

## Identification of New Parameters for Ontology Based Semantic Similarity Measures

Shivani Jain<sup>1,\*</sup>, Seeja K.R<sup>1</sup> and Rajni Jindal<sup>2</sup>

<sup>1</sup>Department of Computer Science & Engineering, Indira Gandhi Delhi Technical University for women (IGDTUW), Delhi, India

<sup>2</sup>Department of Computer Science & Engineering, Delhi Technological University (DTU), Delhi, India

### Abstract

A major challenge among various applications of computational linguistics, natural language processing and information retrieval is to measure semantic similarity accurately. In this research paper, various ontology-based approaches i.e. compute semantic similarity between words have been studied and listed their benefits and shortcomings on the various identified parameters. Earlier, correlation with human judgment was the single criteria for the judgment of good similarity measures. In this paper, more parameters for semantic similarity measures have been identified and a relative analysis of similarity measures is performed on the identified parameters. These identified parameters can be further utilized for formulating the new semantic similarity-measures in the latest research area of text mining, web mining and information retrieval. We have identified various parameters like features-set, applicability on various ontologies such as single ontology, cross ontology or fuzzy ontology, Ontology type, dataset applied and relationship mapping for the various measures. Through detailed analysis we have found that feature based and hybrid approaches has higher accuracy as compare to edge and content based methods and works in different type of ontologies. Recent research drawing interest to find new feature set in this area like fuzzy distance, graph generation and text snippets etc. Max Accuracy was achieved in single ontology 0.87 and 0.83 over cross ontologies. WordNet and MeSH are maximally utilized Global Ontologies.

**Keywords:** Ontology, Natural-Language-Processing, Text-Mining, Knowledge-Based, Semantic-Similarity, WordNet.

Received on 16 October 2018, accepted on 19 January 2019, published on 30 January 2019

Copyright © 2019 Shivani Jain *et al.*, licensed to EAI. This is an open access article distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/3.0/>), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/eai.19-12-2018.156439

\*Corresponding author. Email:shivanijain13@gmail.com

### 1. Introduction

Huge success of WWW and knowledge society quantity of textual electronic information has been expanded significantly[1] that has acquired the interest of many researcher groups in this area. Most of the information on the WWW is represented by structural form like HTML, DHTML and XML formats. This information interpreted by the humans not by the machines. Now a day's web generates

approximately 25 billion textual data on each day. Such a huge data, big data[2,3,4] cannot be organized by humans, so latest research is drawing interest how this textual information can be understood and interpreted by machines. Understanding and Interpretation of textual information revolves around two basic terms i.e. Text Mining [5] and Natural Language Processing(NLP) [6,7]. Textual information is broken into smaller chunks and the smallest chunk is known as terms/words. A major concern in NLP and computational linguistics is to interpret the meaning of each single word. For understanding the meaning of a

word/concept first step is to calculate the semantic similarity among the words or terms [8]. Semantic similarity measures evaluate a numerical value which quantifies the closeness among terms/concepts. Semantic closeness stated that how analytically near, two ideas (words or terms) are, on the ground they share some common part of their meaning [9,10]. Aim of semantic similarity is to measure the accurate closeness among the different words. Author Resnik et al. [11] defined the similarity between the words (“*Car*” & “*Automobile*”) as 1. For measuring the similarity several researchers proposed their similarity measures.

Although massive research have been done in this area, however in this research paper we have conducted a research study on ontology based similarity measures presented in the area of Knowledge Based(KB)[12,13]. KB system stores huge information in un-structural formats. Un-structural information can be processed and presented through an object model known as ontology [14,15]. Ontologies are created for different aspects and domains; they regularly contains overlapping data and information that can be further used for calculating the similarity among the different words [16]. In this paper, a relative study is presented on various parameters in the area of ontology based measures. Similarity is not a new term; semantic similarity was coined in 1995 by the Resnik et al. After that every year new similarity measures were proposed.

First motivation for this study is none of the research papers presented such an extension research on the ontology based measures; they merely illustrated a small amount of measures in this area, however in this paper authors have studied more than 45 research papers in this area. A detailed comparative study is shown.

Second motivation is to identify different parameters; by these approaches can be compared. In the prior studies accuracy was the only criteria for the formulation and judgment of a good similarity measures however, every year; new similarity measures are proposed to deal with the latest research in this area[17,18]. To cope up with these new emerging fields of knowledge base authors have identified more parameters for the formulation of a good similarity measure. By the extension study in this field the authors have identified different parameter such as: features set, type of ontology framework, benchmark dataset used, relationship consideration and Ontology type. Detailed explanation is presented in the forthcoming sections.

This paper is divided into different sections; First section describes the methodology for conducting the research study, in the same section briefly explains similarity taxonomy. Third section shows the various similarity measures in the area of KB. Fourth section characterized the various parameters identified and their respective meaning. Next section presented the detailed relative study of various approaches on identified parameters. Next section illustrates the analysis and finding of the research paper. Last section describes conclusion and future work in this area.

## 2. Research methodology and brief of semantic similarity taxonomy

### 2.1. Research Methodology

For conducting this research study, we have searched various research papers from the online repository like IEEE, Elsevier, Springer, IGI Global, InderScience journal and conference papers from year 1996-2018. Identified keywords for extraction of research papers are; “*Word-Similarity Concept*”, “*Semantic Similarity Meaning*”, “*Corpus-based Similarity*”, “*Text-based Similarity*”, “*Knowledge based similarity*”, “*Ontology based similarity*”, “*Edge-based similarity*” and “*feature based similarity*”. If above defined keywords found in the title and in abstract of the research paper we manually considered them for our research paper. After studying these papers we have identified more related terms that illustrated more depth information in this area. The related keywords are “*WordNet-based similarity measures*”, “*Bio-medical based similarity measures*”, “*Cross-Ontology based Similarity measures*”, “*Fuzzy-ontology based similarity measures*” and “*Formal concept analysis based similarity measures*”. In this research paper we work on the word similarity measures in the area of knowledge base thus we have not included the “string based similarity”, “corpus based similarity”, “Sentence based similarity”, “Paragraph based similarity” and “Document level similarity” for the research paper.

### 2.2. Semantic Similarity Taxonomy

Similarity can be word similarity, sentence similarity or paragraph similarity. In this paper, word similarity is discussed. Two words can be similar lexically or semantically. Lexical similarity represents words having similar string sequence like in {“*Man*”, “*Lan*” and “*Van*”} and also in DNA sequence matching {“*ADCGTDCGTC*” and “*ADCCGTCGCA*”}. Lexical similarity has a wide application in the areas of medical sequence matching [19, 20] and pattern recognition. Various measures are proposed in this area [20].

Whereas semantic similarity deals with the meaning of the two words, how the linguistic meanings of two words are similar, such as how much similar the words {“*mango*” and “*orange*”} as both words belong to class “*Fruit*”. Another term, associated with semantic similarity is “semantic relatedness” [22] not necessarily relay’s only on the taxonomic relationship “*is-a*”, however more relationships were explored in semantic relatedness. For example {“*tier*”, “*pencil*”} was less associated to each other as compared to {“*pencil*”, “*paper*”} in terms of semantic relatedness and “*pencil*” and “*paper*” don’t has “*is-a*” relationship. In this other relationships like “*part-of*”, “*antonym*” is explored.

For calculating the semantic similarity among the words diverse similarity measures have been proposed by different researcher’s groups. In literature two approaches “*Corpus-Based Approaches*” and “*Knowledge-Based Approaches*”

are presented by the researchers [12,23]. In corpus based approaches large amount of textual information is used for calculating the similarity among the words. Latent Semantic Approach (LSA) [24] and Pointwise mutual information(PMI) [25] are the widely used methods for computing the similarity. Fig 1 shows the taxonomy of similarity measures.

Knowledge-Based Methods are an effort to figure out the semantic similarity without human intervention, it uses a vast amount of environment knowledge about the concepts. The major related work on semantic similarity measures utilizing the taxonomy knowledge by the Global ontologies like WordNet[26], MeSH[22] a medical ontologies. Ontologies have been extensively used in the area of knowledge-Base systems.

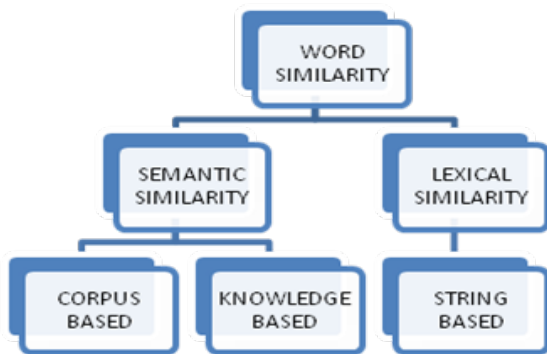


Figure 1. Taxonomy of similarity approaches

Ontologies are represented using classes, attributes and relationships among the attributes. Ontologies characterized as “a formal specification of a shared conceptualization” [27]. WordNet is a domain free and all-purpose thesaurus for English words. It organizes more than ten thousand English words that are semantically structured, and forms ontology. It formulates approximately 10000 English words in a semantic structure that looks like ontology. A graph prototype is generated in form of network and it is used for forming the meaning of concepts. In most of the knowledge based approaches; researchers used WordNet as the base ontology for the computing similarity among the word /concepts. In the next section we have explained the various measures present in WordNet ontology.

### 3. Knowledge based/ Ontology- based approaches

There are mainly three approaches present in literature for KB method [8, 28, 29] and last one is hybrid that is combination of any two approaches. The approaches are,

1. Edge/ path based Approaches
2. Information Content Approaches
3. Features Based Approaches
4. Hybrid Approach

In this section we briefly explain these approaches

Terms associated in semantic similarity measures are:-

Len (C<sub>i</sub>, C<sub>j</sub>) - shortest path length starting from synsets C<sub>i</sub> to another synset C<sub>j</sub> in WordNet.

LCS(C<sub>i</sub>, C<sub>j</sub>)- Lowest Common Subsumer in WordNet (LCS/lso) of C<sub>i</sub>, and C<sub>j</sub>

depth(C<sub>j</sub>)- path length to a synset C<sub>i</sub> from overall root unit and depth of root =1

Max\_deep - taxonomy maximum depth.

Max\_node - maximum number of words present in the ontology.

Sim(C<sub>i</sub>, C<sub>j</sub>) - semantic similarity among the two concept C<sub>i</sub> and C<sub>j</sub>

### 3.1. Edge/ Path based Approach

Edge-Counting approaches were introduced by Author Lin et al [30]. It was a very basic technique to measure the similarity, it calculates the minimum path-distance from the end to end connections in their related ontological model by a “is-a” links. As path length increased the similarity between the terms tends to decreased. If two terms are closed their distance must be low and similarity was high.

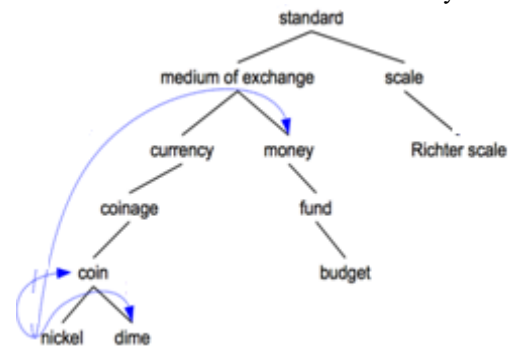


Figure 2. Edge Based Approach presented by Lin [26]

Here the similarity between words (nickel, dime) & (nickel, money) was computed according to edge based approach proposed by Lin as:

$$Sim_{(nickel|dime)} = \frac{1}{1+1} = 0.5$$

$$Sim_{(nickel|money)} = \frac{1}{1+5} = 0.162$$

Mostly used measures in edge based approaches:

**3.1.1** Shortest Path based measures introduced by Researcher Rada et al. [31]. This measure taken len(C<sub>1</sub>, C<sub>2</sub>) into consideration C<sub>1</sub> and C<sub>2</sub> are the two concepts and similarity was computed as:

$$Sim_{path}(C_1, C_2) = 2 * \max\_deep(\text{len}(C_1, C_2)) \quad (1)$$

**3.1.2** Sussna’s Measures [32] : This measure was proposed by Sussna et al. in 1993. It is based on the relative depth in the WordNet Ontology. WordNet Ontology has maximum 16 node depth strength. It was based on the

assumptions that sibling of the concepts in the lower nodes have more similarity than the higher pairs of nodes. In this noun network of WordNet was used, on this; they defined the two relationships among the edges one was direct relationship and another was inverse relationship. Both the relationships lies between the two ranges {min and max} min=1 and max=2<sup>4</sup>.

The weight for edge for a node A is calculated as:

$$wt(A - r) = \max_r - \frac{\max_r - \min_r}{edges_r(A)} \quad (2)$$

The distance between the two adjuncts Node (A, B) was calculated using the weight and depth of node as:

$$dist(A, B) = \frac{wt(A-r) + wt(A \sim r)}{2 \times \max\{depth(A), depth(B)\}} \quad (3)$$

**3.1.3 Wu & palmer's Measure [33]:** In this measure they work on the position for the concept C<sub>1</sub> and C<sub>2</sub> in the taxonomy. For finding the position the relative location was considered for common concept Iso (C<sub>1</sub>, C<sub>2</sub>) which is based on LCS . They concluded similarity among the concepts was based on path length and depth of the ontology. Path length and depth was defined through the links in the ontology.

$$Sim_{w\&p}(C_1, C_2) = \frac{2 * depth(Iso(c_1, c_2))}{len(c_1, c_2) + 2 * depth(Iso(c_1, c_2))} \quad (4)$$

**3.1.4 Leacock & Chodorow Measure [34]:** Maximum depth of the taxonomy was considered and defined the subsequent measure. They considered that if C<sub>1</sub> and C<sub>2</sub> have the same sense or the meaning, then their similarity was defined to be 1.

$$Sim_{l\&c}(C_1, C_2) = -\log \frac{len(c_1, c_2)}{2 * deep_{max}} \quad (5)$$

**3.1.5 Li's Measures [35]:**It works on a hypothesis that sources of information are infinite, and humans tend to evaluate sources of word similarity by means of a limited space between completely similar and non-similar words.

$$Sim_{Li}(C_1, C_2) = e^{-a * len(c_1, c_2)} \frac{e^{b * depth(Iso(c_1, c_2))} + e^{b * depth(Iso(C_1, C_2))}}{e^{b * depth(Iso(c_1, c_2))} - e^{-b * depth(Iso(C_1, C_2))}} \quad (6)$$

This equation symbolize similarity is monotonically increasing with respect to the depth. The value for a and b can be manually adjusted for better results. Here the length, depth and Iso parameters were considered.

### 3.2. Information Content Approach

It was based on the theory that every concept is a combination of a large extent of information in WordNet. It relay on a fact that a node is a distinctive representation of a concept in a particular domain and hold some about of information. Direct link between the two concepts is represented by an edge. Similarity among these two concepts was computed on the shared information in that particular domain. The concepts are more similar if they share related amount of information in a domain.

$$IC(w) = -\log p(w) \quad (7)$$

Here w presents a class/concept.

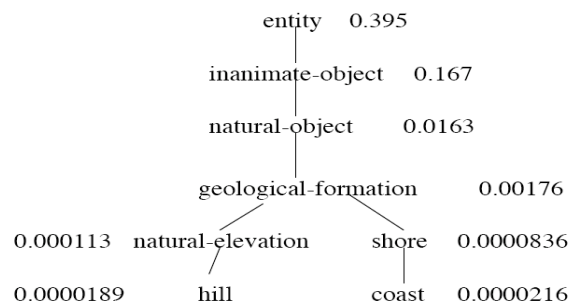
Here, w represent the concept and p(w) probability of that instance.

As the number of nodes increases, the information it contains was found decreases. If there exist a single root node in the ontology, then it is having the highest probability of occurrence, value can defined to 1 and information it contain is lowest and measured to 0.

Most used measures in Information content approaches:

**3.2.1 Resnik Approach[11]:** It relied upon shared information between two concepts. Two terms are dissimilar the most, if LCS doesn't exist. If LCS exists, using information content of the LCS, similarity was calculated as,

$$Sim_{res}(C_1, C_2) = IC(LCS(C_1, C_2)) \quad (8)$$



**Figure 3.** An Information Content Approach presented by Resnik [9]

**3.2.2 Lin Approach [36]:** It was proposed that the similarity among the concepts is measured as the ratio between the extents of information desirable to specify their association.

$$Sim_{lin}(C_1, C_2) = \frac{2 * Sim_{res}(C_1, C_2)}{IC(C_1) + IC(C_2)} \quad (9)$$

**3.2.3 Jiang and Conrath [37]:** It depends on measuring taxonomical length of connections among the IC of a specific word and it's LCS. This formula defines the dissimilarity concept between the two terms.

$$dis\_Sim_{jia\&con}(C_1, C_2) = IC(C_1) + IC(C_2) - 2 * Sim_{res}(C_1, C_2) \quad (10)$$

According to author Resnik et al for each incidence of a particular noun among the corpus is measure as an incident of every taxonomical class and can be formulated as,

$$P(a) = \frac{\sum_{w \in W(a)} count(w)}{N} \quad (11)$$

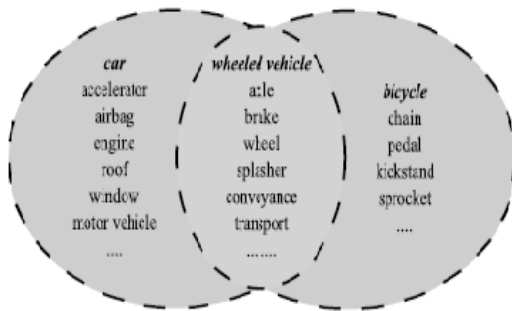
Here W (a) represent as nouns set in a corpus and multiple senses of the word w are denoted by "a". N shows overall number of nouns present in a defined corpus.

### 3.3. Feature Based Approach

Feature based methodology describes similarity among the concepts as a factor of their properties. They relied upon the amount of common and un-common features of evaluated

concepts. Common features enhanced the similarity and exceptional features tend to weaken it.

**3.3.1 Tversky's Model [38]:** It was established on the fact that each concept was depicted through a set of keywords presenting their properties and features. It was represented in the WordNet using definitions and glosses values. Glossary of the word "automobile" describe through the WordNet as "a motor vehicle with four wheels; usually propelled by an internal combustion engine".



**Figure 4** Tversky Model for two words 'car' and 'bicycle'[34]

Two words/concepts are more similar if they contain more common characteristics of words. Exceptional characteristics tend to decrease the similarity.

$$Sim_{Tversky}(C_1, C_2) = \frac{|c_1 \cap c_2|}{|c_1 \cap c_2| + k(|c_1 \setminus c_2|) + (k-1)\{|c_2 \setminus c_1|\}} \quad (12)$$

K is adjustable factor and k can be  $\in [0,1]$

**3.3.2 Rodriguez and Egenhofer [39]:** It proposed a similarity measure, where computation is done on the basis of the individual sum of likeness between the "synsets", their "features" and "neighborhood concept" of the estimated terms as:

$$Sim_{Res}(C_1, C_2) = w.S_{synsets}(C_1, C_2) + u.S_{feature}(C_1, C_2) + v.S_{neighborhood}(C_1, C_2) \quad (13)$$

$S_{synsets}$ ,  $S_{feature}$  and  $S_{neighborhood}$  are the similarity among the "synonym set", their "features" and corresponding "semantic neighborhood" of the computed terms. According to this ( $w, u, v \geq 0$ ) value of w, u and v depends on the particular similarity weights of every specific component. The value relies upon the individuality of the taxonomy. Here S symbolizes the coinciding between the dissimilar features, which are computed using equation (12).

### 3.4. Hybrid Approaches

Hybrid approaches can be any approach where these two or more approaches can be combined. Recent research is drawing interest in these approaches. In the recent time author Cai et al.[40] used a hybrid technique of Edge based and the Information content approach. Detail analysis is shown in Table 6.

## 4. Parameter Identification

### 4.1. Identified Parameters

So far, we have discussed similarity measures in knowledge based system. To cope with the new emerging field of knowledge base system new measures are proposed every year[16]. In this section we have defined some new parameters for similarity measures. After conducting a thorough literature review, the following parameters have been identified for the evaluation of measures. For analysis, we have considered type of measure. Measure belongs to any of three mentioned category or it is based on the hybridization method.

Next, we have chosen feature set of measures that provides insight view of measures. By detailed study we get thirty different features for different measures.

Next parameter is accuracy; it is computed by the correlation value of the computed measure with the human judgment. If the correlation is high, accuracy is high for that measure.

Next parameter is ontology development framework; mainly three type of work present in the literature: single ontology [20], cross ontology [41] and fuzzy ontology [42].

Another parameter is relationships considered among words as *mango* "is-a" *fruit* so *mango* and *fruit* having "is-a" relationship between two words. However, "*mango*" is also related to the word "*seed*". Semantic similarity is also present in above defined words. A good similarity measure also computes the similarity among these two words. So, considering all the relationships among the words is also an important factor for any measure. Some of the measures computes good similarity with a "is-a" relationship but can't computes the similarity with other type of relationships; Some Edge based methods has high accuracy for "is-a" relationship, but not given same results on other type of relationship. A complete understanding of semantic relationship is an important factor for similarity measures.

Next parameter is dataset; three types of standard datasets are available for the WordNet ontology.

Last parameter is the ontology type; in that we have shown measures are applicable on which type of Ontology. As some of measures are applied Global ontology such as WordNet, MeSH and Gene ontology and some are applied for domain ontology like solar ontology or product

ontology. The parameters and their meaning are given in table1.

Table 1. Parameters and Meaning

Sl. No	Parameter	Meaning
1.	Type of Measure	It represents the approach that has followed as Edge based, Content based, Feature based or hybrid approach.
2.	Feature Set	Feature used for computing the similarity.
3.	Accuracy	It computes the correlation with human judgment and the value varies from [0- 1]. Near to 1 signifies high correlation value.
4.	Types of Ontology Framework	The types of ontology framework on which the measure works as single ontology, cross ontology or fuzzy ontology.
5	Benchmark Datasets	It specifies on which dataset the approach has been applied. There are various standard datasets are  1. Miller & Charles dataset[43]  2. Rubenstein & GoodEnough's[44]  3. Word Similarity353-Test Collection data-sets.[45]
6	Taxonomical relations	It specifies the relationship from the WordNet & Mesh such as "is-a", "part-of", "anatomy" etc. examined in the measures.
7	Ontology Type	Global Ontology like WordNet/Mesh/Gene ontology or any domain ontology.

## 4.2. Evaluation of different parameters

Here we have discussed the evaluation of identified parameters.

**Type of measure:** Research shows that first measure was proposed based on edge based and but current research focused on hybrid approaches. A hybrid approach gives more accuracy, flexibility and adaptability to the systems as compared to edge based, content and feature based approach.

**Feature-Set:** This parameter shows feature set used for various measures. Study demonstrates that in starting years of research, the main features were used such as "node", "lcs" and "depth". But in the resent research many different type of features are used such as:"case-based-reasoning", "lexical-pattern", "normalized-dis-

siimilarity", "cluster", "neighborhood concepts", "gloss-information", "fuzzy-distance", "vector-models", "swarm-optimization", "text-snippets", "support vector regression", "named entity overlap", "heuristic approach" and "Wikipedia" and many more. The detailed lists are shown in the table (2-6).

**Accuracy:** The accuracy for the different measures is shown in tables; for this we have considered previous research work presented in this area. The Author Alexendra et al. in 2006 [46] compares result of many previous research work and quoted that Jiang and Conrath [37] performs better than the other approaches at that time. For recent work different researcher[40,47,48] given their measures and compared their results with previous works. We have taken their results for accuracy parameter.

### Types of Ontology Framework:

1. Single Ontology- WordNet, Mesh and SUMO ontology are general purpose ontology. Mostly researchers used WordNet as main ontology for their measures. All the proposed measures work on the single ontology.
- 2 Cross Ontology - Measuring the similarity among the cross ontology has a wide application area of new ontology development. New ontology can be developed using merging of two ontologies. In merging; two words from different ontologies are combined. Like the word "fever" having same synonyms "Pyrogens" in MeSH ontology. For developing a new ontology for medical science we can utilized both the word "fever" and "Pyrogens".
- 3 Fuzzy Ontology- Fuzzy ontology is an emerging field of ontology development. Conventional ontologies are not suitable to manage imprecision or vagueness in information, to deal with probabilistic, uncertainty and vagueness in information, it was introduced Yen et al.[49,50,51]. Fuzzy ontology is the extension of crisp ontology and fuzzy membership is assigned between two concepts for a particular relationship. In recent years fuzzy ontology drawing interest of many research groups and detailed literature review in given by Author Devadoss [52].

**Data-Set:** We have found that mostly three standard dataset are used for different measures, mentioned in the table 1. Mostly researcher considered one or two dataset for their measures. Miller and Charles is the most popular dataset in this area. This dataset has 30 word pairs and three columns; the first column specify first word, second column specify other word and the last column gives the similarity among these two words according to the human judgment. It varies from [0-1] in data set, 1 means highest similarity. In the 353-TC dataset, it has 353 word pairs are present and the similarity varies from [0-4] among the words. Latest researcher considered the 353-TC dataset. In this dataset more emphasis on semantic relatedness as compare to similarity.

**Relationships–Considered:** Most common taxonomical relationship “*is-a*”. However some of the measures also computed the similarity among different words and considered more semantic relationship such as “part-of” and “anatomy”.

**Ontology Type-** Mostly presented work based on the WordNet ontology. Some of the measures considered the MeSH and Gene Ontology (medical science ontology). In

some of the paper’s domain ontologies was also considered.

### 5. A Relative Study on identified parameters for various Measures

Table 2. Edge Based Measures on Various Parameter

Sl. No	Author Name and Year	Feature set	Accuracy (correlation with human judgment)	Type of ontology(Single, cross ,fuzzy)	Benchmark – dataset	Taxonomical-Relationship	Type of ontology
1	Simple-path[31] [Rada,1989]	Only length is considered	0.51	Single	No data set	“ is-a”	WordNet
2	Sussna’s Measure[32] [Sussna,1993]	Length, direct and reverse link	0.53	Single	Miller & Charles	“is-a”	WordNet Ontology
3	Wu&palmer[33] [Wu&Palmer, 1996]	Depth and least common subsequence (lcs)	0.57	Single	Miller & Charles	“ is-a”	WordNet Ontology
4	L&C Measure[34] [Leacock & Chodorow, 1998]	Depth is considered similar word have similarity =1	0.61	Single	Miller & Charles	“is –a”	WordNet Ontology
5	Li’s Measure [35] [Li, Bander,2003]	Depth and length is considered	0.6	Single	Miller & Charles	“is –a”	WordNet ontology and MeSH ontology
6	Al and Ngugen [53] [Nguyen, 2006]	Cluster approach ,length and depth	0.67	Single	Miller & Charles	“is- a”, “part-of”	WordNet and MeSH ontology
7	James and Zhidian [54][Wang,2007]	Ancestor are considered	0.62	Single	No data set	“is-a” and “functional relationship”	Gene Ontology

Table 3. Information content based Measures on various parameters

Sl. No	Author Name and Year	Feature set	Accuracy	Type of ontology)	Benchmark – dataset	Taxonomical-Relation	Type of Ontology
1	Resnik [11] [Resnik, 1995]	Node based approach, lcs is used	0.59	single ,cross	Miller & Charles	“is- a”, ”part-of” relation is considered	WordNet
2	Lin [36] [Lin, 1999]	Based on Resnik approach ,and its ratio information	0.68	single ,cross	Miller & Charles	“is- a”, ”part-of” relation is considered	WordNet & MeSH
3	Ji&Corth[37] [Conrath,1997]	Resnik approach and lcs parameter	0.72	single, cross	Miller & Charles	“is- a”, ”part-of”, antonym	WordNet & MeSH

4	Ontology based concept similarity in FCA[55]	Similarity graph	0.78	single, cross	FCA ,domain ontology	“is- a “ relation is considered	domain ontology
5	A semantic similarity method based on Information content in multiple ontology[56] [Batet, 2011]	Intrinsic IC computing model, least common equivalent subsume	0.76	single, cross	miller & Charles	Only “is-a “relation is shown	MeSh, SNOWED CT ontology
6	Concept similarity in FCA an information content approach[57] [Formica, 2006]	Nouns, synsets of WordNet	Not mentioned	single, cross	Formal concept analysis	All relation can be considered	domain ontology

Table 4. Feature based Measures

Sl .No	Author Name and Year	Feature set	Accuracy	Type of ontology	Benchmark – dataset	Taxonomical-Relation	Type of Ontology
1	Tversky’s [38] [Tversky, 1997]	Common feature and uncommon feature is computed	0.74	single, cross	Miller & Charles	All relation can be used in WordNet	WordNet &MeSH ontology
2	Roand& Eg [39] [Roand,2001]	Neighborhood ,glosses, features synonyms	0.77	single, cross	miller & Charles	All relation can be used in WordNet	WordNet &MeSH ontology
3	X-Similarity: Computing Semantic Similarity between Concepts [45] [Petrakis,2006]	Feature based approach Synset and term description among cross ontology	0.72	Cross Ontology	Miller and Charles	All Relationship	MeSh and WordNet Ontology
4	Ontology –based semantic similarity feature based approach[58] [Sanchez,2012]	Taxonomical features, normalized dissimilarity , sub-sumbers, features generated	0.83	Single	Miller & Charles, Rubenstein& GoodEnough data set	“is-a” , “part – of” relationship	WordNet ontology
5	Towards the estimation of feature – based semantic similarity[59] [Albert,2014]	Common taxonomic subsumbers , static cost between two subsumbers	0.76	Single ,cross	Miller & Charles, Rubenstein& GoodEnough data set	All relation can be considered	WordNet ,MeSH ontology
6	Ontology based similarity for product information [60] [Akmal,,2014]	Case based reasoning , classes, feature replaced by words ,concept lattice	0.79	single, cross	Product dataset	All relation can be considered	Domain ontology
7	Feature based approaches using Wikipedia[61] [Yuncheng,2015]	Feature extraction using Wikipedia, formal representation of Wikipedia	0.83	single, cross	Miller and Charles, 353 data set	All relation can be consider	WordNet

Table 5. Hybrid Measures on various parameters

Sl. No	Author Name and year	Feature set	Accuracy	Type of ontology	Benchmark –dataset	Taxonomical-Relationship	Type of Ontology
1	Similarity using fuzzy distance[62][Valerie, 2004]	Fuzzy min and max distance	0.74	Cross and fuzzy	No dataset	All kind of relationship	WordNet

### Identification of New parameters for Ontology based Semantic Similarity Measures

2	Fuzzy Semantic Distance Measures [63] [Allahverdi, 2007]	Fuzzy distance, vector models, Bhattacharya distance	0.79	single, cross, fuzzy	No dataset	All Relationship	WordNet ontology ,merging multiple ontology
3	Fuzzy Measures on the Gene Ontology for Gene Product Similarity[64] [Mihail, 2006]	Annotation terms, fuzzy measure similarity, Choquet integral	0.83	single, cross ,fuzzy	No dataset	All relation is considered	Gene ontology
4	Evaluating Similarity Measures for Emergent Semantics of Social Tagging[65] [Benjamin,2009]	User-created tag, per user projection ,open dictionary	0.81	single, cross	social mining data	All relation can be considered	WordNet ontology
5	Similarity Measure using Fuzzy Formal concept [66][Selvi, Suresh; 2012]	Page count ,lexical pattern, pattern extraction cluster	0.76	single, cross, fuzzy	Miller and Charles dataset	All relation can be considered	WordNet
6	Semantic similarity estimation from multiple ontologies[56] [Batet,2011]	Ratio of taxonomic feature ,heuristic approach	0.81	single, cross	Miller and Charles dataset	is-a , part -of relationship	MeSh, SNOWED CT, WordNet ontology
7	Semantic similarity assessment of words using weighted WordNet[67] [Mostafa,Ehsan,2014]	Weighted edge distance, depth of the word ,swarm optimization	0.82	single, cross	Miller & Charles	All relation can be used in WordNet	WordNet ontology
8	Combining edge-counting and information content theory[68] [Jian-Bo, Xiao-Hua,2015]	Information content ,edge based weighted shortest path	0.8	single, cross	No dataset	Semantic Relatedness	WordNet ontology
9	A Novel Approach to Find Semantic Similarity Measure between Words[69][Lakshay,2014 ]	Page count, text snippets , edge counting approach, svm	0.76	Single, cross	Miller and Charles	All relation can be used in WordNet	WordNet ontology
10	LIPN-CORE[70] [ Davide,2013]	Support vector regression model, Levenshtein distance, Named Entities overlap, Wu-Palmer	0.83	single, cross	WordNet &MeSH ontology	Length ,distance, Depth, synonym, antonym	WordNet &MeSH ontology
11	FMLNCSIM: fuzzy measure-based similarity calculation[71] [Chen ,2016]	Fuzzy Measure IncRNA functional similarity	0.78	Single	Lnc RNADisease	All	MeSH ontology
12	A hybrid approach for measuring semantic similarity[40] [Cai,2017]	Hybrid Approach Intrinsic- IC and Edge based	0.87	Single Ontology	Miller and Charles TC-353	“is –a”	WordNet
13	Semantic similarity and relatedness between biomedical concepts [48] [Sam,2018]	Frequency count and UMLS hierarchy	0.84	Cross Ontology	Med-Line Text Data	All Relationship	UMLS Hierarchy
14	Distance learning techniques for ontology similarity[72] [Wei,2017]	Graph structure of Ontology Distance structure	0.82	Single Ontology	Textual Data	All Relationship	Go Ontology Plant Ontology

## 6. Discussion

We have given the detailed comparative study in the above section. These finding can be further used for defining the

New similarity measures in this area. Different analysis and results are given below:

In the Edge based measures, feature sets like length, depth, LCS, direct and indirect links were considered. These approaches are quite simple and execution time is less as they have to consider only length or the depth of the ontology. In these measures accuracy lies between the ranges of 0.51 to 0.67. It works only for the single ontologies. These measures considered only the “is-a” relationship among the ontology. They work well for the Global ontologies like WordNet and MeSH ontologies as these ontologies describes the complete structure of the general purpose ontology.

In the information content approach LCS, formal concept analysis, noun, synsets and graph similarity features sets were used. Here the accuracy lies between the 0.59-0.78. The accuracy depends on the coverage and the structure of the ontology. As the depth of ontology increases the similarity increases. They work well in single ontology and also works for the cross ontology using the LCS, if the LCS exists among the ontology, the similarity can be computed accordance with the LCS. In this type of measures “is-a”, “part-of” relationships can be considered if the relationships were present among ontologies. It works for the single, cross ontology framework where complete structure/relationships are defined among the ontologies.

Feature based approaches can be applied for the single, cross and fuzzy ontology framework. There accuracy lies between the 0.67-0.83. In WordNet we have the gloss information that can be used to calculate the similarity however MeSH ontology doesn’t provide such information. In this all the relation can be considered as these approaches not consider length and the structure of the ontology. These measures computed high correlation value among the words. However, none of work was presented, where fuzzy integration with feature based approaches was given. So, this can be future work in this area. These approaches furnish the same result on small or big ontology or on any type of ontology. These measures only depended upon the features of the concepts. For the small size and for the domain ontology they compute good accuracy with less complexity.

In Hybrid method, any two of the above method can be combined. It takes the advantages of any of two techniques. Accuracy lies between 0.76-0.87 in Hybrid approaches. In the hybrid work some researcher introduced the fuzzy concepts in their measures. The accuracy is high as compared to edge based and content base and comparable to the feature based approach. The features set in these techniques were used annotated terms, open dictionary, swarm optimization, heuristic approach and text snippets. These types of approaches are good and give high accuracy and can be used for the all relationship but it adds complexity to the system due to its features set. These approaches can be applied for fuzzy ontology. Very few measures were proposed in the area of fuzzy ontology framework.

Accuracy graph of different Similarity Measures is shown in Fig 5. Maximum accuracy is 0.86 by feature based approach[61] and Hybrid approaches[40,70].

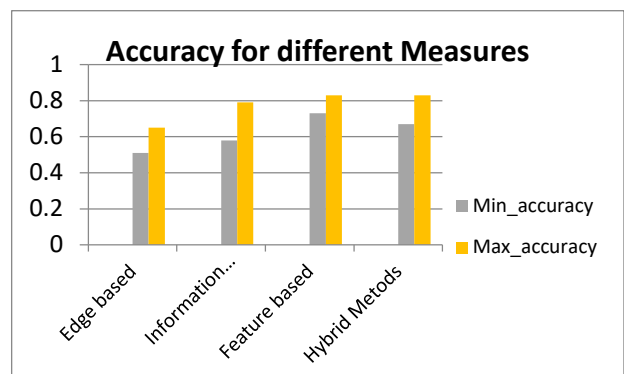


Figure 5 Accuracy Graph for the different measures.

Mostly research drawing interest in Hybrid approaches due to its high accuracy and adaptability to new research areas. Edge based approaches are good for global ontologies and but consider only taxonomical relations and works for single ontology. Information Content approaches works for cross ontologies and where complete structure and relationships are described among the ontologies. Feature based approaches work well in single and cross ontology. Only limitation with this method is to find the complete features of the concepts in the defined domain ontologies. Various measures work on the different ontologies framework is shown in Fig 6.

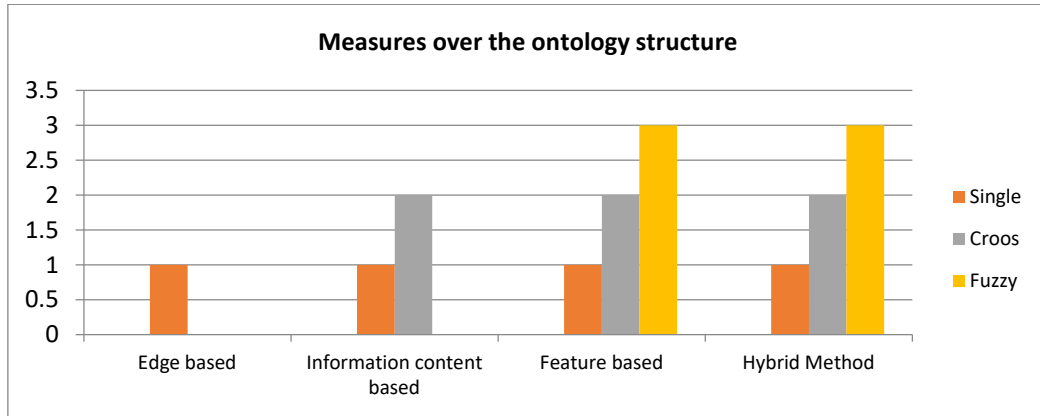


Figure 6 Measures over the Ontology's framework

Charles's dataset[43] is most popular dataset for WordNet. TC-353 dataset is used in recent research, as it has more emphasis on semantic relatedness as compared to similarity.

Data-set used by various measures shown in Fig.7.

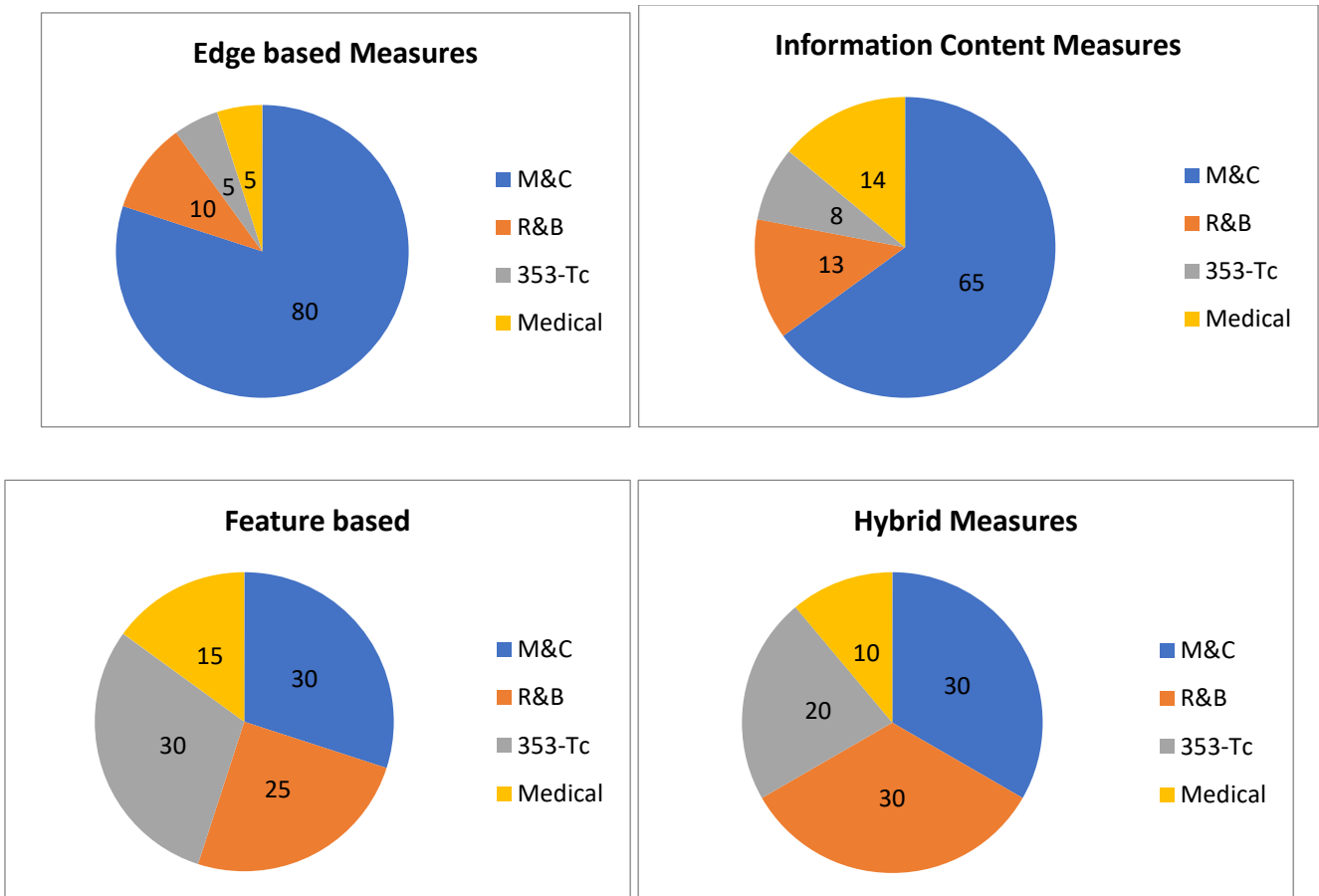


Figure 7 (a), (b), (c), (d) Data Set used in Various Measures.

Taxonomical relationships considered by the different knowledge- based approaches are shown in fig 8.

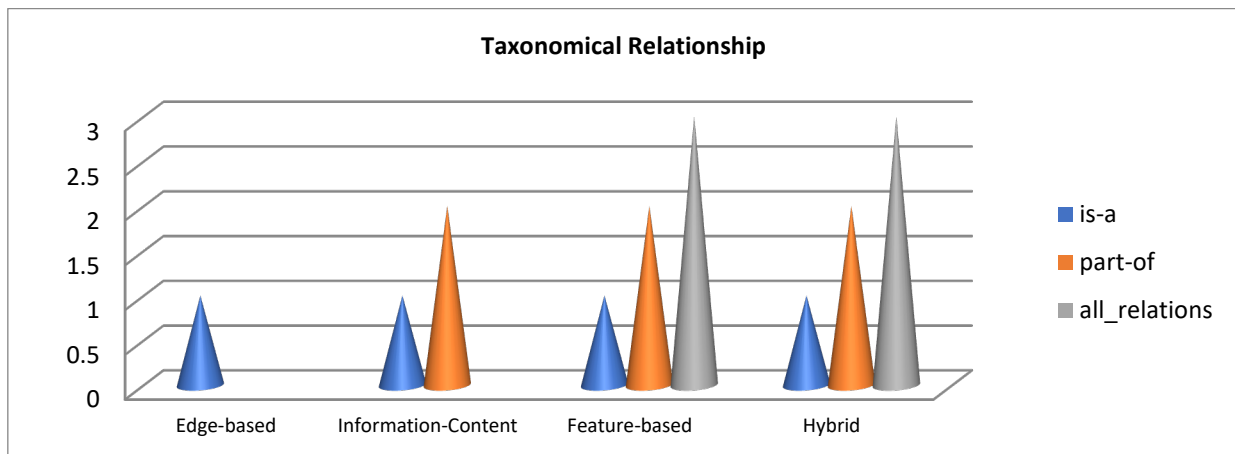


Figure 8 Representation of taxinomical relationships among the measures

## 7. Conclusion and future work

In this research paper, a detailed study of previous two decade work in field of semantic similarity is presented. In this study the state of the art similarity measures are compared based on the seven parameters like Type of Measure, Feature set, Accuracy, Ontology framework, Benchmark Datasets, Taxonomical relationships and Ontology Type. To cope up with new emerging fields new similarity measures are proposed every year. However, there are certain issues in their approaches. If similarity is high, then complexity is high due to their feature set. Similarity measure compute good result among the single ontology but not give the same result over the cross ontologies. Edge based measure works only in single ontology and only taxonomical relationships are considered. Information content methods are works in the area of cross ontology with the existence of LCS. Feature based approaches are found to give high correlation values and give good result on both single and cross ontologies. But some time it is difficult to determine the features of a concept in certain ontologies. Corpus based method like LSA and PMI can be further used for finding the features in the domain ontologies. These measures are used for the domain ontologies and features can be formulated using text processing methods.

To cope up with new emerging field of knowledge based new methods for similarity measured was proposed every year. Further we can identify more parameters in this area like complexity of the measures, tool/ language support for measures, ease of adaptability and ontological coverage.

## References

- [1] P. Kaur, P. Sharma, N. Vohra, R. Campus, R. Campus, and R. Campus, "An Ontology Based Text Analytics on Social Media," vol. 8, no. 5, pp. 233–240, 2015.
- [2] P. Kaur, M. Sharma, and M. Mittal, "Big Data and Machine Learning Based Secure Healthcare Framework ScienceDirect ScienceDirect Big Data and Machine Learning Based Secure Healthcare Framework," *Procedia Comput. Sci.*, vol. 132, no. June, pp. 1049–1059, 2018.
- [3] M. Mittal, H. Singh, K. K. Paliwal, and L. M. Goyal, "Efficient random data accessing in MapReduce," *2017 Int. Conf. Infocom Technol. Unmanned Syst. Trends Futur. Dir. ICTUS 2017*, vol. 2018–January, pp. 552–556, 2018.
- [4] R. Singh, A. Gehlot, M. Mittal, and R. Samkaria, *Smart and Innovative Trends in Next Generation Computing Technologies*, vol. 827. Springer Singapore, 2018.
- [5] A. Stavrianou and N. Nicoloyannis, "Overview and Semantic Issues of Text Mining," vol. 36, no. 3, pp. 23–34, 2007.
- [6] L. Márquez, "Machine Learning and Natural Language Processing," *Apprendizaje automático Appl. al Proces. del Leng. Nat.*, pp. 1–53, 2000.
- [7] A. Aamodt *et al.*, "Requirements Engineering," *Requir. Eng.*, vol. 19, no. 1, pp. 35–46, 2013.
- [8] T. Hong-minh and D. Smith, "Word Similarity In WordNet," *East*, pp. 293–302, 2008.
- [9] D. Bollegala, Y. Matsuo, and M. Ishizuka, "A Web Search Engine-Based Approach to Measure Semantic Similarity between Words," vol. 23, no. 7, pp. 977–990, 2011.
- [10] C. D'Amato, "Similarity-based Learning Methods for the Semantic Web," *Similarity based Learn.*, pp. 1–226, 2007.
- [11] P. Resnik, "Using Information Content to Evaluate Semantic Similarity in a Taxonomy," vol. 1, 1995.
- [12] R. Mihalcea, C. Corley, and C. Strapparava, "Corpus-based and knowledge-based measures of text semantic similarity," *Proc. 21st Natl. Conf. Artif. Intell.*, vol. 1, pp. 775–780, 2006.
- [13] K. R. Uthayan and G. S. A. Mala, "Hybrid Ontology for Semantic Information Retrieval Model Using Keyword Matching Indexing System," vol. 2015, no. 1, 2015.
- [14] S. Weng, H. Tsai, S. Liu, and C. Hsu, "Ontology construction for information classification," vol. 31, pp. 1–12, 2006.
- [15] N. Duhan and M. Mittal, "Opinion Mining using Ontological Spam Detection," 2017.

- [16] S. Harispe, S. Ranwez, S. Janaqi, and J. Montmain, "Semantic Similarity from Natural Language and Ontology Analysis," *Synth. Lect. Hum. Lang. Technol.*, vol. 8, no. 1, pp. 1–254, 2015.
- [17] T. M. Tuan, P. M. Chuan, M. Ali, T. T. Ngan, M. Mittal, and L. H. Son, "Fuzzy and neutrosophic modeling for link prediction in social networks," *Evol. Syst.*, vol. 0, no. 0, p. 0, 2018.
- [18] M. Mittal, L. M. Goyal, J. K. Sethi, and D. J. Hemanth, "Monitoring the Impact of Economic Crisis on Crime in India Using Machine Learning," *Comput. Econ.*, pp. 1–19, 2018.
- [19] A. Hliaoutakis, G. Varelas, E. Voutsakis, E. G. M. Petrakis, and E. Milios, "Information Retrieval by Semantic Similarity," *Int. J. Semant. Web Inf. Syst.*, vol. 2, no. 3, pp. 55–73, 2006.
- [20] A. M. B. Abdelrahman and A. Kayed, "A Survey on Semantic Similarity Measures between Concepts in Health Domain," *Am. J. Comput. Math.*, vol. 05, no. June, pp. 204–214, 2015.
- [21] A. Fahad *et al.*, "A survey of clustering algorithms for big data: Taxonomy and empirical analysis," *IEEE Trans. Emerg. Top. Comput.*, vol. 2, no. 3, pp. 267–279, 2014.
- [22] A. Hliaoutakis, "Semantic Similarity Measures in MeSH Ontology and their application to Information Retrieval on Medline," *Interface*, pp. 1–79, 2005.
- [23] A. Budanitsky and G. Hirst, "Evaluating WordNet-based Measures of Lexical Semantic Relatedness," *Comput. Linguist.*, vol. 32, no. 1, pp. 13–47, 2006.
- [24] T. K. Landauer and P. W. Foltz, "An Introduction to Latent Semantic Analysis," pp. 259–284, 1998.
- [25] B. Xue, C. Fu, and Z. Shaobin, "A study on sentiment computing and classification of sina weibo with Word2vec," *Proc. - 2014 IEEE Int. Congr. Big Data, BigData Congr. 2014*, no. 2013, pp. 358–363, 2014.
- [26] G. a. Miller, "WordNet: a lexical database for English," *Commun. ACM*, vol. 38, no. 11, pp. 39–41, 1995.
- [27] T. R. Gruber, "A translation approach to portable ontology specifications," *Knowl. Acquis.*, vol. 5, no. 2, pp. 199–220, 1993.
- [28] C. Greenbacker, "WordNet Similarity Metrics," pp. 1–36, 2011.
- [29] K. Erk, "Vector Space Models of Word Meaning and Phrase Meaning: A Survey," *Linguist. Lang. Compass*, vol. 6, no. 10, pp. 635–653, 2012.
- [30] D. Lin, "An Information-Theoretic Definition of Similarity," *Proc. ICML*, pp. 296–304, 1998.
- [31] R. Rada, H. Mili, E. Bicknell, and M. Blettnet, "Development and Application of a Metric on Semantic Nets," *IEEE Trans. Syst. Man Cybern.*, vol. 19, no. 1, pp. 17–30, 1989.
- [32] M. Sussna, "Word sense disambiguation for free-text indexing using a massive semantic network," *Proc. Second Int. Conf. Inf. Knowl. Manag. - CIKM '93*, pp. 67–74, 1993.
- [33] Z. Wu and M. Palmer, "Verb Semantics and Lexical Selection," 1994.
- [34] C. Leacock, G. a. Miller, and M. Chodorow, "Using corpus statistics and WordNet relations for sense identification," *Comput. Linguist.*, vol. 24, no. 1, pp. 147–165, 1998.
- [35] Y. Li, Z. A. Bandar, and D. Mclean, "An Approach for Measuring Semantic Similarity between Words Using Multiple Information Sources," vol. 15, no. 4, pp. 871–882, 2003.
- [36] D. Lin, "Automatic identification of non-compositional phrases," *Proc. 37th Annu. Meet. Assoc. Comput. Linguist. Comput. Linguist.*, pp. 317–324, 1999.
- [37] J. J. Jiang and D. W. Conrath, "Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy," *Proc. Int. Conf. Res. Comput. Linguist.*, 1997.
- [38] A. Tversky, "Features of similarity. - 1977 - Tversky.pdf,"
- [39] M. J. Egenhofer, "Determining Semantic Similarity Among Entity Classes from Different Ontologies M. Andrea Rodríguez and Max J. Egenhofer IEEE Transactions on Knowledge and Data Engineering," no. 1.
- [40] Y. Cai, Q. Zhang, W. Lu, and X. Che, "A hybrid approach for measuring semantic similarity based on IC-weighted path distance in WordNet," 2017.
- [41] R. Cao, L. Wu, R. Wang, and C. Yang, "Semantic Similarity Algorithm Based on Generalized Regression Neural Network," no. Icismme, pp. 1333–1336, 2015.
- [42] N. T. Nguyen and H. B. Truong, "A Consensus-Based Method for Fuzzy Ontology Integration \*," pp. 480–489, 2010.
- [43] G. A. Miller and W. G. Charles, "Language and Cognitive Processes Contextual correlates of semantic similarity Contextual Correlates of Semantic Similarity," *Lang. Cogn. Process.*, vol. 6, no. 1, pp. 1–28, 1991.
- [44] H. Rubenstein and J. B. Goodenough, "Contextual correlates of synonymy," *Commun. ACM*, vol. 8, no. 10, pp. 627–633, 1965.
- [45] E. G. M. Petrakis and G. Varelas, "X-Similarity : Computing Semantic Similarity between Concepts from Different Ontologies," vol. 4, no. 4, pp. 233–237, 2006.
- [46] A. Budanitsky and G. Hirst, "Evaluating WordNet-based Measures of Lexical Semantic Relatedness," *Comput. Linguist.*, vol. 32, no. 1, pp. 13–47, 2006.
- [47] Y. Jiang, X. Zhang, Y. Tang, and R. Nie, "Feature-based approaches to semantic similarity assessment of concepts using Wikipedia," *Inf. Process. Manag.*, vol. 51, no. 3, pp. 215–234, 2015.
- [48] S. Henry, A. Mcquilkin, and B. T. Mcinnes, "Artificial Intelligence In Medicine Association measures for estimating semantic similarity and relatedness between biomedical concepts," *Artif. Intell. Med.*, no. August, pp. 1–10, 2018.
- [49] C. S. Lee, Z. W. Jian, and L. K. Huang, "A fuzzy ontology and its application to news summarization," *IEEE Trans. Syst. Man, Cybern. Part B Cybern.*, vol. 35, no. 5, pp. 859–880, 2005.
- [50] D. H. Widyantoro and J. Yen, "A fuzzy ontology-based abstract search engine and its user studies," *10th IEEE Int. Conf. Fuzzy Syst.*, vol. 2, pp. 1291–1294, 2001.
- [51] J. Singh *et al.*, "Fuzzy LogicHybrid Model with Semantic Filtering Approach for Pseudo Relevance Feedback-based Query Expansion," *Ieee*, no. February 2018, 2017.
- [52] N. Devadoss and S. Ramakrishnan, "Knowledge Representation using Fuzzy Ontologies – A Review," vol. 6, no. 5, pp. 4304–4308, 2015.
- [53] H. A. Nguyen and H. Al-mubaid, "A Combination-based Semantic Similarity Measure Using Multiple Information Sources."
- [54] J. Z. Wang, Z. Du, R. Payattakool, P. S. Yu, and C. F. Chen, "A new method to measure the semantic similarity of GO terms," *Bioinformatics*, vol. 23, no. 10, pp. 1274–1281, 2007.
- [55] A. Formica, "Concept similarity in Formal Concept Analysis : An information content approach," vol. 21, pp. 80–87, 2008.
- [56] M. Batet, D. Sánchez, and A. Valls, "An ontology-based measure to compute semantic similarity in biomedicine," *J. Biomed. Inform.*, vol. 44, no. 1, pp. 118–125, 2011.

- [57] A. Formica, "Ontology-based concept similarity in Formal Concept Analysis," *Inf. Sci. (Nij.)*, vol. 176, no. 18, pp. 2624–2641, 2006.
- [58] D. Sánchez, M. Batet, D. Isern, and A. Valls, "Ontology-based semantic similarity: A new feature-based approach," *Expert Syst. Appl.*, vol. 39, no. 9, pp. 7718–7728, 2012.
- [59] A. Solé-ribalta, D. Sánchez, and F. Serratos, "Knowledge-Based Systems Towards the estimation of feature-based semantic similarity using multiple ontologies," vol. 55, pp. 101–113, 2014.
- [60] S. Akmal, L. H. Shih, and R. Batres, "Ontology-based similarity for product information retrieval," *Comput. Ind.*, vol. 65, no. 1, pp. 91–107, 2014.
- [61] Y. Jiang, X. Zhang, Y. Tang, and R. Nie, "Feature-based approaches to semantic similarity assessment of concepts using Wikipedia," *Inf. Process. Manag.*, vol. 51, no. 3, pp. 215–234, 2015.
- [62] V. Cross, "Fuzzy semantic distance measures between ontological concepts," *IEEE Annu. Meet. Fuzzy Information, 2004. Process. NAFIPS '04.*, vol. 2, pp. 1–6, 2004.
- [63] R. Saracoglu, K. Tutuncu, and N. Allahverdi, "A fuzzy clustering approach for finding similar documents using a novel similarity measure," *Expert Syst. Appl.*, vol. 33, no. 3, pp. 600–605, 2007.
- [64] M. Popescu, J. M. Keller, and J. A. Mitchell, "Fuzzy measures on the gene ontology for gene product similarity," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 3, no. 3, pp. 263–274, 2006.
- [65] B. Markines *et al.*, "Evaluating similarity measures for emergent semantics of social tagging," *Proc. 18th Int. Conf. World Wide Web, WWW' 09, ACM*, pp. 641–650, 2009.
- [66] K. Selvi and R. M. Suresh, "Context similarity measure using fuzzy formal concept analysis," *ACM Int. Conf. Proceeding Ser.*, vol. 3, no. 12, pp. 420–426, 2012.
- [67] M. Ghazizadeh Ahsae, M. Naghibzadeh, and S. E. Yasrebi Naeini, "Semantic similarity assessment of words using weighted WordNet," *Int. J. Mach. Learn. Cybern.*, vol. 5, no. 3, pp. 479–490, 2014.
- [68] J. B. Gao, B. W. Zhang, and X. H. Chen, "A WordNet-based semantic similarity measurement combining edge-counting and information content theory," *Eng. Appl. Artif. Intell.*, vol. 39, pp. 80–88, 2015.
- [69] L. Sahni, A. Sehgal, S. Kochar, F. Ahmad, and T. Ahmad, "A Novel Approach to Find Semantic Similarity Measure between Words," *Comput. Bus. Intell. (ISCBI), 2014 2nd Int. Symp.*, pp. 89–92, 2014.
- [70] D. Buscaldi, J. Le Roux, J. J. G. Flores, and A. Popescu, "LIPN-CORE: Semantic Text Similarity using n-grams, WordNet, Syntactic Analysis, ESA and Information Retrieval based Features," *Proc. 2nd Jt. Conf. Lex. Comput. Semant.*, vol. 1, pp. 162–168, 2013.
- [71] X. Chen, Y.-A. Huang, X.-S. Wang, Z.-H. You, and K. C. C. Chan, "FMLNCSIM: fuzzy measure-based lncRNA functional similarity calculation model," *Oncotarget*, vol. 7, no. 29, 2016.
- [72] W. Gao, M. Reza, F. Adnan, and A. Sunilkumar, "Distance learning techniques for ontology similarity measuring and ontology mapping," *Cluster Comput.*, 2017.