The Nucleus **54**, No. 4 (2017) 258-265

www.thenucleuspak.org.pk

The Nucleus

ISSN 0029-5698 (Print)
ISSN 2306-6539 (Online)

Statistical Analysis Approach to Determine Language Quality of Students' Responses Using WorldNet Similarity Techniques

Farhan Ullah[1,2], M. Farhan[1*], K. R. Malik[1], M. M. Iqbal[3], M. Ibrar[1] and Z. Rahman[2]

[1]*Department of Computer Science, COMSATS Institute of Information Technology, Sahiwal, Pakistan*

[2] *College of Computer Science, Sichuan University, 610064, Chengdu, China*

[3] *Department of Computer Science and Engineering, University of Engineering and Technology, Taxila, Pakistan*

ARTICLE INFO

Article history :
Received : 21 April, 2017
Accepted : 17 January, 2018
Published : 31 January, 2018

Keywords:
E-Assessment
Electronic Learning
Machine Learning
WordNet
Technology Enhanced Assessment
Semantic Similarity

ABSTRACT

This study aims to electronically assess (e-assessment) students' replies in response to teachers' question. It can be useful to systematize the question answering context regarding matching text semantically through WordNet semantic similarity techniques. WordNet is a lexical database of words' synonyms. It uses group of synonyms called synsets for semantical operation of English text. For this purpose, a new methodology is proposed to automate e-assessment in the field of education. The collected dataset contains 210 pairs of words extracted from different undergraduate students' replies in contradiction of teacher's question statement. Further WordNet similarity measures i.e. Path Length. Moreover, Lin, Wu & Palmer and Hirst & St-Onge are used to compute the semantic relatedness score. In the pilot study 42 pair of words were extracted from 8 students' replies, which are marked using semantic similarity measures and equated with teacher's marks. Teachers are provided with four boxes of the mark while our developed method provides a precise measure of marks. The experiment is shown with comprehensive dataset resulting with words' frequencies in similarity measures.

1. Introduction

E-assessment and electronic learning (e-learning) is becoming more common in the last few years, but effectiveness and practicality regarding students' e-assessment are not well understood. Teacher uploads a query online in e-learning schooling, and then students give replies in response to teacher's query. There is a necessity to process all these responses and rank the most related reply automatically regarding semantically similar keywords. The traditional technique is, that teacher reads all replies manually and then assigns marks to related reply physically. There is a need for ane-assessment method to read all replies automatically using machine learning methods and assign marks to related replies in standings of semantics. WordNet thesaurus is used to match keywords by synonyms or semantic similarity functions. With the quick development of the internet, there is a need to evaluate a massive amount of data and extract meaningful text in response to user's query. Text similarity plays a vital role in text mining and information retrieval to afford related text. It can be functional in different tasks such as plagiarism detection, question answering, paraphrase identification, textual entailment [1, 2] and so forth.

In Natural Processing Language (NLP), several types of text represent objects, and these objects' structures are changed to numbers through Vector Space Model (VSM) with Term Frequency-Inverse Document Frequency (TF-IDF) value. However, these methods emphasize mainly keyword-based similarity and ignore the semantic relationships between texts. To capture the semantic relations between words using WordNet [1], or some other semantic catalog is used. In these dictionaries, all words with complete information like root word, synonyms details, tagging information are there. Interpreted corpus is used to translate the keywords in standings of semantics. Answer selection in question answering context is an open research task to rank the related reply to the user. For example, yahoo answer in which a user input a query and then professionals give responses after that the user manually searches for related answers. Similarly, blogs and public media on which different users from the world provide information or newsflash about various matters need to be gone through to determine which information/newsflash is more related to us? Again, the user searches the related information. Although the set of documents, which are repossessed by the search engine comprise much information about the specified topic. There are lots of search engines accessible such as Google, Yahoo, Bing and many others. All these search engines have no doubt many important proficiencies but the problem with these search engines is that they do not have assumption capabilities.

[*]Corresponding author : farhansajid@gmail.com

The main problem is that in question answering framework the input question and reply responses have no syntactic relation among words in a sentence. The authors [3] proposed *sn-grams* from the text that there is a syntactic relationship among neighbors of words from syntactic trees and not from the text directly. If we compute syntactic relations of input question and syntactic relations of dissimilar answers from professionals, then after computing resemblance the most user related answer can rank on top. In this paper a new methodology has been proposed for e-assessment in students' answers in response to teacher's question. Teacher posts a question on LMS (Learning Management system) and students from different areas gave answers online. Then the proposed methodology is used to find the most similar answer based on WordNet similarity techniques.

The paper consists of related work in section 2, Material in section 3, methods in section 4, Results, and discussion in section 5, Experimental in section 6 and Conclusion in section 6.

2. Related Work

Luaces et al. [4], defined Massive Open Online Courses (MOOCs) to assess students with particular knowledge requirements. Automatic e-Assessment is completed with a huge number of projects through multiple choice queries. Vector Space Model (VSM) method is used to provide similar text and further to match replies.

Monolingual and cross-lingual semantic textual similarity in e-assessment is a big challenge in question answering context. Agirre et al. [5] presented a question answering opportunity set in which text similarity is done in English lingual data and cross-lingual Spanish data. The authorsconsidered semantic similarity in snippet text sets.

Kastner et al. [6] applied e-assessment to select the most related answer in research question answering context by complex suggestion. Synthesis method was used to rank the alike text in answerto a precise question. Author anticipated an abstract algorithm to synthesize the optimum knowledge of a research query.

Kang et al. [7] proposed e-assessment of 6-12 years, Korean children, to evaluate the value of life with sensitive rhinitis. The number of students was 277 from middle schools. These students were separated into three groups; allergic-rhinitis (AR), non-allergic rhinitis (non-AR), and controls.

E-assessment can be assisted by multiple choice questions or short questions answers that are easy to understand. Burrows et al. [8] projected an idea of automatic short answer grading (ASAG) in e-assessment. The authors proposed that short answer question should consider outside information, query response, answer length and should emphasize on text content.

Kim et al. [9] planned e-assessment in speech act analysis in web media. Online debates were shown through a set of speech act patterns. It defined how different speech patterns identify conversation threads and how they smooth automatic question answering context.

Knowledge-based information retrieval plays a significant role in e-assessment in question answering framework. Otegi et al. [10] did e-assessment in question answering through text extension in the text by WordNet ontology. The pseudo-relevance feedback (PRF) used for query extension is better for simple questions and the projected model has robust results for tough queries. The author showed good results using Wikipedia as acknowledge-based data structure.

Partalas et al. [11] evaluated the text classification in question answering and proposed LSHTC (Large Scale Hierarchal Text Classification) for a huge number of programs in a dataset. Corpora of Wikipedia and web encyclopedia were used to evaluate the hierarchal text arrangement. The drill dataset of LSHTC is accessible online and may be downloaded for more experimentation.

Cigdem and Oncu [12] proposed a TAM2 model through which an e-assessment methodology is investigated among student through e-quizzes The students answers' are assessed according to their age. Results showed that behavioral intention considerably increased the perception of question's content.

Cosine text similarity using Vector Space Model has been broadly used and researched to rank related text in a document. Chen and Durme [13] proposed a model that inputs text into a context sensitive environment and then finds similarities in text pairs. They used an unsupervised approach which ranks similar text using cosine similarity measure.

In the last few years, Tree Edit Distance (TED) became more widely used to find similarities in dissimilar documents. Sidorov et al. [2] defined the application of TED to treasure similar text between syntactic n-grams to calculate the soft similarity between documents. Syntactic n-grams are non-linear tree structure, and TED is the accurate measure to recover similar text. Many other authors have approximately explained the applications of TED in different situations for similar mining text [14].

In Natural Language Processing (NLP) text similarity plays a vital role like question answering, entity disambiguation, author attribution and so on. Sidorov et al. [1], presented an idea to use soft cosine similarity measure using VSM between syntactic n-grams. In earlier years words and n-grams were used for VSM to find similar text but the authors have presented syntactic n-grams which is based on parse trees. There will be a syntactic relation amongst words which will be further managed in VSM.

The authors also used machine learning algorithms for the transformation of VSM to find similarity.

In the current research, question answering is a hot zone in NLP to retrieve the utmost related response. Tymoshenko et al. [15] accomplished kernel based classifier on SemEval dataset and then their scores were used to rank the interrelated answer on top. They used intra-pair comparisons and inter- pair kernel method to rank the associated answer.

Goyal et al. [16] labeled K-means cluster method to cluster papers to find similar text. Furthermore, the author used cosine and fuzzy procedures to draw an assessment in accurately. Text similarity methods are more significant for academic papers and patents and are used to find similar ideas and rank the status of the text.

Xu et al. [17] linked four similarity measure (Latent Semantic Analysis) LSA based on words, LSA based on terms, VSM based on words and VSM based on terms to find similarity between academic papers and patents. The author also presented that term based VSM measure gave more precise outcome than others.

Jiang et al. [18] proposed an idea in recurrent question answering in the Chinese language. Pair wise objects were presented in VSM in different dimensions. Authors used unigram and bigram to find similar text in documents vectors. The author used shallow lexical semantics and document length information to monitor the the performance of VSM.

In a period of big data, understanding of huge dataset is a hot topic. Although there are several submissions developed to crack the problem, but still, it is an open research challenge. Recently a calculating clustering platform, Spark, has been proposed for large scale data dispensation and big data analytics. Bao et al. [19] used VSM and TF-IDF (term frequency-inverse document frequency) to extract structures from objects and used hashing methods to extract similar text from large scale data processing by Spark.

Mihalcea et al. [20] targeted the semantic similarity between small text (abstract of technical articles, tweets, and blogs) using corpus-based and information-based similarity. They defined that semantic similarity using a bag of words approach was significantly better than traditional lexical matching. However, a bag of words approaches disregards dependency structure between words in a sentence. Semantic parse tree will give better outcome in text similarity. Text similarity in plagiarism finding’s one of the core are as in NLP.

Acree [21] used cosine similarity, substring, and weighted cosine similarity to measure plagiarized text in political data. They proposed that all these techniques are not useful for detecting substantive difference between texts however these methods are more useful for data processing to save similar text.

Contents extracting from question and answer are a critical task to understand the global meaning of small text. Zhang et al. [22] used statistically based similarity measure to determine the structure of the text semantically to improve the performance of “ Wh” question in question answering structure.

3. Experimental

Statistical Package for the Social Sciences (SPSS) was used to compute statistical measures of the whole dataset. Here 210 pairs of words dataset were examined in SPSS to extract meaningful statistic measures and frequency dispersal of Path Length, Lin, Wu & Palmer, and Hirst & St-Onge. Fifty different students’ answers are examined as shown in Table 4. It was investigated that mean, median and mode of Wu & Palmer measure gave improved results as related to other measures. The variance and standard deviation are applied on the dataset. It was observed that different students’ answers have different statistical values. Frequency graph of each similarity measure against assigned marks are shown in Figs. 3(a), 3(b), 3(c) and 3(d). In Figs. 3(a) and 3(b), it is shown that Path Length and Lin measure calculate 1.5 to 2 marks to most replies. In Fig. 3(c) it is shown that Wu & Palmer calculate 3 to 4.5 marks to most replies which proved that it calculates the highest semantic similarity measures between a pair of words.

Table 4: Descriptive statistics for the analysis of 50 students’ replies

	Path Length [25]	Lin [26]	Wu & Palmer [27]	Hirst & St-Onge [28]
Valid	50.00	50.00	50.00	50.00
Missing	0.00	0.00	0.00	0.00
Mean	2.07	1.98	3.63	1.60
Std. Error of Mean	0.05	0.08	0.07	0.07
Median	1.95	1.83	3.57	1.47
Mode	1.77	1.72	3.26	1.47
Std. Deviation	0.38	0.59	0.52	0.51
Variance	0.15	0.35	0.27	0.26
Range	1.76	2.41	2.28	1.88
Sum	103.58	98.86	181.56	80.00

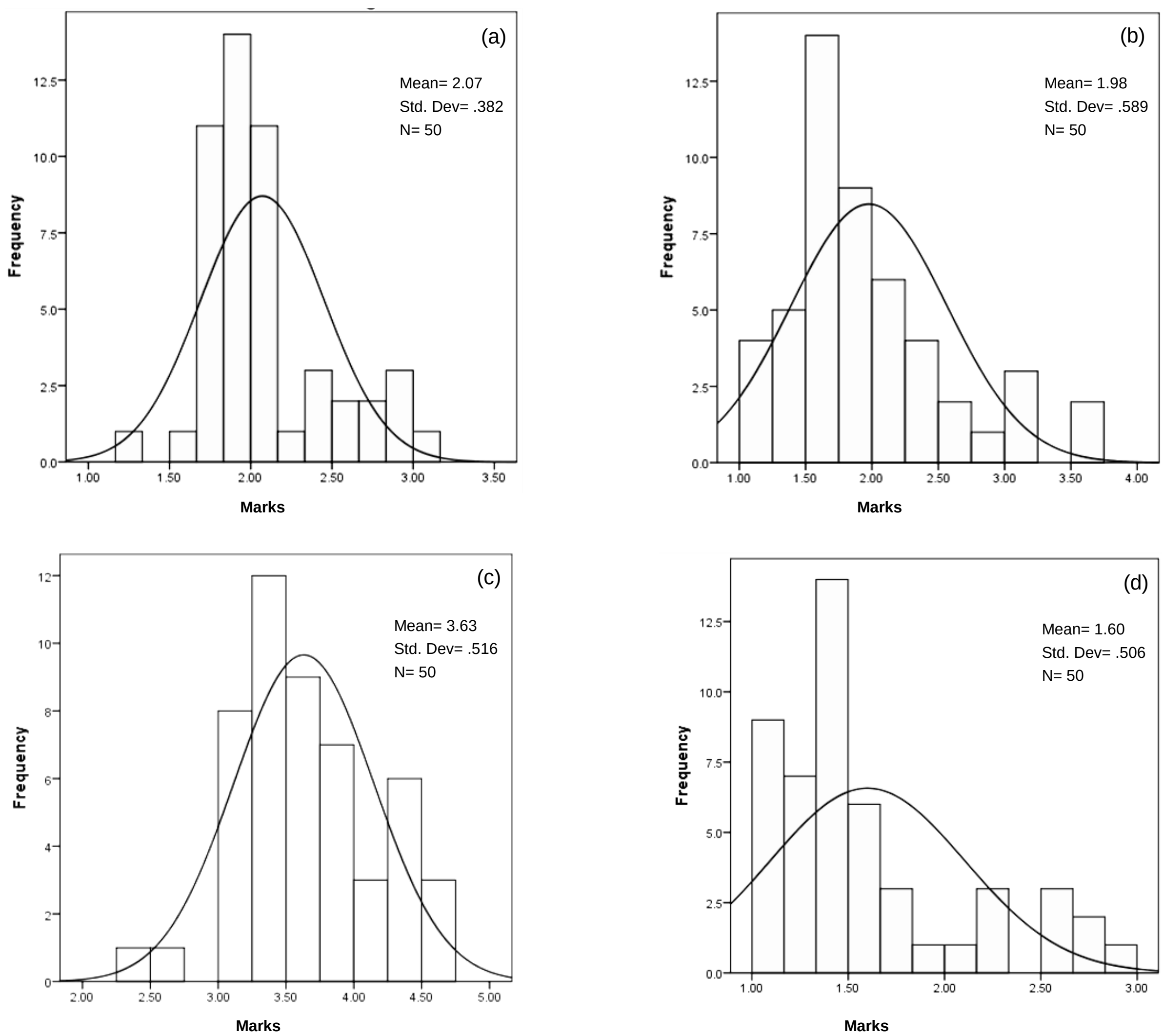

Fig. 3: Frequency distribution of marks using WordNet similarity techniques (a) Path length, (b) Lin,, (c) Wu & Palmer and (d) Hirst & St-Onge

4. Material and Methods

Text similarity is an elementary and vital task in NLP because the user is attracted to see the matching or most related text. WordNet [1], or some further semantic database is used to capture semantic relations between words. Delen [23], defined the student's behaviour in computer based training and testing in two different circumstances and then compared the resulting scores. It was shown in this research that, we could improve the results of the computer-based test if the student has an optimal response time. Case study was carried out to evaluate undergraduate students in Open Source Web Application Development (PHP, PERL, CGI and MYSQL) course from Virtual University of Pakistan (VU). Case study was divided into diverse phases for semantics and marks ranking scrutiny.

4.1 Methodology

Semantic similarity is restrained between teacher's question and student answers to rank the maximum related reply automatically. De Boni and Manandhar [24], proposed an algorithm for semantic similarity between sentences using different linguistic evidence in query responding context. WordNet semantic similarity methods have been used to measure the semantic relatedness between scripts. The first teacher uploaded a query in an undergraduate course on LMS page of VU. In this system 50 students answered in response to the query. The teacher also allocated marks in Teachers' score table to students' answers manually in a series of 1 to 5 (1.25, 2.50, 3.75, 5.00). Keywords have been mined from teacher's query to allocate marks automatically in standings of semantic relatedness; Students' answers to translate these to Question Keywords Vector (QKV) and Students' Answers Vector (SRV). Then WordNet semantic similarity methods, i.e., Path Length [25], Lin [26], Wu & Palmer [27] and Hirst & St-Onge [28] were employed to measure the similarity scores which were stored in a table called grading score table. Path Length, Lin, Wu & Palmer provided a score in the series of 0 to 1, and Hirst & St-Onge gave a score in the series of 0 to 16. Normalization technique was used to gauge the semantic relatedness score in each series using Eq. (1).

$$Nval = (Cval / SM_{\max}) + SM_{\min} \qquad (1)$$

Where Nval signifies normalized value, Cval is Calculated Value by Similarity Measures, SM_{max} Similarity Measure maximum range and SM_{min} Similarity Measure minimum range.

The teacher allocates marks in teacher's score table manually to all student's answers. The WordNet semantic similarity methods allocate marks automatically to students' replies. Semantic measure score table is calculated using Eq. (2).

$$ScoreTable = \sum_{i=1}^{n} \sum_{k=1}^{k=4} ScoreTable_{i,k} \qquad (2)$$

The assessment has been done to authenticate the resultant scores in both tables. It has been detected that grading score table contributed good results. The teacher has only 4 choices to grade the students'answer. Our projected methodology gave more exact score depending upon the method used as shown in Fig. 1.

4.2 *Dataset Collection*

Online assessment of 50 undergraduate students' dataset was taken from Virtual University of Pakistan (VU). The dataset comprised of an online assessment of students in undergraduate course i.e. Open Source Web Application Development, spring semester, 2016. The test was conducted from August 11, 2016, to August 12, 2016. The teacher posted a question on Learning Management System (LMS) from Open Source Web Application Development course and 50 students gave replies in response to the question. Then, the teacher read all answers manually and gave marks in 1 to 5 series to the related reply based on the value of the text answers.

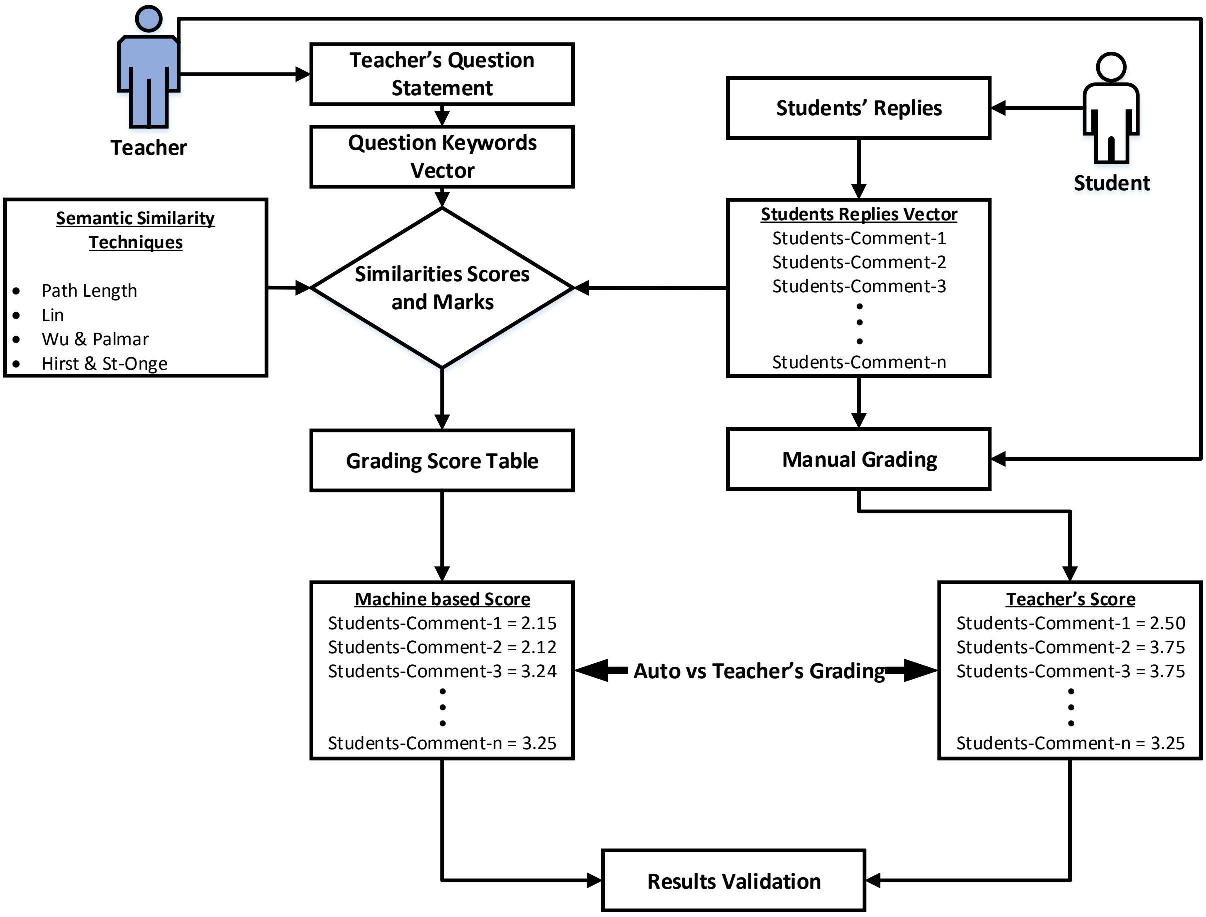
Teacher
Teacher's Question Statement
Question Keywords Vector
Students' Replies
Student
Semantic Similarity Techniques
Path Length
Lin
Wu & Palmar
Hirst & St-Onge
Similarities Scores and Marks
Students Replies Vector
Students-Comment-1
Students-Comment-2
Students-Comment-3
Students-Comment-n
Grading Score Table
Manual Grading
Machine based Score
Students-Comment-1 = 2.15
Students-Comment-2 = 2.12
Students-Comment-3 = 3.24
Students-Comment-n = 3.25
Auto vs Teacher's Grading
Teacher's Score
Students-Comment-1 = 2.50
Students-Comment-2 = 3.75
Students-Comment-3 = 3.75
Students-Comment-n = 3.25
Results Validation

Fig. 1: Methodology of student replies assessment using WordNet based similarity calculation and teacher' marks

4.3 WordNet Similarity Techniques

Text similarity is a big challenge for researchers to rank the most relevant text by intelligence and context of words. WordNet is a treasure of words which comprises synonyms details which can be used for text similarity regarding semantics. Different types of WordNet similarity methods are also used for semantically matching text. First stop words like POS (Part Of Speech) are detached; then keywords a remind from the text, which clarifies the complete sense of all text. Then stemming process is applied to decrease the words into its root term. For example, worked and working goes to one root word work. So, there may be more than one keyword fiting into one stem word. After that, Path Length, Lin, Wu & Palmer and Hirst & St-Onge semantic similarity approach is applied on our dataset to mark answers in 1 to 5 marks. Then the manually allocated marks from a teacher are associated with our automatically allotted marks from our test results.

Path Length**:** It works on the counting of nodes along the shortest path between synset1 and synset2. The semantic relatedness is inversely related to several nodes in the shortest path. The relatedness cost will be in the choice of 0 to 1. If two synsets are the same, then maximum relatedness value will be 1 [29].

Lin: The Lin measure works on information content and semantic relatedness value will be in the range of 0 to 1. If information content of synset1 and synset2 is zero, then semantic relatedness value returned from Lin measure is 0 [29], as shown in Eq. (3).

$$Lin = 2 * IC(LCS) / (IC(synset1) + IC(synset2)) \qquad (3)$$

Where IC is information content, and synset is synonyms' information of specified word and LCS (Least Common Subsumer) of synset1 and synset2.

Wu & Palmer**:** It computes semantic relatedness value by seeing the depths of two synsets in WordNet taxonomies to the depth of LCS [29], as shown in Eq (4).

$$Wu\_\&\_Palmer = 2 * depth(LCS) / (depth(s1) + depth(s2)) \qquad (4)$$

The affiliation score is in the range of 0 to 1. In this method, the score will not be 0 because the depth of LCS is never 0 among the two synsets, so it contributes a better result in our research.

Hirst & St-Onge: It works by extracting lexical information between the two-word intelligence. The semantic relatedness score is in the range of 0 to 16. It has additional three classes; extra strong, medium strong and strong which are used for scheming semantic relatedness score [29].

5. Results and Discussion

Four semantic similarity techniques have been applied to the data set to assign marks automatically in 1 to 5 range. Then the results are compared with manually assigned marks from the teacher. Our proposed methodology assigns marks in 1 to 5 range.

5.1 Pilot Study

In this study, keywords were mined from 8 students' answers. In the next step, WordNet semantic similarity techniques wer applied to find the semantic relatedness score among query keywords and students' answer keywords. The information of ICS reply-1 keywords with queries' keywords with semantic relatedness scores is shown in Table 1.

In this study, the Wu & Palmer technique presented better outcomes than other techniques as it reflects depths of two synsets in WordNet taxonomies along the depth of LCS. It can be seen from Table 2 that reply 9 is more related to the question and assigned highest value 4.63 using Wu & Palmer technique while reply 10 is less similar to the query which has 1.04 value using *Hirst & St-Onge* method but teacher allocated 2.5 marks to answer10 manually. This research has allocated peak value of 4.63 to reply 9 and teacher also allocated high marks 3.75, which is clear that semantic relatedness score gave improved relevancy grades than the manually allocated marks from the teacher as shown in Table 2.

Table 1: Similarity scores of reply 1

Question Keywords	Reply Keywords	Path Length	Lin	Wu & Palmer	Hirst & St-Onge
Dominate	Demand	0.33	0.00	0.50	0.00
Server	Website	0.33	0.00	0.91	5.00
	System	0.20	0.16	0.78	3.00
Script	Web	0.14	0.09	0.22	0.00
Language	Website	0.07	0.00	0.25	0.00
	Scaled values (1-5)	2.08	1.24	3.66	1.50

Table 2: Comparison of WordNet similarity measure and teacher's marks

Students' Replies	Path Length	Lin	Wu & Palmer	Hirst & St-Onge	Teacher's Marks
Reply 1	2.53	2.25	3.26	2.25	2.5
Reply 2	2.75	3.13	4.28	2.64	3.75
Reply 3	2.40	2.91	4.51	2.17	3.75
Reply 4	1.91	1.92	3.46	1.47	2.5
Reply 5	1.82	1.99	3.57	1.16	3.75
Reply 6	1.95	1.73	3.84	1.55	3.75
Reply 7	1.75	1.72	3.28	1.39	3.75
Reply 8	1.88	1.58	3.08	1.31	5
Reply 9	2.50	3.53	4.63	2.02	3.75
Reply 10	1.89	1.23	2.62	1.04	2.5

In Fig. 2 reply facts are given horizontally and standing of students' answers are given vertically. Ranking is provided in the range of 1 to 5 as shown in Fig. 2, in which highest mark is 5. Semantic relatedness score for keywords is shown in dissimilar colors. Yellow colour shows teacher's marks, dark blue colour shows Hirst & St-Onge

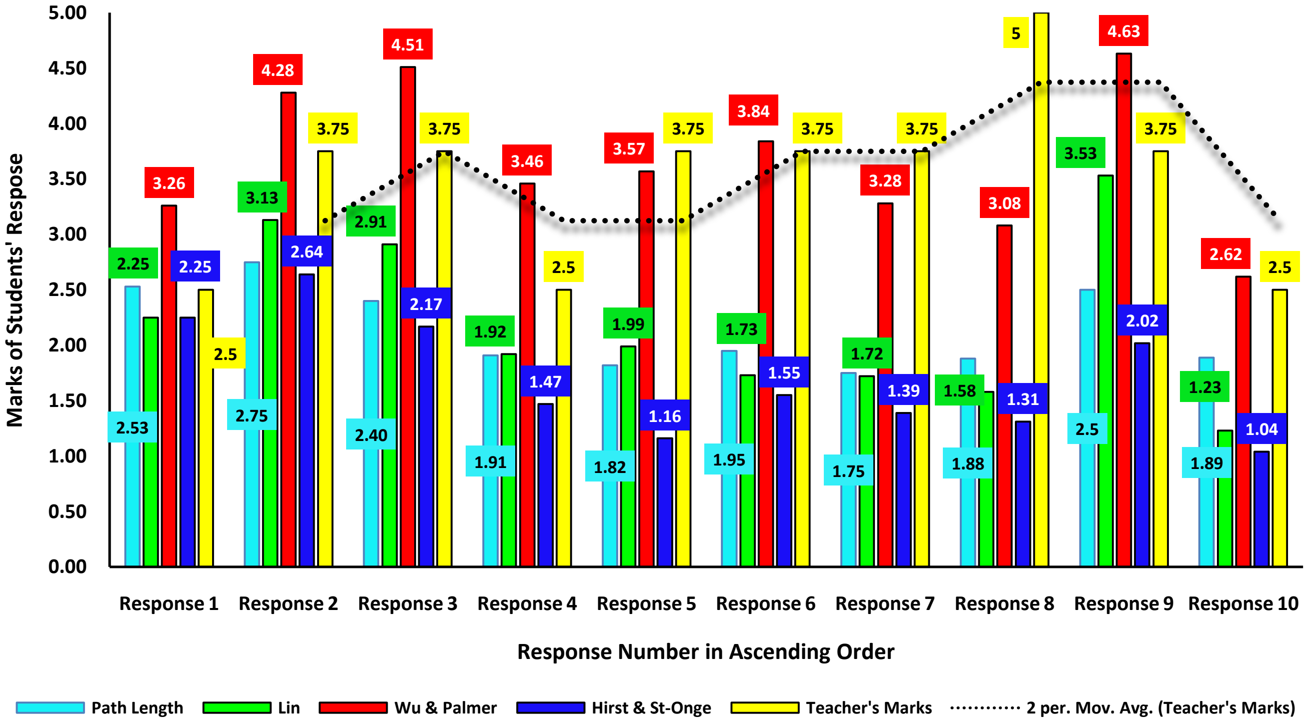

Fig. 2: Assessment and analysis by the teacher and WordNetsimilarity measures scoring

score, cyan shows Path Length score, red colour shows Wu & Palmer score and lime green colour shows Lin scores. This figure shows an assessment between teacher's marks and semantic relatedness scores from given techniques. The Wu & Palmer scores gave the best relevancy score as associated with teacher's marks than other used procedures because it computes semantic relatedness score between the depths of two synsets along the WordNet taxonomy of words. The spotted curve in Fig. 2 shows the moving average or running average or moving mean from statistics, which examines data facts by scheming a series of averages from dissimilar subsets of the dataset. It takes the average of first two answer's scores. Then it takes the answer 3 score and calculates the average with previous and so on. Finally, average score's points in spotted curve show the assessment among answer's scores with teacher score. It generates different subsets from bigger dataset to comprehend the overall performance of dataset. Mean similarity scores was computed for all techniques to associate with the extreme score. In Table 3 it is shown that extreme scores are close to manually allocated marks thus offer improved results.

Table 3: Mean and max similarity measures and teacher's marks

Students' Replies	Max out of all measures	Mean of Similarity Measures	Teacher's Marks
Reply 1	3.26	2.57	2.5
Reply 2	4.28	3.20	3.75
Reply 3	4.51	3.00	3.75
Reply 4	3.46	2.19	2.5
Reply 5	3.57	2.14	3.75
Reply 6	3.84	2.27	3.75
Reply 7	3.28	2.04	3.75
Reply 8	3.08	1.96	5
Reply 9	4.63	3.17	3.75
Reply 10	2.62	1.70	2.5

6. Conclusion

E-assessment of students can be mechanized using WordNet semantic similarity measures. It is examined that as an alternative of manual assignment of marks to students' replies, WordNet dictionary can be used to extract semantic sense from the text relatively rather than just corresponding keywords. In this research WordNet, semantic similarity procedures were used to mark the students' answers by semantics matching. The dataset contains 210 pairs of words from 50 undergraduate students. The teacher has only 4 baskets for the assignment of marks, but our proposed methodology gives a correct measure by the semantics of complete text of student's reply.

In future, our proposed methodology can be enhanced by automatic selection of keywords which give complete sense and structure of answer. An algorithm can be designed to select best keywords automatically based on similarity measures, i.e., Path Length [25], Lin [26], Wu & Palmer [27], and Hirst & St-Onge [28]. Further, we will apply more WordNet semantic similarity measures and will examine results statistically.

References

[1] G. Sidorov, A. Gelbukh, H. Gómez-Adorno and D. Pinto, "Soft similarity and soft cosine measure: Similarity of features in vector space model", Computación y Sistemas, vol. 18, pp. 491-504, 2014.

[2] G. Sidorov, H. Gómez-Adorno, I. Markov, D. Pinto and N. Loya, "Computing text similarity using tree edit distance", Fuzzy Information Processing Society (NAFIPS) held jointly with 5th World Conference on Soft Computing (WConSC), Annual Conference of the North American, pp. 1-4, 2015.

[3] G. Sidorov, F. Velasquez, E. Stamatatos, A. Gelbukh and L. Chanona-Hernández, "Syntactic dependency-based n-grams as classification features", Mexican International Conference on Artificial Intelligence, pp. 1-11, 2012.

[4] O. Luaces, J. Díez, A. Alonso-Betanzos, A. Troncoso and A. Bahamonde, "Content-based methods in peer assessment of open-response questions to grade students as authors and as graders", Knowledge-Based Systems, vol. 117, pp. 79-87, 2017**,** 2016.

[5] E. Agirre, C. Banea, D. Cer, M. Diab, A. Gonzalez-Agirre, R. Mihalcea, et al., "Semeval-2016 task 1: Semantic textual similarity, monolingual and cross-lingual evaluation", Proc. of the 10th International Workshop on Semantic Evaluation, pp. 509-523, 2016.

[6] M. Kastner, J. Antony, C. Soobiah, S. E. Straus and A. C. Tricco, "Conceptual recommendations for selecting the most appropriate knowledge synthesis method to answer research questions related to complex evidence", J. Clinical Epidemiology, vol. 73, pp. 43-49, 2016.

[7] H. Y. Kang, S. H. Moon, H. J. Jang, D. H. Lim and J. H. Kim, "Validation of "quality-of-life questionnaire in Korean children with allergic rhinitis" in middle school students", Allergy, Asthma & Respiratory Disease, vol. 4, pp. 369-373, 2016.

[8] S. Burrows, I. Gurevych, and B. Stein, "The eras and trends of automatic short answer grading", Int. J. of Artificial Intelligence in Education, vol. 25, pp. 60-117, 2015.

[9] J. Kim, G. Chern, D. Feng, E. Shaw, and E. Hovy, "Mining and assessing discussions on the web through speech act analysis", Proc. of the Workshop on Web Content Mining with Human Language Technologies at the 5th International Semantic Web Conference, 2006.

[10] A. Otegi, X. Arregi, O. Ansa, and E. Agirre, "Using knowledge-based relatedness for information retrieval" Knowledge and Information Systems, vol. 44, pp. 689-718, 2015.

[11] I. Partalas, A. Kosmopoulos, N. Baskiotis, T. Artieres, G. Paliouras, E. Gaussier, et al., "Lshtc: A benchmark for large-scale text classification", arXiv preprint arXiv:1503.08581, 2015.

[12] H. Cigdem and S. Oncu, "E-assessment adaptation at a military vocational college: student perceptions", Eurasia J. Math., Sci. & Tech. Education, vol. 11, pp. 971-988, 2015.

[13] T. Chen and B. Van Durme, "Discriminative information retrieval for knowledge discovery", arXiv preprint arXiv:1610.01901, 2016.

[14] P. Bille, "A survey on tree edit distance and related problems", Theoretical Computer Science, vol. 337, pp. 217-239, 2005.

[15] K. Tymoshenko, D. Bonadiman and A. Moschitti, "Learning to Rank Non-Factoid Answers: Comment Selection in Web Forums", Proc. of the 25th ACM Int. on Conf. on Information and Knowledge Management, pp. 2049-2052, 2016.

[16] M. M. Goyal, N. Agrawal, M. K. Sarma and N. J. Kalita, "Comparison Clustering using Cosine and Fuzzy set based Similarity Measures of Text Documents", arXiv preprint arXiv:1505.00168, 2015.

[17] H. Xu, W. Zeng, J. Gui, P. Qu, X. Zhu and L. Wang, "Exploring similarity between academic paper and patent based on latent semantic analysis and vector space model", 12th Int. Conf. on Fuzzy Systems and Knowledge Discovery (FSKD), 2015, pp. 801-805, 2015.

[18] R. Jiang, S. Kim, R. E. Banchs and H. Li, "Towards improving the performance of vector space model for Chinese frequently asked question answering," Int. Conf. on Asian Language Processing (IALP), pp. 136-139, 2015.

[19] X. Bao, S. Dai, N. Zhang and C. Yu, "Large-Scale Text Similarity Computing with Spark", Int. J. of Grid and Distributed Computing, vol. 9, pp. 95-100, 2016.

[20] R. Mihalcea, C. Corley and C. Strapparava, "Corpus-based and knowledge-based measures of text semantic similarity", Proc. of the 21st National Conference on Artificial Intelligence, Boston, Massachusetts, USA, pp. 775-780, 2006.

[21] B. Acree, E. Hansen, J. Jansa, and K. Shoub, "Comparing and Evaluating Cosine Similarity Scores, Weighted Cosine Similarity Scores, & Substring Matching "http://hansen.web.unc.edu/files/2014/12/AHJS_Weighted_Cosine.pdf, 2016.

[22] W.-N. Zhang, Z.-Y. Ming, Y. Zhang, T. Liu and T.-S. Chua, "Capturing the semantics of key phrases using multiple languages for question retrieval", IEEE Transactions on Knowledge and Data Engineering, vol. 28, pp. 888-900, 2016.

[23] E. Delen, "Enhancing a Computer-Based Testing Environment with Optimum Item Response Time", Eurasia J. Math., Sci & Tech. Education, vol. 11, pp. 1457-1472, 2015.

[24] M. De Boni and S. Manandhar, "The use of sentence similarity as a semantic relevance metric for question answering", New Directions in Question Answering, pp. 138-144, 2003.

[25] J. J. Jiang and D. W. Conrath, "Semantic similarity based on corpus statistics and lexical taxonomy", arXiv preprint cmp-lg/9709008, 1997.

[26] D. Lin, "An information-theoretic definition of similarity", Icml, pp. 296-304, 1998.

[27] Z. Wu and M. Palmer, "Verbs semantics and lexical selection", Proc. of the 32nd Annual Meeting on Association for Computational Linguistics, pp. 133-138, 1994.

[28] G. Hirst and D. St-Onge, "Lexical chains as representations of context for the detection and correction of malapropisms", WordNet: An electronic lexical database, vol. 305, pp. 305-332, 1998.

[29] T. Pedersen, S. Patwardhan and J. Michelizzi, "WordNet: Similarity: measuring the relatedness of concepts," Demonstration papers at HLT-NAACL 2004, pp. 38-41, 2004.