International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249 – 8958 (Online), Volume-9 Issue-5, June 2020

Application of WordNet for Text Analysis in Different Domains

Suyash Lakhani, Ridhi Jhamb, Mayank Arora, Saravanakumar Kandasamy

Abstract*: The following paper examines and illustrates various problems which occur in the field of Natural Language Processing. The solutions used in these papers use WordNet in one way or the other to enhance or improve the efficiency of the projects.WordNet can therefore be viewed as a combination and an augmentation of a word reference and a thesaurus. While it can be used by developers and programmers via a web browser, its prime use is in automatic text analysis and applications based on AI.

Keyword*: Cosine sentence similarity, semantic significance, superposition effect, structural dependencies, Semantic relatedness, Ontology, Word gloss, Essence Comparison, Human similarity

I. INTRODUCTION

WordNet is a huge lexical database of English language. All sets of noun, adjectives, verbs and adverbs are joined together into series (sets) of synonyms (synsets), each expressing a unique concept. Synsets are linked with each other by methods for reasonable semantic and lexical relations. WordNet is likewise unreservedly and freely accessible for download and can be used by researchers, scientists, developers to make breakthroughs. WordNet has been used alongside Wikipedia for the process of query expansion [1]. While performing query expansion, we have to focus on one main job of associating one word with its set of synonyms which helps in provisioning a broader perspective of the query. WordNet has also been used with Big Data environment to eradicate plagiarism. It helps in easy detection of cross language plagiarism. In this scenario Arabic and English languages have been studied and detailed analysis have been conducted upon the set of keywords to detect any kind of cross language plagiarism. Wordnet is an important resource when looking into the domain of text analysis based on semantic similarity and relatedness measures. It can be used for multiple purposes including answer sheet evaluation.

Revised Manuscript Received on July 22, 2019.

* Correspondence Author

Suyash Lakhani*, Student of VIT,Vellore, Tamil Nadu, persuing B-tech, CSE department III year. Mail ID- Suyash.lakhani2017@vitstudent.ac.in.

Ridhi Jhamb, Student of VIT,Vellore, Tamil Nadu, persuing B-tech, CSE department III year.Mail ID- ridhi.jhamb2017@vitstudent.ac.in.

Mayank Arora, Student VIT,Vellore, Tamil Nadu, persuing B-tech, CSE department III year. Mail ID- mayank.arora2017@vitstudent.ac.in.

Saravanakumar kandasamy, Teacher at VIT,Vellore, Tamil Nadu, B-tech, CSE department.Mail ID- ksaravanakumar@vit.ac.in

Comparison of wordnet graphs is the simplest way for carrying out such analysis. It has proven to be better than any of the state of art techniques in the domain from statistical analysis. However, there are certain domains which are too sensitive for the similarity calculation using wordnet. Domains like privacy fail at effectively using wordnet to calculate text analysis and similarity calculations. These domains are human view dominated. This paper discusses such domains in detail. Text classification and filtering is a very important application of wordnet in today's date. With increasing size of information available online and concepts like big data and data science coming into play. Approaches to classify and filter all this information is much needed. Wordnet is an effective methodology in this direction. A proposed approach for email classification is discussed in the further sections of the paper. Another major domain which we surveyed was the vast world of Sentiment Analysis. WordNet consisting of a huge lexical database not only helps with query expansion and plagiarism detection but also plays an important role in analyzing what kind of opinions have been passed or given by people. A new approach has also been to devised to generate a similarity core of words and synsets using Set Theory and WordNet. This paper not only helps in guiding the reader about the process of generating a similarity core but also is extremely fundamental for understanding all the papers effectively. In the process of cross language based likeness identification where words boundaries cannot be determined clearly what's more, the convergence of implications of words are fluffy, the fluffy set hypothesis is by all accounts the most suitable approach to treat such case. Fluffy set hypothesis was presented by Lofti Zadeh in 1965 dependent on his numerical hypothesis of fluffy sets. Fluffy set hypothesis could be utilized in a wide scope of spaces particularly for dealing with questionable and loose information that connected with CLPD. Automatic methods for expansion of query were proposed in 1960 by Maron and Kuhns [2]. Modern methods which help in query expansion either imply document collection analysis or are dictionary based.

Rocchio decided to adjudicate by himself some of the retrieved documents and use this feedback information for query expansion [3]. Only the top-most retrieved documents are considered as relevant. The technique is so called pseudo-relevance feedback (PRF). Pseudo-relevant documents are used to look for expansion candidate terms that occur together with many query terms. Another direction for query expansion is the application of word embeddings [4].

II. WORDNET BASED MEASURES

A. Path Length - It is a similarity measure which uses the structure of the wordnet and is said to be the most intuitive out of the four measures. According to this method, the similarity value is the backwards of the briefest way between two words in the wordnet graph.

B. WUP Measure - Similarity based on depths of two words and the depth of their lowest common parent.

C. Leacock and Chodorow measure – This method finds the shortest path between two words and measures it with maximum depth they occur in.

D. Jiang–Conrath - It uses corpora as an ontology. According to the measure, similarity value of two words is the measurement between that child node and its parent. It is found out to be the best among other similarity measures. [34]

E. Hirst– St-Onge - It is a relatedness measure based on the fact that if two words are associated by a way which isn't excessively long and its course isn't changed excessively, at that point the two words have some semantic connection.

F. Word gloss - According to the measure, similarity of two words is proportional to the degree of their gloss overlap. It is not based on the wordnet structure, but more on the wordnet semantics. It's considered to be the best relatedness measure.

G. Information Content based[50] - This is a measure for idea explicitness. In light of the way that the idea that happens all the more much of the time is less explicit and the other way around.

H. Resnik measure - Ascertains the data substance of the most minimal basic parent of the two words. Data content is contrarily corresponding to the likelihood of a term t in a given report containing N unmistakable terms.
IC(t)=-log P(t)
P(t)=frequency (t)/N

III. ADDRESSED DOMAINS

A. EDUCATION

Context identification and text similarity are certain domains under natural language processing which play crucial role in text analysis

- **WordNet graphs for answer sheet evaluation**

WordNet graphs can also be used for answer sheet evaluation for short answer type questions. This can be done by estimating content similitude between the appropriate response gave by the understudy and the perfect answer gave by the instructor. This application will not only help the examiners to focus better and reduce their burden but also make the process impartial. It will help in not only accurate but also time efficient evaluation procedure. Fuzzy WordNet graphs act as the base concept for the study. [26]
According to the methodology, the ideal answer is POS tagged, followed by selection of keywords (all nouns for the sake of simplicity). Wordnet graph consisting of synsets (set of at least one equivalent words that are tradable in some setting without changing reality estimation of the suggestion in which they are inserted) of all synonyms of these keywords is constructed. Another graph is constructed to represent the answer for evaluation using depth first search. WordNet graph generated uses semantic measures like hypernym, hyponym, meronym and holonym.
Presently the closeness between these two charts is figured dependent on common node appearances. i.e. text similarity is calculated proportional to the degree of commonality between the two graphs which is determined directly by the number of common nodes.Marks are given proportional to the ratio of matching nodes to the total number of nodes in the wordnet (similarity measure between the two graphs).
This technology has a lot of future growth scope in many different ways like by amalgamating handwriting recognition in the process. This can be done bygaining a model that learns on the past information dependent on the penmanship of an individual like cosine sentence likeness. [27] Or by building up a calculation that targets using sentence-based outline methods for short answer assessment. [28] A definite examination of different existing procedures for leading short answer type assessment regarding evaluating by methods for correlation and differentiating plots the cons of the current condition of-craftsmanship strategies and features the requirement for such AI based innovations. [29]
Consequently in this AI situated WordNet diagram based strategy, AI and characteristic language handling meet up in the proposed approach and use chart hypothesis, semantic noteworthiness and basic conditions in the wordnet charts to tackle the issue being addressed.

- **Text Summarization**

The successive employments performed by an administrator, when managing an enormous number of archives are rundown of a record and concentrate a confined arrangement of sentences or thoughts, extraction of applicable catchphrases, arranging reports according to pertinence. The assignments don't underline on essential measurable tasks like checking fitting words just, yet more in separating conventional terms. Separating a record watchword set is utilized for ordering of a specific archive, looking through general catchphrases, calculation of semantic separation between the report and a forced arrangement of words so as to rank the report and sort a lot of reports by pertinence. Language abilities and the utilization of a language framework are amazingly urgent for deciding an archive's importance since we will manage unmistakable word structures, association words and potential linguistic and syntactic issues. At whatever point we are thinking about any language framework, we need to allude to its center crucial capacity: to give a type of articulation (composed or represented) an idea and feeling. On the off chance that we talk about the technique wherein words are consolidated into sentences, we definitely allude to language. We can see language structure "as a coin whose different sides are included importance and articulation and whose assignment is to methodicallly join the two". The information on framing appropriate sentences is called fitness, and having the option to absolute them is known as execution. Skilled speakers think that its helpful to change the words and

structure of a sentences since they realize how to precisely utilize the littlest units of punctuation (morphemes).
Linguistic structure is amazingly basic in passing on a message. Records expounded by local speakers can comprise of phonetic style figures and they are the greater part of the occasions difficult to be investigated by a programmed device in light of the fact that the outcomes are a long way from the normal yields. What a word implies is generally characterized by its relationship to different words in its nearby region.
There are non-progressive relations which "fundamentally structure lexical things into the arrangements of equivalent words and different types of resistances." We can separate the three significant various leveled relations which are alluded as scientific categorizations, meronomies and corresponding arrangement. Scientific classifications (likewise alluded to as hyponymy) partner a hyponym (a particular thing) to a superordinate (an increasingly conventional thing or hypernym). This implies some portion of a word's significance isn't just associated with antonymy and synonymy, yet in addition to the way in which it fits into the real jargon chain of command. Things, action words and modifiers can be "grouped" as hyponyms (explicit words) and hypernym or superordinates (nonexclusive words) based on lexical and semantic relations, so would they be able to be concentrated based on pecking order.

- **Plagiarism Detection**

An Ontology is a theoretical model speaking to this present reality, this theoretical model is comprised of officially sorted out arrangement of ideas (or elements), traits, relations and adages. The fundamental bit of leeway of utilizing ontologies is to make conveyance, sharing and reusing of information reasonable. Cosmology coordinating wipes out heterogeneity between information sources and makes them plausible for general purposes which can be gotten to and traded by various applications.
It is a significant activity in building of present day ontology in light of the presence of heterogeneous conditions in which ontologies are made and created, and expected to work. The coordinating is key to empower joint effort and incorporation of frameworks utilizing various ontologies. This is a result of the way that interoperability among ontologies and between frameworks that utilization them, become conceivable just when unmistakable ontologies are brought into shared accord. The fundamental undertaking of coordinating is to create closeness results between two unique components in the information ontologies. All the more absolutely metaphysics coordinating is the way toward harboring ontologies into common understanding by playing out the programmed revelation of coordinating between related ideas. The ontologies themselves are unaffected by the procedure of arrangement. Heterogeneity between ontologies can happen just when various dialects, phrasings, ideas, and displaying are utilized.

- **Extending WordNet with UFO foundational ontology**

Social speculations of lexical semantics hold that any word can be characterized regarding different words to which it is connected. Other than the arrangement of words and the importance, every synset has semantic relationsship that are utilized to interlink them to different synsets, bringing about a thick and critical framework that assistants imparting the semantic data about a particular word in a given setting. These relations are the most huge component given by WordNet and what remembers it from other open lexical databases, which regularly work similarly as thesaurus just as vocabularies. At this moment, WordNet give semantic relations, for instance, super-subordinate, part-whole, troponymy and derivational association, among others.
The semantic relations in WordNet were picked considering the way that they apply widely all through the language and in light of the fact that they are notable, suggesting that a client doesn't must have helped planning in phonetics to get them. Every one of these semantic relations can be spoken to by pointers between word structures or between synsets. The most as often as possible utilized connection among synsets is the very subordinate connection. Things, action words and descriptive words likewise present another particular connection, the derivationally related structure, which centers around interlinking synsets from beginning starting point. A derivational association appears, for example, that an activity word was the source to the course of action of a thing, for instance, the activity word to make is the root for the thing headway.

Mapping Process of ontologies

- Simple mappings - A few guidelines can be made by alluding to the meanings of supersenses and semantic sorts. We consider such mapping a Simple Mapping, which for the most part happens in two circumstances. At the point when both the supersense and the semantic sort have semantically equivalent definitions. This basically target portraying a similar class of ideas and it is conceivable to see that any thing identified with the supersense (by a hypernym connection) can be identified with the predetermined semantic sort. Second circumstance is the point at which the definitions given to the supersense and the semantic sort are not semantically indistinguishable, yet when taken a gander at growing the semantics of the definition, unmistakable from different supersenses, just a particular Semantic Type qualifies in such manner.
- Complex mappings - Inverse to basic mappings, complex mappings are those that don't happen in an immediate way. Here, the supersense can have more than one potential choice of correspondence to semantic kinds. At times it is even conceivable to have a supersense being mapped all the while to two semantic sorts. This kind of mapping can likewise have two structures. The standard characterizing the Semantic sort to be utilized relies upon data other than just definitions, for example, the nearness of specific synsets in the hypernym deduction tree or different sorts of semantic relations accessible on WordNet, for example, the derivational relation.The set of Semantic Types characterized by Dixon doesn't give an alternative able to do totally communicating the Supersense meaning, yet it is conceivable to decide a Semantic Type that communicates its majority and point out other Semantic Types that finishes the portrayal.

- **Query Expansion**

Query expansion (QE) is a notable method used to improve the yield produced by data recovery. Query expansion overhauls the underlying inquiry by adding comparative terms to the first inquiry that help in recovering increasingly applicable outcomes.
The most widely recognized inquiry development system which is delivering a solitary extended question that contains every single extended term. Another strategy that was presented is extending each term each in turn delivering various inquiries, and afterward consolidating these questions results into a solitary outcome list. The points of interest were Handling Generalization, Handling Morphological Variations, Handling Concept Matches, Handling Synonyms with Correct Sense. An elective methodology could be to include "Subcategories", and include Wikipedia "Gloss" when there is no "Redirect" page accessible at that point permit the client either to extend all terms in a solitary question, or to grow each term independently creating different inquiries. The outcome arrangements of these various inquiries are then joined into a solitary outcome list.
A few such methodologies have been proposed in writing delivering very attractive outcomes, yet they are not equitably affable for a wide range of questions (individual and expression inquiries). One of the primary purposes behind this is the utilization of precisely the same sort of assets and weighting plan strategies while extending both the person just as the expression inquiry terms. Thus, the all encompassing relationship among the inquiry terms isn't very much produced or gathered.
To address this issue, a novel methodology has been introduced for inquiry extension utilizing WordNet and Wikipedia as information sources. Wikipedia produces a rich arrangement of extension terms for state terms, while WordNet does likewise for each individual terms. A tale weighting plans has additionally been proposed for development terms: in-interface score (for terms extricated from Wikipedia) and a tf-idf based plan (for terms separated from WordNet). In the proposed Wikipedia-WordNet-based inquiry extension strategy (WWQE), the development terms are gauged twice: first, they are scored by the weighting plan each in turn, and afterward, the weighting plan is liable for scoring the chose development terms concerning the whole question utilizing connection score.
The proposed approach increases enormous upgrades of 24% on the MAP score and 48% on the GMAP score over unexpanded questions on the FIRE dataset. Test results accomplish an important improvement over individual development and other related methodologies. They likewise examined the impact on viability of recovery of the proposed method by changing the quantity of development terms.

- **Religious Research**

Semantic Query for Quranic Ontology

This paper is based on designing the semantic web index for the content of the Quran utilizing Quranic ontology. The sematic engine can search for the Arabic texts and the texts of Quran by the use of meanings as well as words. The Sematic search engine is made using a tool called Apache Lucene.[58,59]
This follows a set of steps which is illustrated as follows:

- Presenting of resources and illustrating the Quranic ontology.
- Describing the search engine system to develop the query processing.

Ontology Conception and classes

There are 6 concepts as listed below:
Semantic idea: It is a class which speaks to all phrasing utilized in the metaphysics.
Semantic Field: Class speaking to all current semantic fields inside the sacred Quran.
General significance: Entitling each semantic field as a worldwide sense.
Comparative Semantic Field: It is a class speaking to the comparable of the semantic field.
Semantic unit: Refers to a class speaking to a solitary word.
Semantic Domain: It is a class speaking to an area.
Creating a semantic ontology - Initial, an assortment all the pre-owned word settings is finished. At that point the importance in each setting wherein it shows up is given. From that point forward, we assemble the words with the equivalent or related implications in one semantic field. At that point Entitle each semantic field as a worldwide importance is entitled. At long last, formation of semantic and coherent connections between the fields' significance is finished.
Ontology evaluation - Assessment of how well the ontology can introduce the word's significance (word development and its relations) by means of semantic examination is done in this progression. Additionally, the testing the ontology with all the expressions of the heavenly Quran is finished.

B. Privacy Measurement

Use of semantic dependencies between words with the help of wordNet to measure privacy leakage. Words might come from various sources but come down to one common sense. This can lead to data leakage. This is also known as the superposition effect[39]. This effect can be studied by derivation of semantic relationships amongst words. This can be done by using existing ontologies like semantic networks, thesauri etc. or by analyzing properties of words from corpora.
In such a situation, knowing the suitability of similarity calculations using wordnet in the domain of privacy becomes trivial. Hence the methodology proposes a comprehensive comparison between the similarity calculations by the wordnet measures against that of the human rating scores, which is considered to be the benchmark dataset. Proposed approach is based on semantic similarity and relatedness. Comparability is a smaller idea between lexical words that have a comparable significance and can be fill in for one another. It depends on the various leveled structure of the wordnet. Though relatedness, is a more extensive idea that have any sort of lexical or practical connections between two words. Alongside progressive structure, it additionally considers some different connections like meronymy, homonymy and so forth.

Human Rating Score

Procedure of differentiating the results and human rating scores is followed, for it can give the best evaluation of the tolerability of a measure since human choices of resemblance and relatedness are respected to be correct. [30] To avoid human bias scientific sampling is used to derive the data from a large dataset. Selection of the subject is also a key factor, as the background they come from affect their privacy mindset.
In the experimental procedure, a questionnaire with 136 word pairs was developed with 17 words based on P3P[35]. As per P3P, protection data is partitioned into five classes. Fundamental individual data, budgetary data, occupation data, address data and other data. Words are classified under these five categories. The subjects assign scores based on the probability that one piece of the information can be deduced, given the other. Score range was defined as 0-4. A total of 150 questionnaires were sent, out of which 32 came out to be valid.
Having such huge dataset, human rating score proved to be better and more relevant than any other benchmark like R&G, M&C etc. in the field of privacy [31,32]
For the comparison analysis of the methodology and the benchmark. The order, especially the pattern created by human rating scores will be contrasted with the outcomes acquired from the WordNet based measures. Correlation of the examples is better than simply comparing numeric values as it highlight essential differences between the two procedures more explicitly. Further Spearman's Correlation and top k ranking algorithm is used to compare the results. Results show that wordnet is not suitable for measuring privacy leakage as it does not coincide with the privacy idea of human ratings [36,37]. Human rating scores can be used to improvise wordnet by narrowing the difference between the similarity values calculated by the two approaches[38].

IV. COMMUNICATION

- **Email Classification**

Filtering emails is a topic of increasing concern. In today's date there are uncountable mails sent everyday without any checking mechanism. This leads toan over-burden to the email servers, and superfluous utilization of system transfer speed and capacity limit. Subsequently utilizing wordnet, a novel way to deal with order spam and ham Emails dependent on the Email body is introduced. It essentially workers the semantic relations and proportions of the WordNet cosmology to diminish the Email highlight semantically[41].
It has two stages which are preparing and testing. The preparation stage comprises of following modules
Pre-processing - It includes cleaning the mail format by tokenization. Irrelevant tokens, symbols like symbols and numbers are expelled. Tokens are removed from both the body and title of the Email. From that point onward, the stop words are dispensed with. To make sure only the meaningful and required words stay, wordnet morphology is used.
Feature Weighting - Weight of every feature is calculated using frequency/inverse document frequency method. Frequency of each term is calculated by the number of times it occurs in the mail. Inverse document frequency measures the importance of a word, by checking how rare its use is in the mail.

$$tf = \frac{f(t)}{\max(f(t))}$$
$$IDF = \log\left(\frac{N}{dt}\right)$$
$$W = tf * IDF$$

dt is the number of Emails,'t', and 'N' is the total number of Emails
f(t) is the Frequency of each term.
Feature Reduction - This module reduces the email extracted features using different reduction techniques [47,48]. One of these is by using semanticbased reduction under which extracted features are replaced by their synsets. These synsets of each term in the Email are used to group the terms that have common synonyms. Hence semantic similarity between words is calculated by applying measures based on wordnet
Features weights updating - To this reduced feature set, new weight is assigned based on semantic similarity values. Updation happens only when the path distance between the two nodes is less than a particular threshold (normalized to be within the range [0.1, 1]) value.
Updated value F=(wi,wj)*(1-dist) , Where wi,wj are initial weights of the features/terms.
Feature selection - This process selects an optimal subset of relevant features. Two different techniques that can be used for the same are:

- Principal component analysis (PCA): Helps diminish the dimensionality of information. It removes significant data from the information and presents them as principal components[52]
- Correlation Feature Selection (CFS): It calculates correlation between subset of features by measuring the interaction with them. Chooses features with high correlation with the class features [53]

Classification - Supervised classification is being used for the proposed system [42]. Class labels are assigned to all data using the learning set. Classifiers used are:

- Naïve Bayes - Based on conditional probability of the features and the class calculated by Baye's formula [43-45].
- SVM – Separates data as spam or ham by creating a hyperplane [46].
- J48 – Based on decision tree. A binary tree is created, at every node of whose, an attribute is selected which splits the instances into groups. Recursively repeats until no more split possible.
- Logistic regression – Predicts association between variables. Followed by grouping these variables to calculate probability of an event [45].
- Random Forest – Decision tree based.
- RBF (radial-basis function) Network – Based on neural networks [40]

Output of this are mails classified as spam or not spam

V. EVALUATION TECHNIQUES

A. Root Mean Square Error

Dataset used is synthetic dataset, hence not very reliable. Dataset consisted of answer sheets of 400 students. These sheets were checked, and their content was changed over into a machine-intelligible organization utilizing OCR (Optical Character Recognition). The sheets were first checked by instructors and important scores were apportioned. Followed by answers in these sheets being examined by the proposed technique and being rethought. The Root Mean Square Error (RMSE) value of the proposed method marks and the actual marks was calculated for comparison analysis.
Approach gives higher accuracy in lesser time in comparison with existing technologies likeIndusMarker [24], Superlative model [25], Fuzzy WordNet diagram based watchword determination [26]. IndusMarker produces the word cloud in a computerized way which an evaluator examinations physically. Proposed strategy straightforwardly produces the wordnet diagram, completes the examination and appoints the score. The consequences of the assessment show that the proposed approach of answer sheet assessment yields promising outcomes as far as exactness and time
. Hence proves to be superior to other existing technologies.

B. Spearman's Correlation

To analyze the consequences of the wordnet measures and the human rating scores. The level of connection between's the two outcomes is dissected utilizing Spearman's Correlation.
Calculate the Spearman's Correlation Coefficient for each of the individual wordnet measure against the human rating scores. This results in 32 coefficients for each word pair. Correlation of the coefficients shows that aftereffects of the WordNet based measures don't agree well with human rating scores. In any case, it tends to be seen that the JC strategy works the best among the four WordNet based similitude and relatedness gauges as it is generally predictable with the human rating scores.

C. Top k ranking

This analysis is directed to outline the level of top-k fortuitous event between the outcomes from the WordNet based measures and from the human rating scores. It means to distinguish the quantity of word combines inside the top-k rank of the WordNet based similitude or relatedness quantifies that fall into the top-k rank of the human rating scores as well). This is called as the essence comparison. It is done for each wordnet measure against the average similarity of all word pairs from the 32 questionnaires of the human rating score. This analysis could be done in two ways. One when we ignore the ordering while taking the top k ranks. The other if we are looking for high precision, by taking ordering into account. This method also shows JC to be the closest to the human rating scores. As it had most common top-k members with human ratings. [30,34]
From the evaluation it can be concluded that the WordNet based measures don't give precise measures on relatedness between words for security concerns. As don't agree with the unprejudiced perspective on people about security. The contrast between the wordnet based measures and the avg. human rating similitude esteems being negative demonstrates that. Be that as it may, the human rating based scores can be utilized to improve the wordnet based measures by fusing important alterations to limit the contrasts between human rating scores and comparability measuresusing wordnet. [36-38]

D. Terrier Retrieval

Varieties of solutions and techniques have been proposed for evaluation purposes involving either Wikipedia or WordNet or both. Every paper proposes a unique solution or an outcome is produced on the basis of previous findings and outcomes.
One of the articles presents a new way of query expansion using Wikipedia and WordNet as standalone entities [5]. In order to expand the initial query generated by the user, they first pre-processed the initial query to extract relevant keywords, terms and phrases.
The acquired expressions and individual terms are later utilized for creation of comparative terms from Wikipedia and WordNet. After this progression, they attempted to relegate the comparability score to these got terms utilizing the proposed weighting score. This is done autonomously for terms from Wikipedia and WordNet. Next, the top n results (which contain significant terms) from Wikipedia and WordNet are short-recorded and chosen as extension terms. These development terms are re-weighted dependent on the connection score. The proposed approach comprises of four fundamental advances:

- Pre-processing of the initial query
- Query Expansion using Wikipedia
- Query Expansion using WordNet
- Re-weighting of the expanded terms

List of datasets used - They used a renowned benchmark dataset Forum for Information Retrieval Evaluation (FIRE) to assess our WWQE approach [6]. FIRE assortments comprises of an enormous measure of records on which the procedure of data recovery is performed [7]. The FIRE dataset comprises of an enormous assortment of articles from newswire from two sources, specifically BDnews24 12 and The Telegraph 13, gave by Indian Statistical Institute Kolkata, India [8]. The above mentioned dataset is used as an evaluation metric because the dataset consists of a myriad of documents on which the process of information retrieval is performed.

E. Similarity score

The measure of information produced each day turns out to be incredibly pivotal with the end goal of examination; this quick increment in the volume of information has made an issue to discover fascinating information among this colossal measure of chaotic content [9]. To defeat this issue, scientists attempt to discover ideal strategies and methods for looking through significant information dependent on spaces, for example, data recovery, content characterization, record grouping, point identification, theme following, questions age, question replying, exposition scoring and so on [10].

The paper proposes an altogether novel methodology and assessment measurements to discover semantic similitude between related words utilizing WordNet and the ideas of Set Theory [11]. In this paper, they proposed another methodology for computing semantic likeness between two unmistakable ideas [12]. The proposed technique depends on ideas of set hypothesis and properties of WordNet, by figuring the relatedness between the shines and synsets of the two concepts.[13]

F. Evaluating Alignment Evaluation Initiative

Ontology coordinating is the procedure that decides correspondences between comparable ideas in two particular ontologies of the comparable space of talk to take care of information heterogeneous issues [20]. They proposed a programmed closeness based coordinating calculation that utilizes practically a wide range of substance portrayals just as their relations to successfully ascertain the correspondences between the two to be coordinated ontologies [21].

The iterative calculation proposed not just goes about as a successful measurement of assessment yet additionally helps in figuring each proportion of comparability independently and afterward joins them in a straight blend to form the last similitude score. The measures utilized arrangement with phonetic, semantic, and auxiliary just as numerous different measures to pick up effectiveness [22].

A typical application in ontology coordinating could be continued as follows: In request for a privately owned business to participate in the commercial center, it needs to decide correspondences between passages of their inventories and sections of a typical inventory of a general commercial center [23]. Having adjusted the indexes, clients of a commercial center have a uniform access to the items, which are discounted. Various coordinating calculations have been proposed in the hypothesis, they all need productivity since they utilize obsolete information sources [24]. In the above framework, they utilized all the information that they could to induce significant data from the plan of the information ontologies, syntactic, etymological, semantic, too are auxiliary [25]. The outcomes are assessed and are put before a few benchmark informational indexes suggested by the OAEI (Ontology Alignment Evaluation Initiative). The outcomes grandstand a colossal improvement over existing framework

G. Domain Specific Extraction

The given calculation begins by gathering a domain explicit phrasing and it at that point finds a pre-characterized set of theoretical connections among the area wordings used to make a space explicit ontology. The assessment of the calculation is finished by utilizing a dataset removed from BibSonomy. The given calculation is planned for lessening normal issues identified with label uncertainty and interchangeable labels which cause a great deal of vagueness on account of ontologies.

Information taken-domain name; Output anticipated area ontology.

This calculation speaks to assets of folksonomy as an undirected weighted chart. Folksonomy can be characterized as a client produced arrangement of isolating and sorting out online substance into various classes by the utilizing metadatas[64]such as electronic labels. The down to top methodology of folksonomies has end up being a fascinating option in contrast to the present exertion at semantic web ontologies in light of the fact that folksonomies give a rich wording produced by huge client networks. Many research contemplates have planned for catching semantics in folksonomies, some of which have prevailing with regards to creating ontologies from folksonomy. The proper explicit area ontology, nonetheless, comprising of space subordinate relations has not been explored at this point.

At that point, it gathers a space phrasing by exploring the assets diagram depending on a lot of area catchphrases which are consequently removed from titles of Wikipedia sections. At long last, the extraction of semantics information[63] about the gathered space phrasing is finished by connecting it to relating Wikipedia passages. This incorporates recognizing the significance, qualities and equivalents of the area phrasing and furthermore finding semantic connections among the domain terminologies.

Table- I: Comparison between various methods for text classification

Sub problems addressed	Email classification	WordNet Ontology	Semantic Similarity	Feature Reduction	Graph Theory/ Set Theory	Machine Learning/ Artificial Intelligence	Text Similarity	Query Expansion	Information retrieval
Extracting domain specific ontology from folksonomy	No	Yes	Yes	No	No	No	No	No	No
A New method for Calculating the Semantic Similarity between the Words Using the WordNet and the Set Theory	No	Yes	Yes	No	Yes	No	No	No	No

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Semantic query for Quranic ontology	No	Yes	Yes	No	No	No	No	No	No
A new approach for the query expansion using Wikipedia and the WordNet	No	No	No	No	No	No	No	Yes	Yes
Extracting keywords using WordNet	No	No	No	No	No	Yes	No	No	Yes
A robust approach to ontology problem	No	Yes	Yes	No	No	No	Yes	No	No
Extending WordNet with UFO foundational ontology	No	Yes	Yes	No	No	No	No	No	No
A Machine Learning Approach for the Automated Evaluation of the Short Answers Using Text Similarity Based on the WordNet Graphs	No	No	No	No	Yes	Yes	Yes	No	No
On the Suitability of Applying WordNet to Privacy Measurement	No	Yes	Yes	No	No	No	Yes	No	No
Efficient email classification approach based on semantic methods	Yes	Yes	Yes	Yes	No	No	No	No	No

Table- 2: Comparison of the base papers based on the given criteria

	Application Domain	**Sub problem addressed**	**Implementation Methods**	**Evaluation Techniques**	**Tools and Dataset used**
Bridging the Gap between the Social and the Semantic Web: Extracting domain specific ontology from the folksonomy	Communication: This can be applied to various domains of social media and user generated tags can be added to ontologies and thereby bridging the gap between social ad semantic web.	Social networking	They have proposed an algorithm which first deals with collection of domain terminologies, then it discovers a pre-defined set of conceptual relationships among domain terminologies. The evaluation of algorithm is done by using a set of data extracted from BibSonomy.	Domain specific extraction	BibSonomy

A New Approach for Calculating Semantic Similarity between Words Using WordNet and Set Theory	Education: Generation of similarity statistic between words using WordNet	Query Expansion	In this paper, they proposed a new approach for calculating semantic similarity between two distinct concepts. The proposed method is based upon concepts of theory of sets and properties of WordNet, by calculating the relatedness between the glosses and synsets of the two concepts.	Similarity Score	--
Semantic query for Quranic ontology	Education: This search engine can be applied to various texts of The Quran and various other related Arabic texts. It also helps in finding the meaning of such texts. (Semantic query expansion)	Religious Research	At first, presenting the resources and illustrates the quranic ontology is done. Second, we describe the system to increase the query. Third refers to including the discussions and the explanations of the search result analysis and finally, summarizing the illustration of the program is done.		Words from The holy Quran
A new approach for query expansion using Wikipedia and WordNet	Education: Google searches, Wikipedia searches, Dbpedia searches	Query expansion	These expansion terms are weighted again based on the scores of corelation.	Terrier Retrieval	FIRE dataset, TREC dataset, CHiC(Cultural Heritage in CLEF) datasets
Extracting keywords using WordNet	Education	Text summarization	By statistically analyzing the entire set of words, on all levels and considering the weights we will have an image about the words of the document	Majority Voting to Obtain Domain Relevancy	Classification and Information filtering, Clustering tool
A robust approach to ontology problem	Education: Ontology matching	Plagiarism Detection	The iterative algorithm calculates each measure of similarity separately and then combines them in a linear combination	Ontology Alignment Evaluation Initiative	Datasets from OAEI campaign, tools for ontology alignment: SMART, PROMPT, PROMPTDIFF
Extending WordNet with UFO foundational ontology	Education: To show its applicability, the proposal was applied to the task of automatically learning a well-found domain specific ontology.	Plagiarism Detection	The extension of WordNet with UFO foundational ontology can be implemented using simple and complex mappings between ontologies and also evaluating WordNet's mapping to UFO	Correctness based on evaluator agreement	--

A Machine Learning Approach for Automated Evaluation of Short Answers Using Text Similarity Based on WordNet Graphs	Education: Answer sheet evaluation based on text similarity.	Automated Answer sheet evaluation	Wordnets for answers to be evaluated are constructed and compared with wordnet of the ideal answer. Text similarity is calculated proportional to the degree of commonality between the two graphs which is determined directly by the number of common nodes.	Root mean square error calculation	Synthetic Data set, Natural Language Tool Kit (NLTK) libraries of python.
On the Suitability of Applying WordNet to Privacy Measurement	Privacy Measurement	To what extent can the Similarity measures of WordNet be applied in the domain of privacy.	Comparison is done using two methodologies. Spearman's Correlation and top k ranking algorithm. Spearman's correlation consists of calculating the correlation values for each word pair from the wordnet based measures and human rating scores. Top k ranking appears to be more efficient as it simply compares the rank vectors of all the wordnet based measures with those of the human rating and give accurate results.	Spearman's Correlation, Top k ranking	Human rating score as benchmark dataset, WordNet::Similarity is a free software tool of perl for relatedness and similarity calculations. Java WordNet Similarity (JWS)

VI. CONCLUSIONS

The above base papers effectively discuss the implementation methods and different approaches carried by multiple authors which solve problems in the field of Natural Language Processing using WordNet. These papers not only distinguish between previously existing techniques and methods but also provide detailed reasons about their approaches and why these new approaches are better. WordNet being a huge domain can be used to tackle majority of problems in NLP. It can solve simple problems ranging from generation of a similarity core to more complex problems like cross-language plagiarism detection and query expansion. There are plenty of more problems which can be tackled using WordNet, however the above papers were the most relevant for our survey purposes as they not only span across multiple domains but also provide a in-depth analysis of the problem worked upon.

In the above papers, we present an assortment of ideas for closeness coordinating strategy dependent on data content utilizing the first chain of importance of WordNet. The outcomes delivered can be utilized for portraying the similitude measures among words. We can utilize this idea of semantic closeness for supplanting unique unaltered inquiry with set of equivalent words based on comparability score which can without a doubt upgrade the data recovery (IR) task. Clients much of the time neglect to portray the data they precisely need to recover in the inquiry question which can prompt creation of pointless and obscure yields.

A social examination performed on a lot of English action words has uncovered a portion of the striking manners by which action words contrast from things and modifiers all in all. The connection between action words are not the same as those between words having different grammatical features. All in all, their semantics are extensively more frustrate. The power of unmistakable relations and lexicalization designs in different semantic domains has likewise been talked about.
Wordnet could be improved by inclusion of technical terms so that it can be used more in the technical and engineering domains. It can be made to consider other relationships other than synonyms, hyponyms etc. such that similarity values can be calculated irrespective of spelling errors also. Wordnet based similarity calculation measures can be made more accurate by wordnet joining essential alterations to limit the contrasts between human rating scores and comparability gauges about the relatedness of word sets this can be done by adding direct paths between words otherwise not related.
Aside from these talked about domains, WordNet can likewise be utilized in assortment of different domains to take care of heap of issues like tackling the issue of cross language unoriginality utilizing keen multilingual literary theft identification utilizing Big information innovations (Hadoop, HDFS and MapReduce). Information content and the assortment corpus can be made up from two inaccessible dialects like Arabic and English, the production of an appropriate objective information of each report is simple, numerous content preprocessing methods defined on common language handling can likewise be actualized for future upgrades. Notwithstanding this group arrangement can be tried alongside the conduct of the current semantic methodology for various datasets.

VII. FUTURE WORK

For future work, another domain of sentimental analysis can also be kept under consideration. The domain has plenty of scope of improvement and WordNet can improve the quality of projects tremendously. The desired solution can be obtained by usingpreparing information improvement of NB learning for IAT recognizable proof by advancing the up-and-comer verifiable angle list with a lot of removed words from WordNet relations, principally equivalent and definition. These examinations can be reached out to the semantic coordinating methodology by processing semantic comparability among various existing ontologies. The methodologies introduced here can be additionally improved exponentially with fusing Word Sense Disambiguation (WSD). With the processed comparability, WSD can be performed by utilizing relatedness for the age of the ideas compulsory for the question extension module

ACKNOWLEGEMENTS

The authors wish to thank other members of the Natural language processing groups at Vellore institute of technology for their help throughout the course of this work.

REFERENCES

1. Azad, H. K., & Deepak, A. (2019). A new approach for query expansion using Wikipedia and WordNet. Information Sciences, 492, 147-163.
2. Maron, M. E., & Kuhns, J. L. (1960). On relevance, probabilistic indexing and information retrieval. Journal of the ACM (JACM), 7(3), 216-244.
3. Rocchio, J. (1971). Relevance feedback in information retrieval. The Smart retrieval system-experiments in automatic document processing, 313-323.
4. Bhogal, J., MacFarlane, A., & Smith, P. (2007). A review of ontology based query expansion. Information processing & management, 43(4), 866-886.
5. Carpineto and Romano reviewed major QE techniques, data sources and features in an information retrieval system.
6. http://fire.irsi.res.in/fire/static/data%20.
7. https://bdnews24.com/%20.
8. https://bdnews24.com/%20
9. EZZIKOURI, H., MADANI, Y., ERRITALI, M., & OUKESSOU, M. (2019). A New Approach for Calculating Semantic Similarity between Words Using WordNet and Set Theory. Procedia Computer Science, 151, 1261-1265.
10. Youness, M., Mohammed, E., & Jamaa, B. (2018). Semantic Indexing of a Corpus. International Journal of Grid and Distributed Computing, 11(7), 63-80.
11. Leacock, C., & Chodorow, M. (1998). Combining local context and WordNet similarity for word sense identification. WordNet: An electronic lexical database, 49(2), 265-283.
12. Gupta, D., Vani, K., & Singh, C. K. (2014, September). Using Natural Language Processing techniques and fuzzy-semantic similarity for automatic external plagiarism detection. In 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI) (pp. 2694-2699). IEEE.
13. Bollegala, D., Matsuo, Y., & Ishizuka, M. (2007). Measuring semantic similarity between words using web search engines. www, 7, 757-766.
14. Jyoti Yadav, Yogesh Kumar Meena, Use of fuzzy logic and wordnet for improving performance of extractive automatic text summarization, ICACCI 2016, 21-24 Sept. 2016, DOI: 10.1109/ICACCI.2016.7732356
15. J. C. Lee, Yu-N Cheah, Paraphrase detection using semantic relatedness based on Synset Shortest Path in WordNet, Advanced Informatics: Concepts, Theory And Application (ICAICTA), 2016 Int. Conf. On, ISBN: 978-1-5090-1636-5
16. Mohamed Ben Aouicha, Mohamed Ali Hadj Taieb, Sameh Beyaoui, Distributional semantics study using the co-occurrence computed from collaborative resources and WordNet, INISTA, 2016, DOI: 10.1109/INISTA.2016.7571831
17. Saint-Dizier, P. (ed.) Predicative Forms in Natural Language and in Lexical Knowledge Bases, Springer-Science+Business Media Dordrecht, B.V., 1999.
18. Sneha S. Desai, J. A. Laxminarayana, WordNet and Semantic similarity based approach for document clustering, CSITSS, 6-8 Oct. 2016.
19. Steve Legrand, 2006, Word Sense Disambiguation with Basic-Level Categories, Advances in NLP Research in Computing Science 18, 2006, pp. 71-82
20. Steve Legrand, JRG Pulido, 2004, A Hybrid Approach to Word Sense Disambiguation: Neural Clustering with Class Labeling. ECML and PKDD Pisa, Italy September 24, 2004.
21. Ehrig, M. et Stabb, S. QOM□quick ontology mapping. In : The Semantic Web□ISWC 2004. Springer Berlin Heidelberg, 2004. p. 683-697.
22. Euzenat, J. et Valtchev, P. (2004). Similarity-based ontology alignment in OWL Lite. In Proceedings of the 15th European conference on Artificial Intelligence (ECAI), pages 333-337, Valence, Spain.
23. Euzenat, J. (2007). Semantic precision and recall for ontology alignment evaluation. In Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI' 2007), pages 348-353, Hyderabad, India.
24. Euzenat, J. et Shvaiko, P. (2013). Ontology Matching. Springer-Verlag Berlin Heidelberg
25. Euzenat, J., et al. Results of the Ontology Alignment Evaluation Initiative 2009. In Proceedings of the 4th International Workshop on Ontology Matching (OM-2009). Vol. 551.
26. Vii, S., Tayal, D., & Jain, A. (2019). A fuzzy WordNet graph based approach to fnd key terms for students short answer evaluation. In 2019 4th international conference on internet of things: Smart innovation and usages (IoT-SIU) (pp 1–6). IEEE
27. Sijimol, P. J., & Varghese, S. M. (2018). Handwritten short answer evaluation system (HSAES).
28. Van Hoecke, O. D. C. S. (2019). Summarization evaluation meets short-answer grading. In Proceedings of the 8th workshop on NLP for computer assisted language learning, pp. 79–85
29. Roy, S., Rajkumar, A., & Narahari, Y. (2018). Selection of automatic short answer grading techniques using contextual bandits for diferent evaluation measures. International Journal of Advances in Engineering Sciences and Applied Mathematics, 10(1), 105–113.
30. Budanitsky, A., & Hirst, G. (2001). Semantic distance in WordNet: An experimental, applicationoriented evaluation of five measures. In Proceedings of the workshop on WordNet and other lexical resources, 2nd meeting of the North American chapter of the association for computational linguistics (Vol. 2(12), pp. 29–34). Pittsburgh, PA.
31. Gao, J. B., Zhang, B. W., & Chen, X. H. (2015). A WordNet-based semantic similarity measurement combining edge-counting and information content theory. Engineering Applications of Artificial Intelligence, 39, 80–88.

32. Han, L. S., Finin, T., McNamee, P. L., Joshi, A., & Yesha, Y. (2013). Improving word similarity by augmenting PMI with estimates of word polysemy. IEEE Transactions on Knowledge and Data Engineering, 25(6), 1307–1322.
33. Patwardhan, S., & Pedersen, T. (2006). Using WordNet-based context vectors to estimate the semantic relatedness of concepts. In Proceedings of the EACL 2006 workshop on making sense of sense: Bringing computational linguistics and psycholinguistics together (pp. 1–8). Trento.
34. Budanitsky, A., & Hirst, G. (2006). Evaluating WordNet-based measures of lexical semantic relatedness. Computational Linguistics, 32(1), 13–47.
35. Cranor, L., Langheinrich, M., Marchiori, M., Presler-Marshall, M., & Reagle, J. (2002). The platform for privacy preferences 1.0 (p3p 1.0) specification. W3C Recommendation.
36. Maalej, M., Mtibaa, A., & Gargouri, F. (2015). Enriching user model ontology for handicraft domain by FOAF. In Proceedings of the 2015 IEEE/ACIS 14th international conference on computer and information science (pp. 651–655). Las Vegas, NV.
37. Ell, B., Hakimov, S., & Cimiano, P. (2016). Statistical induction of coupled domain/range restrictions from RDF knowledge bases. In Proceedings of the 15th international semantic web conference, lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics) (#10579, pp. 27–40). Kobe.
38. Jilek, C., Maus, H., Schwarz, S., & Dengel, A. (2015). Diary generation from personal information models to support contextual remembering and reminiscence. In Proceedings of the 2015 IEEE international conference on multimedia and expo workshops. Turin.
39. Terzi, D. S., Terzi, R., & Sagiroglu, S. (2015). A survey on security and privacy issues in big data. In Proceedings of the 2015 10th international conference on internet technology and secured transactions (pp. 202–207). London.
40. Del Castillo M, Dolores, Ignacio Serrano J. An interactive hybrid system for identifying and filtering unsolicited e-mail. Intelligent Data Engineering and Automated Learning–IDEAL. Springer, Berlin Heidelberg; 2006. p. 779–88.
41. Hristea FT. Semantic wordnet-based feature selection. In: The Naïve Bayes model for unsupervised word sense disambiguation. Heidelberg: Springer, Berlin; 2013. p. 17–33.
42. Khan Aurangzeb et al. A review of machine learning algorithms for textdocuments classification. J Adv Inform Technol 2010;1(1):4–20.
43. Islam MS, Al Mahmud A, Islam MR. Machine learning approaches for modeling spammer behavior. Information Retrieval Technology. Heidelberg: Springer Berlin; 2010. p. 251–60.
44. Blanzieri E, Bryl A. A survey of learning-based techniques of email spam filtering. Tech. rep. DIT-06-056, University of Trento, Information Engineering and Computer Science Department; 2008.
45. Mitchell T. Generative and discriminative classifiers: naive Bayes and logistic regression. Manuscript available at http://www.cs.cm.edu/~tom/NewChapters. html; 2005.
46. Islam Rafiqul, Yang Xiang. Email classification using data reduction method. In: Proceedings of the 5th international ICST conference on communications and networking in China. IEEE; 2010. p. 1–5.
47. Bahgat EM, Rady S, Gad W. An E-mail filtering approach using classification techniques. In: The 1st international conference on advanced intelligent system and informatics, Beni Suef, Egypt. Springer International Publishing; 2016. p. 321–31.
48. Bahgat Eman M, Moawad Ibrahim F. Semantic-based feature reduction approach for e-mail classification. International conference on advanced intelligent systems and informatics. Springer International Publishing; 2016.
49. Kolhatkar Varada. An extended analysis of a method of all words sense disambiguation Diss. University of Minnesota; 2009.
50. Pedersen Ted, Patwardhan Siddharth, Michelizzi Jason. WordNet:: Similarity: measuring the relatedness of concepts.'' Demonstration papers at HLT-NAACL 2004. Association for Computational Linguistics; 2004.
51. Slimani Thabet. Description and evaluation of semantic similarity measures approaches. arXiv preprint arXiv:1310.8059; 2013.
52. Sharma Amit Kumar, Yadav Renuka. Spam mails filtering using different classifiers with feature selection and reduction technique. 2015 Fifth international conference on communication systems and network technologies (CSNT). IEEE; 2015
53. Karegowda Asha Gowda, Manjunath AS, Jayaram MA. Comparative study of attribute selection using gain ratio and correlation based feature selection. Int J Inform Technol Knowledge Manage 2010;2(2):271-7.
54. Enron-Spam datasets. CSMINING group [accessed July 7, 2016] http:// csmining.org/index.php/enron-spam-datasets.html.
55. Shams Reza, Mercer Robert E. Classifying spam emails using text and readability features. 13th international conference on data mining (ICDM). IEEE; 2013.
56. Siddiqi, R., Harrison, C. J., & Siddiqi, R. (2010). Improving teaching and learning through automated short-answer marking. IEEE Transactions on Learning Technologies, 3(3), 237–249.
57. Jayashankar, S., & Sridaran, R. (2017). Superlative model using word cloud for short answers evaluation in eLearning. Education and Information Technologies, 22(5), 2383–2402.
58. Apache, S., 2016. Ultra fast search library lucene 2016. Available from: http://jakarta.apache.org/lucene.
59. Mahgoub, Ashraf Y., Rashwan, Mohsen A., Raafat, Hazem, Zahran, Mohamed A.,Fayek, Magda B., 2014. ''Semantic query expansion for arabic information retrieval 2014 Cairo, Egypt
60. Alruqimi, M., Aknin, N., 2015. Semantic emergence from social tagging systems. Int. J. Org. Collective Intelligence 5 (1), 16–31.
61. Font, F., Serrà, J., Serra, X., 2015. Analysis of the impact of a tag recommendation system in a real-world folksonomy. ACM Trans. Intelligent Syst. Technol. https://doi.org/10.1145/2743026.
62. Uddin, M.N., Duong, T.H., Nguyen, N.T., Qi, X.M., Jo, G.S., 2013. Semantic similarity measures for enhancing information retrieval in folksonomies. Expert Syst. Appl. https://doi.org/10.1016/j.eswa.2012.09.006.
63. Wang, S., Wang, W., Zhuang, Y., Fei, X., 2015. An ontology evolution method based on folksonomy. J. Appl. Res. Technol. 13 (2), 177–187. https://doi.org/10.1016/J.
64. Mathes, A. (2004). Folksonomies - Cooperative Classification and Communication Through Shared Metadata. Retrieved July 2, from http://www.adammathes.com/academic/computer-mediated
65. Font, F., Serrà, J., Serra, X., 2015. Analysis of the impact of a tag recommendation system in a real-world folksonomy. ACM Trans. Intelligent Syst. Technol. https://doi.org/10.1145/2743026.
66. S. Hertling, H. Paulheim, WikiMatch - using wikipedia for ontology matching, in: Proceedings of the 11th International Semantic Web Conference 2012, ISWC'12, Boston, USA, 2012.

67. F.C. Schadd, N. Roos, Coupling of WordNet entries for ontology mapping using virtual documents, in: Proceedings of the 11th International Semantic Web Conference 2012, ISWC'12, Boston, USA, 2012
68. M. Ciaramita, M. Johnson, Supersense tagging of unknown nouns in wordnet, in: Proceedings of the 2003 conference on Empirical Methods in Natural Language Processing, EMNLP, 2003, pp. 168–175.
69. N.F. Padilha, F.A. Baião, K. Revoredo, Ontology alignment for semantic data integration through foundational ontologies, in: Advances in Conceptual Modeling - ER 2012 MORE-BI Workshop, 2012, pp. 172–181.
70. D. Nadoveza, D. Kiritsis, Ontology-based approach for context modeling in enterprise applications, Special Issue on The Role of Ontologies in Future Web-based Industrial Enterprises, Comput. Ind. 65 (9) (2014) 1218–1231.
71. C.D. Manning, P. Raghavan, Introduction to Information Retrieval, first ed., Cambridge University Press, New York, 2008

AUTHORS PROFILE

Suyash Lakhani: Student of VIT,Vellore, Tamil Nadu, persuing B-tech, CSE department III year.
Mail ID- Suyash.lakhani2017@vitstudent.ac.in.

Ridhi Jhamb: Student of VIT,Vellore, Tamil Nadu, persuing B-tech, CSE department III year.
Mail ID- ridhi.jhamb2017@vitstudent.ac.in.

Mayank Arora: Student VIT,Vellore, Tamil Nadu, persuing B-tech, CSE department III year.
Mail ID- mayank.arora2017@vitstudent.ac.in.

Saravanakumar kandasamy: Teacher at VIT,Vellore, Tamil Nadu, B-tech, CSE department.
Mail ID- ksaravanakumar@vit.ac.in