- Page count based co-occurrence measure methods
- Frequent pattern finding based methods
- Trend Analysis based methods

4.1. Snippet based methods

These kinds of approaches consist of capturing the text snippets which are generated by the search engines like Google when producing the result, just after searching for these terms. These text snippets can be processed in order to compare distinguished algorithm for determining the semantic similarity between two terms based on their associated text snippets**.**

Sahami *et al*. [33], proposed a semantic similarity measure between two queries using snippets returned for those query by a WSE. For each query, snippets are collected from a WSE and represent each snippet as a (Term Frequency-Inverse Document Frequency) weighted term vector. In this method collected snippet captured more of the semantic context based similarity measures rather than the taxonomy based similarity measures. As a result, high TF provides more semantic similarity and IDF provides less semantic similarity. The major drawback of this method is that only the top ranking results for a query can be processed efficiently.

H. Chen *et al.* [13], proposed a double-checking method using text snippets returned by the WSE. Two objects are considered to be related if one can be found at from the other using web search engine. For example, the Co-Occurrence Double-Checking (CODC) is defined as:

$$CODC(w1, w2) = \begin{cases} 0, & w1@w2 = 0 \\ \exp(\log[\frac{f(w1@w2)}{H(w1)} \times \frac{f(w1@w2)}{H(w2)}])^{x}, & otherwise \end{cases} \quad (5)$$

The major drawback of this method is that we cannot assure the occurrence of one word in the snippets for the other event even though they are related.

4.2. Page count based co-occurrence measure methods

It consists of measuring the probability of co-occurrence of the terms on the Web based on the page count. For the given two words w1 and w2 page counts are returned by WSE when these words are given as an input. On the Web, the probability of term co-occurrence can be identified by hits. In fact, these formulas are measures for the probability of co-occurrence of the terms w1 and w2. [10]. The probability of a specific term is specified by the number of hits returned when a given WSE is presented with this search term divided by the overall number of web pages possible returned. The combined probability p (w1, w2) is the number of hits returned by a WSE, including both search term w1 and w1 divided by the overall number of web pages returned.

Normalized Google Distance (NGD) [10] is considered one of the most outstanding works in this field. The NGD is a measure of semantic similarity derived from the Google search engine (GSE) for a given set of keywords.

$$D(w1, w2) = \frac{\max\{\log hit(w1), \log hit(w2)\} - \log hit(w1, w2)}{\log M - \min\{\log hit(w1), \log hit(w2)\}} \quad (6)$$

There are four famous other measures of this kind are Jaccard, Overlap (Simpson), Dice, and Pointwise mutual information (PMI), all of which are described by Bollegala et al. [7]. When these measures are used it is necessary to include the prefix Web e.g. WebJaccard, WebDice and so on. All of these measures are used to compute the probability between terms using page count. These are their corresponding formulas:

$$WebJaccard(w1, w2) = \log \frac{p(w1, w2)}{p(w1) + p(w2) - p(w1, w2)} \quad (7)$$

$$WebOverlap(w1, w2) = \log \frac{p(w1, w2)}{\min(p(w1, w2))} \quad (8)$$

$$WebDice(w1, w2) = \log \frac{2.p(w1, w2)}{p(w1) + p(w2)} \quad (9)$$

$$WebPMI(w1, w2) = \log \frac{p(w1, w2)}{p(w1).p(w2)} \quad (10)$$

The main idea behind these measures is that terms with similar meanings tend to be close to each other because it appears to be empirically supported that synonyms often appear together in web pages [10], whereas terms with dissimilar meanings tend to be farther apart, and therefore, present low similarity values.

4.. Frequent pattern finding based methods

The techniques of this group belong to the field of machine learning that consists of looking for similarity patterns in the websites that are indexed by particular WSE. One of the famous technique was proposed by Bollegala et al. [7], which consists of looking for such regular expressions as “w1 also known as w2”, “w1 is w1 w2”, “w1 is an example of w2”, and so on. This is because this kind of expression indicates the semantic similarity between the two (set of) terms.

A high number of occurrences of these types of patterns give us with evidence for the similarity between the two terms, but it is essential to perform some preliminary studies about what is ‘a high number’ according to the problem that we wish to address. This can be done, for example, by studying the number of results offered by particular WSE such as Google for perfect synonyms. Furthermore, it is necessary to take into account that these expressions should be tested in two ways because the similarity between w1 and w2 is by definition equal to the similarity between w2 and w1.

4.4. Trend Analysis based methods

Trends analysis based methods used for extracting an underlying pattern of behavior in time series, i.e. collections of observations of well-defined data items obtained through repeated measurements. WSE stores the queries in this way in order to exploit this information in an efficient manner in the future. Over the years many methods have been proposed to measure the correlation between search patterns and utilize for the computing semantic similarity [15]. Jorge Martínez Gil [15] has proposed method using Pearson's correlation coefficient which is closely related to the Euclidean distance over normalize vector space. This measure provides the mean in shape of time series instead of quantitative values. Therefore, the similar concepts may present almost exactly the same shape in their associated series and semantic similarity between them is presumed to be very high. This coefficient can be computed as follows:

$$\rho_{w1,w2} = \frac{\mathrm{cov}(w1,w2)}{\sigma_{w1}\sigma_{w}{}^{2}} = \frac{E[(w1-\gamma a)(\mathrm{w2}-\gamma \mathrm{w2}]}{\sigma_{w1}\sigma_{w2}} \quad (11)$$

This measure of correlation uses the Spearman correlation coefficient which is used to assess how well the association between two variables can be described using the monotonic function. If there are no repeated values, a perfect Spearman correlation occurs when each of the variables is a perfect monotone function for other. This is the formula used to compute it:

$$\rho_{w1,w2} = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \quad (12)$$

5. STATISTICAL EVALUATION

Human has an innate ability to judge the semantic similarity of terms. Therefore, to evaluate the semantic similarity we have created a new dataset that has been judged by the 30 people who came from the several countries and various fields, indicating a value 0 for dissimilar terms and 1 for completely similar terms. Our main aim in designing this new dataset is to evaluate terms that are not commonly included in dictionaries but frequently used by people today. To do this, we will be able to determine the most appropriate method for comparing the semantic similarity of promising terms. This could be greatly utilized in various dynamic domains such as education, medical, finance, business, marketing, social networks, emerging technologies, and so on. Table 1 demonstrates the term pairs and the mean for the values obtained after requesting the people to make the judgment on their similarity.

To do the comparison between this dataset and produce our results is made using Pearson’s Correlation Coefficient, which is a statistical measure for the comparison of two matrices of numeric values. Consequently, the results can be in the interval (-1, 1), where 1 represents’ the best case and -1 represents the worst case. This correlation coefficient permits us to measure the strength of the relation between human ratings of similarity and computational values. Pirro G. [25] stated that it is also essential to estimate the significance of this relation. To do that, we have used the p-value technique, which shows how unlikely a given correlation coefficient will occur given no relation in the population. We have obtained that, for our sample, all values above 0.3 are statistically significant. A larger dataset would be necessary to confirm the significance of the rest of the tests. However, Pirro stated that it is also necessary to evaluate the significance of this relation [25]. To do that, we have used the p-value technique, which shows how unlikely a given correlation coefficient will occur given no relation in the population. We have obtained that, for our illustration, all values above 0.3 are statistically significant. A larger dataset would be required to confirm the significance of the rest of the tests.

To make the comparison among methods with the existing ones; we have considered techniques which are based on dictionaries. We have selected the Path Length algorithm which is simple node counting approach. The similarity score of terms is inversely proportional to the number of nodes along the shortest path between the definitions. The shortest path takes place when the two definitions are the similar [26]. Lesk et al. [23] have proposed a dictionary-based approach which consists of finding overlaps in the definitions of the two terms. The relatedness score is the sum of the squares of the overlap lengths. Rensik [28] has proposed an information-based approach which is used to computes the common information between concepts. Finally, Pedersen et. al. [26] has proposed the vector pair method in which the basic idea is to compare the co-occurrence vectors from the WordNet definitions of concepts.

Table 1. Benchmark of datasets holding the similarity scores / human (mean) for a set of terms and expressions that not often covered by the dictionaries

Term / Expression w1	Term / Expression w2	Score Human Judgment(Mean)
Delay	racism	0.119
Production	hike	0.175
Volunteer	motto	0.256
Prejudice	recognition	0.300
tweet	snippet	0.315
Btw	by the way	0.451
investigation	effort	0.459
Slumdog	underprivileged	0.492
Secretary	senate	0.506
PDA	computer	0.565
Life	lesson	0.594
Weapon	secret	0.606
Skin	eye	0.622
TBA	to be announced	0.625
Governor	office	0.634
FYI	for your information	0.635
Qwerty	keyboard	0.665
FAQ	frequently asked questions	0.71
Thx	thanks	0.785
Treatment	recovery	0.791
Credit	card	0.806
War	troops	0.813
Nature	environment	0.831
Weather	forecast	0.834
Seafood	lobster	0.870
Liquid	water	0.885
Wi-Fi	wireless network	0.915

Table 2. Ranking for the algorithm tested using dataset (Highlights shows the name of the algorithm considered in this study)

Ranking	Algorithm	Score
1	WebOverlap	0.535
2	Patterns	0.528
3	Co-Occurrence	0.520
4	WebDice	0.423
5	WebJaccard	0.401
6	WebPMI	0.391
7	ESA	0.383
8	Google Normalized Distance	0.315
9	Snippet Comparison	0.283
10	Vector Pairs	0.215
11	Pearson Coefficient	0.121
12	Lesk	0.92
13	Path Length	0.71
14	Prediction	0.031
15	Outlier Comparison	0.009
16	Leacock	0.006
17	Spearman	0.001
18	Rensik	-0.15

Table 2 shows the results of applying various methods to estimate the semantic similarity for the benchmark dataset. As can be seen, the emerging methods are giving the much better result as compare to the dictionaries-based methods. The big reason is that by using WSE it is possible to have fresh and up to date contents. On the other hand, we can see that the best methods are those which are based on WSE such as co-occurrence, frequent pattern finding and trend analysis. Snippet based method seems to less efficient, but these results may be influenced by the fact that our employed method is simple. More complex methods which are based on this idea could be better, at least when solving specific circumstances. Finally, we have realized that the classical methods (Vector pairs, Lesk, Path Length and Resnik) which are based on the dictionaries are much worse than the majority of the emerging ones; consequently, our initial assumption is confirmed. Furthermore, it is essential to take into account that most of the methods explained here are appropriate for optimization, even though this step is beyond the scope of this work.

6. CONCLUSION AND FUTURE RESEARCH DIRECTIONS

The semantic similarity measure is playing an important role in many applications nowadays. In this survey, firstly we have reviewed various ontologies used for semantic similarity and traditional methodology used for semantic similarity measure. Secondly, we have described and evaluated various promising novel techniques for determining the semantic similarity between words which consist of using knowledge from WSE. All the techniques reviewed have been evaluated using a benchmark dataset for terms which are not often included in dictionaries, taxonomies or thesaurus. As a result, we have confirmed experimentally that some of the WSE based methods significantly outperform existing methods when evaluating this kind of dataset.

For future work, the effectiveness of semantic similarity strongly depends on the richness of the fact that people rate our term pairs in many different ways according to their cultural background. Therefore, in future, we want to avoid the cognitive bias as for the particular terms that are a synonym to some person, but a person from another culture may not agree (and vice versa). It is necessary we should have a common agreement on the data used for evaluating different approaches. One direction of future trends to use the automatic construction of ontologies (KBs) will possibly lead to improving the accuracy of the semantic similarity measure. The latter is related to agent communication, where agents need to share knowledge. The main objective is to determine which approaches are the best for resolving this problem and implement them in real information systems where the automatic computation of semantic similarity between terms may be necessary.

Acknowledgments: We are thankful to the Basic Science Research Unit, Deanship of Scientific research, Majmaah University, Kingdom of Saudi Arabia for providing all the research facilities.

REFERENCES

1. http://mcmcweb.er.usgs.gov/sdts.
2. http://umlsks.nlm.nih.gov
3. https://www.nlm.nih.gov/
4. https://www.nlm.nih.gov/mesh/meshhome.html
5. "WordNet Statistics", WordNet.Princeton.edu, Retrieved 2018-06-22.
6. Budanitsky, A. and Hirst, G., "Evaluating WordNet-based measures of lexical semantic relatedness," *Compute Linguistics*, **32**(1):13–47(2006), https://doi.org/10.1162/coli.2006.32.1.13
7. Bollegala, D. and Matsuo, Y., Ishizuka M., "Measuring semantic similarity between words using web search-egines," *In Proceedings of WWW*, pp:757–766 (2007) doi>10.1145/1242572.1242675
8. Beckwith, R. and Miller, G. A., "Implementing a Lexical Network", *Technical Report 43, Princeton University*, (1992)
9. Barbara, R., Katarzyna P., Krystyna, C., Wojciech, W., and Piotr, K., "SemEval-2016 Task 1: Necessity for diversity; combining recursive auto encoders, wordnet and ensemble methods to measure semantic similarity," In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval 2016). San Diego, CA, USA (2016) DOI: 10.18653/v1/S16-1091
10. Cilibrasi, R. and Vitányi, PM., "The google similarity distance," *IEEE Trans Knowl Data Eng*, **19**(3):370–383(2007) **DOI:** 10.1109/TKDE.2007.48
11. G. Miller, "WordNet: a lexical database," *Communications of the ACM*, **38**(11), pp. 39-41 (1995) doi>10.1145/219717.219748
12. Gabrilovich, E. and Markovitch, S., "Computing semantic relatedness using Wikipedia-based explicit semantic analysis," *In Proceedings of the 20th international joint conference on Artificial intelligence, IJCAI'07. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA*, pp 1606–1611 (2007) https://dl.acm.org/citation.cfm?id=1625535
13. Chen H., Lin, M., and Wei, Y., "Novel Association Measures using Web Search Engines with double checking," *Proc. 21st Int'l Conf. Computational Linguistics and 44th Ann. Meeting of the Assoc. For Computational Linguistics (COLING/ACL '06)*, pp. 1009-1016 (2006) doi>10.3115/1220175.1220302
14. Junfeng, T. and Man, L.," Leveraging word embedding from macro and micro views to boost performance for semantic textual similarity" *ECNU at SemEval2016 Task 1: In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval 2016), San Diego, CA, USA*, (2016) http://www.aclweb.org/anthology/S16-1094
15. Jorge MG., "an overview of textual semantic similarity measures based on web intelligence," *Artif. Intell. Rev.*, **42**(4), 935-943 (2014) DOI: 10.31219/osf.io/ka26r
16. Jiang, J., Conrath, D., "Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy," *In: Intern. Conf. on Research in Computational Linguistics, Taiwan*, (1998) http://www.aclweb.org/anthology/O97-1002
17. Knight, K. and Luk, S., "Building a Large-Scale Knowledge Base for Machine Translation," *In Proceedings of thr National Conference on Arti¯cial Intelligence (AAAI'94)*, Seattle, WA,(1994) https://dl.acm.org/citation.cfm?id=199456
18. Li, Y., Bandar, Z.A. and McLean, D., "An Approach for Measuring Semantic Similarity between Words Using Multiple Information Sources," *IEEE Trans. on Knowledge and Data Engineering*, **15**(4), 871–882(2003) **DOI:** 10.1109/TKDE.2003.1209005
19. Leacock, C., Chodorow, M., "Combining Local Context and WordNet Similarity for Word Sense Identification in WordNet," *In Fellbaum, C., ed.: An Electronic Lexical Database, MIT Press*, 265–283(1998) https://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6287675
20. Lord, P., Stevens, R., Brass, A., and Goble, C., "Investigating Semantic Similarity Measures across the Gene Ontology: The Relationship between Sequence and Annotation, Bioinformatics", **19**(10), 1275–83 (2003) https://doi.org/10.1093/bioinformatics/btg153
21. Lin, D., "Principle-Based Parsing Without Over generation," *In: Annual Meeting of the Association for Computational Linguistics (ACL'93), Columbus, Ohio*, 112–120 (1993), doi>10.3115/981574.981590
22. Nelson, S.J., Powell, T. and Humphreys, B.L., "The Unified Medical Language System (UMLS) Project," In A. Kent and C.M. Hall, editors, *Encyclopedia of library and Information Science*, Marcel Dekker, Inc., New York, 369-378 (2004), DOI: 10.1093/nar/gkh061
23. Lesk, M., "Information in Data: Using the Oxford English Dictionary on a Computer," *SIGIR Forum*, **20**(1-4): 18-21 (1986), doi>10.1145/15497.15502
24. Nelson, S.J. Johnston, D. and Humphreys, B.L., "Relationships in Medical Subject Headings," In C.A. Bean and R. Green, editors, *Relationships in the Organization of Knowledge*, pages 171(184), Kluwer Academic Publishers, New York (2001), https://www.springer.com/gp/book/9780792368137
25. Pirro G., "a semantic similarity metric combining features and intrinsic information content," *Data Knowledge Engineering*, **68**(11):1289–1308 (2009) doi>10.1016/j.datak.2009.06.008
26. Pedersen, T., Patwardhan, S., and Michelizzi, J., "WordNet: Similarity - Measuring the Relatedness of Concepts, "*AAAI*,1024-1025(2004) https://dl.acm.org/citation.cfm?id=1597310
27. Reed, S. and Lenat, D., "Mapping Ontologies into Cyc.," In *Proceedings of the AAAI'02 Conference Workshop on Ontologies for The Semantic Web*, Edmonton, Canada, July (2004), doi>10.1145/219717.219745
28. Resnik, O., "Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity and Natural Language," *Journal*

of Artificial Intelligence Research, 11, 95–130 (1999), https://dl.acm.org/citation.cfm?id=3013547

29. Rodriguez, M., and Egenhofer, M., ”Determining Semantic Similarity Among Entity Classes from Different Ontologies,” *IEEE Trans. on Knowledge and Data Engineering,* **15**(2), 442–456 (2003), **DOI:** 10.1109/TKDE.2003.1185844
30. Richardson, R., Smeaton, A., and Murphy, J., “Using WordNet as a Knowledge Base for Measuring Semantic Similarity between Words,” *Techn. Report Working paper CA-1294*, School of Computer Applications, Dublin City University, Dublin, Ireland (1994), http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.49.6027
31. Rada, R., Mili, H., Bicknell, and E., Blettner, M., “Development and Application of a Metric on Semantic Nets,” *IEEE Trans. on Systems, Man, and Cybernetics*, **19**(1), 17–30 (1989), **DOI:** 10.1109/21.24528
32. Sánchez, D., Batet, M., Isern, D., and Valls, A., “Ontology-based semantic similarity: A new feature-based approach,” *Expert Systems with Applications*, **39** (9),7718-7728(2012a), https://doi.org/10.1016/j.eswa.2012.01.082
33. Sahami M. and Heilman T., “A Web-based Kernel Function for Measuring the Similarity of Short Text Snippets,” *15th International Conference on World Wide Web*, pp. 377-386 (2006), doi>10.1145/1135777.1135834
34. Thabet Silamani, ”Description and Evaluation of Semantic Similarity Approaches,” *International Journal of Computer Applications*, **80**(10):25-33 (2013), DOI:10.5120/13897-1851
35. Tversky, A., “Features of Similarity, Psycological Review, ”**84**(4),327–352(1977), http://www.cogsci.ucsd.edu/~coulson/203/tversky-features.pdf
36. Wu, Z., and Palmer, M., “Verb Semantics and Lexical Selection,” In *Annual Meeting of the Associations for Computational Linguistics (ACL’94)*, Las Cruces, New Mexico, 133–138 (1994), doi>10.3115/981732.981751