Special Issue Published in International Journal of Trend in Research and Development (IJTRD), ISSN: 2394-9333, www.ijtrd.com

# A Survey: Efficient Semantic Comparison& Effective Information Retrieval System

[1]M.Muralidharan and [2]Dr.V.Valli Mayil
[1]Research Scholar, Periyar Maniammai University, Vallam Thanjavur, TamilNadu, India
[2]Associate Professor and Head, Department of Computer Science and Applications, Periyar Maniammai University, Vallam Thanjavur.

***Abstract:*** **S**emantic comparison relates to processing the comparison among sympathetically comparative than not basically lexically relative terms. The comparison among biomedicinal terms/conceptsis a awfully commanding task for biomedicinal evidence extraction and knowledge discovery. The procedures and tests are tools appliedto characterize how to quantity the goodness of ontology or its properties. In this paper, a proportional study on characteristic methods such as pathway based, information content based, highlight based and crossover comparison measures is done for identifying semantically comparative ideas in ontology. The center is on more than one ontology techniques since it is interesting than the single ontology and semantic similarities are processed among terms stemming from distinctive ontologies (WordNet and MeSH, SNOMED-CT, ICD) in this work). The reason of this survey is to explore how these comparison calculation techniques could assist to improve the retrieval adequacy of Information retrieval models based on web Ontology.

***Key terms:*** *Ontology, Highlight Based Measure, Information Retrieval, Retrieval Effectiveness, Semantic Comparison Measures Crossover Measure;*

## I. INTRODUCTION

Comparison plays a central role in information management, especially in the concontent of environment like the semantic web where information may originate from distinctive sources and has to be combined and integrated in a flexible way. Semantic comparisonis a metric ended a usual of forms based on the likeliness of their significance, which alludes to evaluationamong two philosophiesin a taxonomy or ontology and it is achieved through ontology or taxonomies to characterize a separation among words or utilizing statistical means. comparisonamong ideas is a quantitative measure of information, processed based on the properties of the ideas and their relationships. With the initiation of Semantic Mesh, the semantic comparisonmeasures are fetchingimperative workings in Information Extraction (IE), Information Retrieval (IR) and furthercunningknowledgebased schemes. Potential application for these measures includes search, learning disclosure in database and information mining or decision support frameworks that utilize ontology. Semantic comparison alludes to the comparison of two ideas within a given ontology or taxonomy.

Semantic comparisonamongphilosophiesis a procedureto measure the semantic comparison, or the semantic separation among two philosophiescompatibleto a assumed ontology. In additionalrelations, semantic comparisonis utilized to separatephilosophieshaving steady"characteristics". While human do not recognize the officialdescription of connectionamongperceptions, can reviewerunderstandingamong them. For specimen, a slightkid can communicate that “apple” and “peach” devise more connected to each other than “apple” besides “tomatoes”. These collectionsof philosophies stay correlated to each other and its constructiondescription is properlyso-called “is-a” grading. Semantic comparison methods fetching intensively exploitedfor supreme applications of cunningknowledge-based and semantic data retrieval structures(distinguish an ideal competition among query terms and papers), intellect disambiguationalso Bioinformatics. Semantic comparison and semantic understanding are two connected words, but semantic comparisonis further exactconnectionbesides can be measured as a typeof semantic comparison. Aimed atdesign‘Student’ and ‘Professor’ are the connectedrelations, which are not related. Totally the comparative ideasare connected and the immorality versa is not constantlyaccurate.

Semantic comparisonand semantic separation are characterized equally. Leasestay$C1$ and $C2$ two ideas that fit to two distinctive hubs $n1$ and $n2$ in a assumed ontology, the separation among the hubs ($n1$ and $n2$) limits the comparisonamong these two ideas $C1$ and $C2$. Both n1 and n2 can be reflected as an ontology (alsonamedidea nodes) that covers a usual of relationsidentical and accordingly. Doublerelations are identical if they are in the equal node then their semantic comparisonis exploited.

The procedure of ontologies to denote the philosophiesor relations (humans or PCs) describingindividualinteractivecauses are beneficial to brandknowledgefrequentlylogical. Furthermore, it is conceivable to usage ofdistinctive ontologies to denote the ideas of everylearning cause. Later, the planning or ideas relatingconstructed on the matching or distinctive ontologies certifieslearning allotmentamong concepts. The planningwants to discovery the comparisonamong the relations or ideas constructed on space definiteontologies. The comparisonamongideas or elements be able todistinguished if they offer regular features or if they are connected to further semantically related essentialsin an ontology. On behalf ofinstance, the planningamong the KIMP ontology and MeSH ontology supports to distinguish the connectionthrough the homogenousmedicinal relations which expands the reusability then the disclosure of the furtheralliedideas. This paper center on semantic comparison. It computes four groupingsof semantic comparisonmeasures depicted in nonfictions. Everytactic of semantic comparisonmeasure has been associatedtoward others in the equalclassification and estimated.

This paper is organized as follows. Area 2 portrays some examples of perceived ontologies utilized with semantic comparisonmeasures. Area 3 offerings the classifications of semantic comparisonmethods. Area 4 gives an assessment of the depicted semantic comparisonmeasures. Area 5 is the conclusion and future.

## II. RELATED WORK

Topicsallied to semantic comparison calculations alongside with problemslinked to processing semantic comparisonon Word Net also MeSH are conversed below.

### *A. Word Net*

Word Net is an on-line philologicalsituationframework created at Princeton University. Nouns, verbs, adjectives and adverbs are congregated into substitutegroups (synsets). The synsets are additionallycontrolled into faculties (i.e., matching to distinctive imports of the equalperiod or idea). The synsets are associated to further synsets advanced or inferior in the gradingcategorizedby distinctive categories of affairs. The maximumregular connections are the Hyponym/Hypernym (i.e., Is-A connection), and the Meronym/Holonym (i.e., Part-of connection). There are nine thing and some verb Is-A hierarchies (adjectives also adverbs are not ordered into Is-A hierarchies).

### *B. MeSH*

MeSH is a ordered hierarchy of medicinal and biological relations (or notions) recommended by the U.S National Library of Medicine (NLM). MeSH positions are planned in Is-A classificationsthroughextracommonrelations[“chemicals and drugs”] higher in a classification than newdetailedterms [“aspirin”]. Every MeSH duration is portrayedby numerouspossessions, the greatestauthoritativeof them actuality the MeSH Heading (MH) [i.e., term name or identifier], Possibility Note [i.e., a content report of the tenure] and EntrancePositions [i.e., generally synonym relations to the MH]. In this work, entrancepositions are preserved as synonyms.

### *C. Semantic comparison*

Numeroustechniques aimed atdeterminingsemantic comparisonamongrelations have been projected in the writing plusselected of them have been confirmedarranged Word Net. We existent an assessment for a addedcomprehensivebesides up-to-date established of techniques also we research cross ontology methods. Comparative outcomes on MeSH have not stayeddescribed in the nonfiction. Comparisonmeasures put onfirst for nouns [also verbs in Word Net] and for Is-A relations. Ordered chattelssimilarunity, individuality and variancethings for adverbs and adjectives ensure not occur. Semantic comparison techniques are categorized into subsequentcentraltechniques:

(i) Edge Counting Techniques
(ii) Information Content Techniques
(iii) Hybrid measure
(iv) Feature based measure

## III. SEMANTIC COMPARISONMEASURES IN INFORMATION RETRIEVAL

Assessing semantic comparisonof ideas is a difficult that takesexistedexpansivelyresearched in the writing in distinctive zones, such as industrialintellect, intellectualknowledge, databases and software industrial. Semantic comparisonnarrates to processing the comparisonamongattentively comparative then not essentially lexical comparative relations. Now, it is increasingpopularstanding in distinctive situations, such by way ofarithmeticalarchives, heterogeneous recordsalso in specific the Semantic Mesh. In such situations, frequentlydesignsare plannedconcurring to taxonomy [or a hierarchy]. We studymethodologies to process the semantic comparisonamongregular morphologicalrelations. In this paper, the new methodology for computing semantic comparisonamong words and orderedconstruction is utilized to currentinfocontented. This paper, extant anaspectengine utilizing Google API that develops the client question constructed on comparisonmarks of every term of client’s query.Clients question arguments are changed with synonyms revealedafter the comparisonmeasures alsoinfo to the Google look API.

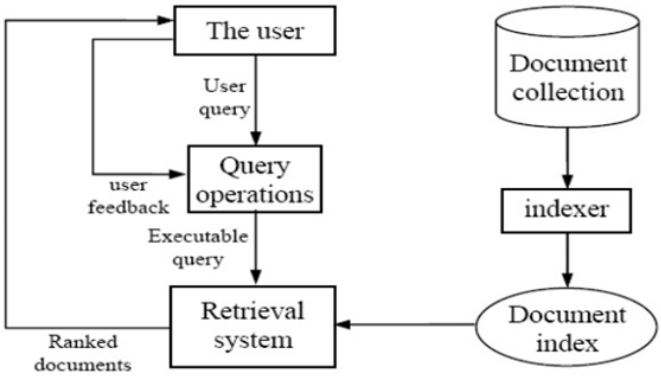



Figure 1: Architecture of Information Retrieval

### *A. WordNet Taxonomy*

WordNet is a lexical record for the English language . It groups English words into collections of synonyms called synsets, gives small, commonmeanings, and records the different semantic familiesamong these synonym groups. The exactmeaning of one word under one sort of POS is called a s ense. Each synsethas a glossthat describes the idea it signifies. For example, the words night, nighttime, and dull establish a single synset that has the following shine: the time after sunset and before sunrise while it is dull outside. The reason is twofold: to produce a mixture of vocabulary and thesaurus that is more naturally usable, and to support automatic content investigation and manufactured knowledge uses.

### *B. WordNet Database*

For every syntactic group, two records signify the WordNetrecord — index.pos and data.pos, where pos is either noun, verb, adj or adv. The record is in an ASCII layout that is social- and machine-readable, and is certainlyavailable to those who hope to use it with their own uses. The index and information records are interconnected. The WordNet morphological processing role, morphy(), handles a extensivekind of morphological changes. During WordNet growth synsets are structured into forty-five lexicographer records based on syntactic classification and logical groups. grind() processes these records and produces a recordappropriate for use with the WordNet library, interface code, and other uses.A record number relates to each lexicographer file. Record numbers are prearranged in numerous parts of the WordNet framework as an productive mode to show a lexicographer record name.

### *C. WordNet as an ontology*

The hypernym/hyponym connections between the thing synsets can be translated as specialization familiesamongtheoretical groups. In other words, WordNet can

be translated and utilized as a lexical ontologyin the computer science logic.

The WordNet dictionary covers the faculties of words. The recurrence of specific sense is given in enclosure and "n" show the thing (n in parenthesis).

### *D. Semantic comparison utilizing Information Content*

WordNet associatesideas or wisdoms, but maximum words have additional than one sense. Word comparisoncan be determined by the finesttheoretical comparison esteem among all the idea (sense) pairs.

In this paper, they present a idea comparison coordinating technique based on information content utilizing the order of WordNet. The outcomes give the comparisonprocesses of words. We have found that substitutingquestion with set of synonyms based on the matchscore can trulyincrease the information retrieval (IR) task. Clients repeatedly fail to define the information they want to recover in the look query.

**Result**

In upcomingeffort, we are prolonging the semantic coordinating method by processing semantic comparisonbetweendistinctive ontologies. The algorithm exhibited here can be further improved with integrating Word Sense Disambiguation (WSD). With the processed match, in the comparison calculation unit, WSD can be achieved by maximizing similarity for the group of the ideas necessary by the question developmentunit.

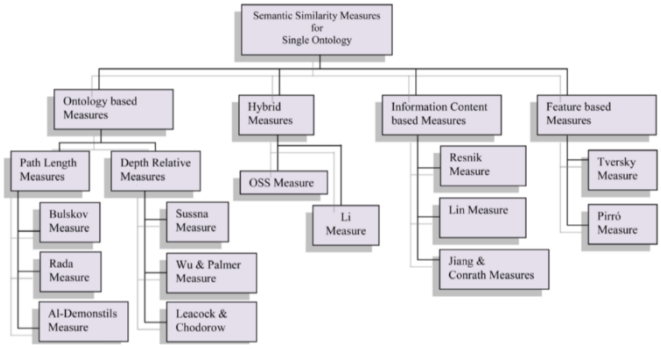
## IV. ASSESSMENT OF SEMANTIC COMPARISONMEASURES APPROACHES

In current years, semantic parallelmeasure has a excessive interest in Semantic Web and Regular Language Processing (NLP). Several comparisonmeasures have been established, being given the presence of a organized learning representation offered by ontologies and corpus which enable semantic explanation of terms. Semantic comparisonmeasures process the comparisonamongideas/terms included in learning sources in order to execute estimations. This paper discusses the current semantic comparison techniques based on construction, information content and highlight methods.Moreover, we present a dangerousassessment of numerousclassifications of semantic comparisonmethods based on two standard benchmarks. The point of this paper is to give an productive assessment of all these processes which help specialist and experts to select the measure that topsuitable for their necessities.

### *A. Semantic Measure Approaches*

Several techniques of deciding semantic measures have been planned in the last few periods. Three aspectsrelated with the ontology ordered order can be listed: The path length factor, profundity component and neighborhood thickness component in the order do affecting (although not meaningfully) the semantic separation measure. The thickness of two ideas C1 and C2 is the number of lads of the ideas which belong to the straight path from the root to the most exact regular subsumer of two ideas C1 and C2.

The comparisonmeasures can affected by the regular qualities of the matched concepts. The differences among the ideas cause the measures to reduction or to rise with unity. In addition, the comparisonmeasures and the taxonomy can be linked (ordered relations), i.e. the position of the ideas in the taxonomy and the number of hierarchic joins are measured. Additionally, comparisonmeasures take into account the information content of the ideas, whether they are coveredor infinite values, whether they are symmetric and whether they give distinguishingangles. All the proprieties will be deliberated in each class of comparisonmeasure.




*(i) Structure-based measures*

Structure-based or edge counting measures signify the measures that use a capacity that processes the semantic comparisonmeasure in ontology hierarchy structure (is-a, part-of). The capacity processes the extent of the routeconnecting the terms and on the location of the terms in the taxonomy. Thus, the more comparative two ideas are, the more joins there are between the ideas and the more closely linked.

*(ii) Information Content Measures*

Information content (IC) based measures are those measures that use the information content of ideas to measure the semantic comparisonamong two concepts/terms. The information content esteem of a idea is processed based on the recurrence of the term in a given filegroup. The next Area presents a great number of semantic comparisonmeasures. All of them use the information content of the shared guardian of two terms C1 and C2 (Equation 6), where S(C1; C2) is the set of ideas that subsume C1 and C2. The two ideas can offer parents by multiple routes. The minimum p(C) is utilized when there is more than one shared guardian where C is the most enlightening subsume (MIS).

To compute the comparisonof two words, the information content of the most enlightening subsume is used.

*(iii) Feature-Based Measures*

The study of the structures of a term is extremely significant, because it coversvalued information regardinglearning about the term. Highlight based measure assumes that each term is depicted by a set of terms specifying its properties or structures. The comparisonmeasure among two terms is characterized as a capacity of their properties (e.g., their definitions or "glosses" in WordNet) or based on their connections to other comparative terms in hierarchical structure.

*(iv) HybirdMeasures*

Crossover measures combine the structural qualities depicted above (such as route length, profundity and neighborhood thickness) and some of the above exhibited methods. Although, their exactness for a concrete condition is

higher than more basic edge-counting measures, which depend on the experimentalchange of weights concurring to the ontology and info terms.

**Result**

Semantic comparison assessment is a great component included in many applications encased in the manufactured knowledge relook area. Based on the theoretical values and the way in which ontologies are researched to process comparison, distinctive types of techniques can be recognized. This paper gives an innovativeexamined of the most perceived semantic comparisonmeasures that can be utilized to evaluation the similarityamongideas or terms. This paper has examined, with the point of giving some visions on the correctness, the typology and the key properties of the depicted measures under each group. In addition, an productive comparison of all these measures in a realsite is presented, utilizing the two extensivelyutilized benchmarks. The benefices decided from those examines would help the specialist and experts to select the measure that well fits with the prerequisites of a real use.

## CONCLUSION

In this work, we discuss the basics of semantic comparisonmeasures, the classification of single ontology comparisonmeasures. We plan a brief overview of the different semantic comparisonmeasures in web ontology and health. We moreover plan to implement a web-based client interface for all these semantic comparisonmeasures and to make it available freely to specialists over the Internet. That will be much helpful for interested specialists in the field of bioinformatics content mining.

Ontology portrays the space of discourse, intended for sharing among distinctive applications and it is expressed in a language that can be utilized for reasoning. By utilizing distinctive calculations the efficiency can be improved for searching the information's. By utilizing the distinctive ontology techniques we can map or match distinctive ontologies.We tested with numerous semantic comparison techniques for processing the theoretical comparison among regular language terms utilizing WordNet and MeSH.

### *References*


[1] Itziar Aldabe; Montse Maritxalar, "Semantic comparison Measures for the Generation of Science Tests in Basque", IEEE Transactions on Learning Technologies Year: 2014, Volume: 7, Issue: 4 Pages: 375 – 387.

[2] Sheau-Ling Hsieh; Wen-Yung Chang; Chi-Huang Chen; Yung-Ching Weng, "Semantic comparison Measures in the Biomedical Domain by Leveraging a Web Search Engine", IEEE Journal of Biomedical and Health Informatics Year: 2013, Volume: 17, Issue: 4 Pages: 853 – 861.

[3] Danushka Bollegala; Yutaka Matsuo; Mitsuru Ishizuka, "A Web Search Engine-Based Approach to Measure Semantic comparisonamong Words", IEEE Transactions on Knowledge and Data Engineering, Year: 2011, Volume: 23, Issue: 7, Pages: 977 – 990.

[4] Hisham Al-Mubaid; Hoa A. Nguyen, "Measuring Semantic comparisonAmong Biomedical Concepts Within Multiple Ontologies", IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) Year: 2009, Volume: 39, Issue: 4 Pages: 389 – 398.

[5] Xuebo Song; Lin Li; Pradip K. Srimani; Philip S. Yu; James Z. Wang, "Measure the Semantic comparison of GO Terms Using Aggregate Information Content", IEEE/ACM Transactions on Computational Biology and Bioinformatics Year: 2014, Volume: 11, Issue: 3 Pages: 468 – 476.

[6] Hsun-Hui Huang; Yau-Hwang Kuo, "Cross-Lingual Document Representation and Semantic comparison Measure: A Fuzzy Set and Rough Set Based Approach", Hsun-Hui Huang; Yau-Hwang Kuo IEEE Transactions on Fuzzy Systems Year: 2010, Volume: 18, Issue: 6 Pages: 1098 – 1111.

[7] Y. Li; Z. A. Bandar; D. Mclean, "An approach for measuring semantic comparisonamong words using multiple information sources", IEEE Transactions on Knowledge and Data Engineering Year: 2003, Volume: 15, Issue: 4 Pages: 871 – 882.