iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** XI **Month of publication:** November 2017

DOI: http://doi.org/10.22214/ijraset.2017.11123

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue XI November 2017- Available at www.ijraset.com

Monitoring Classroom Teaching Relevance Using Speech Recognition Document Similarity

Raja Mathanky S[1]
[1]*Computer Science Department, PES University*

Abstract: ***In any educational institution, it is imperative to maintain teaching quality. One main factor influencing teaching quality in a classroom is the relevance of the content taught by the teacher to the information provided by the suggested textbook. This paper presents an application to evaluate a teacher's performance by calculating a similarity measure between the contents of the lecture and the textbook. The video (or audio) of the lecture is obtained, and speech recognition engines are used to convert the speech to text. This transcript is then cleaned and compared against the uploaded textbook. A semantic document similarity technique is then used to arrive at a similarity measure that mirrors the relevance of the lecture. The results obtained are 77% accurate, and this accuracy depends on the speech recognition engine and the semantic similarity algorithm used.***
Keywords: ***Semantic Similarity, Document Similarity, Speech Recognition, Teaching Relevance, WordNet***

I. INTRODUCTION

1) In any school or college, the most important school-related factor affecting student performance is the quality of teaching. One main factor that determines the productivity of a classroom lecture is its relevance to the subject. In every educational institution, a curriculum is drafted for each subject. The curriculum embodies all the important concepts of the subject in a sequential manner, designed carefully so that students possess all the prerequisite knowledge required to understand a particular concept. It is extremely important for a teacher to follow the sequence prescribed by the curriculum within every concept and across all concepts. This improves the flow of concepts, hence increasing the understanding of the subject among students.

2) Additionally, the curriculum, which includes content from textbooks and reference materials, becomes the primary source of the subject in a classroom. A teacher must stick to the facts provided by these materials, and not add or remove concepts based on unwarranted assumptions. Since student performance is of utmost importance to an educational institution, it is necessary to monitor the relevance of classroom teaching to the prescribed curriculum. This will help the institution assess teacher performances and train the teachers who are not adhering to the syllabus.

3) The application presented in the paper is a web-based tool that calculates the pertinence of a class room lecture to the prescribed textbook. This is done by obtaining the video (or audio) of the lecture, converting the speech in the video to text and then comparing this text to the content present in the textbook using semantic document similarity techniques.

4) This paper also examines the accuracy of some popular semantic similarity measures such as Cosine Similarity for Vector Space Models, TF-IDF Victimization, Latent Semantic Analysis (LSA) and Singular Value Decomposition(SVD) and WordNet based similarity.

5) In the second section of this paper, the various steps involved in the application are delineated and the flow of the model is presented. In the third section of the paper, conclusions are made about the accuracy of the model and ways to make it more efficient. Possible directions of future research in the areas of Speech Recognition and Document Similarity are also examined.

II. METHODOLOGY

6) In this section, the various processes involved in the model are presented. Fig 1 illustrates the framework in the form of a flowchart.

©IJRASET (UGC Approved Journal): All Rights are Reserved

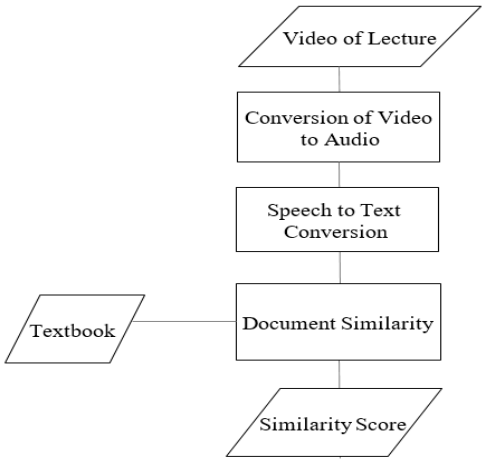


Fig 1: Flowchart of Methodology Used

A. Conversion of Video to Audio

7) This application takes the video of the classroom lecture as the input. The main assumption made is that each class is recorded using a camera and a microphone. This video is then converted to an audio file, which can be used for speech recognition.

B. Speech to Text Conversion

1) A speech recognition engine is used to convert the audio file of the classroom lecture to a transcript, which is then used to assess the similarity to the prescribed textbook. Common measures of accuracy of the speech recognition engine are:

2) Character Error Rate (CER %): This metric is used to measure error rate at the syllable level. This metric is not very useful if the speech is in English, as each character is phonetically different from the others. But in a language such as Mandarin, which has different characters with the same pronunciation, this measure plays a vital role in determining the precision of the speech recognition engine. Since there are no word or sentence boundaries in Mandarin, interpretation of the characters play a prominent role in determining these boundaries, hence the meaning of the utterance.

3) Word Error Rate (WER%): This metric is the most common performance indicator of a speech recognition engine. When words from the transcripts obtained after the conversion of speech to text is compared to the expected transcripts, three types of error can arise:

Insertion – a character is present in a word in the speech recognition output, but not in the reference transcript. The number of such occurrences is denoted as I.

Deletion- a character is present in a word in the reference transcript, but not in the speech recognition output. The number of such occurrences is denoted as D.

Substitution- a character in the reference transcript is misinterpreted (substituted) for another in the speech recognition output. The number of such occurrences is denoted by S.

4) The word error rate is given by the formula:

5) $WER = \frac{I+D+S}{N}$

6) where N is the total number of words in the reference transcript.

7) F-measure: The WER and CER of an automatic speech recognition engine provide an adequate measure for applications such as sub-titling, where the correct transcription of every word is of importance. However, these metrics falls short of capturing the essence of performance in other applications where the detection of key-terminology is of primary importance.

8) The F-measure is a function of precision and recall. Precision and Recall of a speech ecognition engine depends on the following parameters:

a) True Positives (TP) – number of keywords which occur in the audio and which are detected by the system.

b) False Positives (FP)– number of keywords which are detected by the system but which aren't actually uttered by the speaker.

c) False Negatives (FN)– number of keywords which are uttered by the student but not detected by the system

d) True Negatives (TN)– Number of keywords which are not uttered by the student and are not detected by the system

©IJRASET (UGC Approved Journal): All Rights are Reserved

9) Precision is the fraction of retrieved documents that arerelevantto the query. It is given by the formula:

10) $PRECISION = \frac{TP}{TP+FP}$

11) Recall is the fraction of relevant documents that are successfully retrieved. It is given by the formula :

$$RECALL = \frac{TP}{TP + FN}$$

12) Precision and recall are usually related in an inverse manner: higher precision typically results in lower recall and vice-versa. The F-measure combines these two measures to arrive at a combined rate; often the point of equal-error-rate (EER) is cited in performance evaluations. It is given by the formula:

$$F - MEASURE = \frac{2 \times PRECISION \times RECALL}{PRECISION + RECALL}$$

13) A speech recognition engine of high linguistic performance has a high F-measure value. WER must be low, not exceeding 35%. Lower CER and WER measures are indicative of a more accurate speech to text conversion.

C. Document Similarity

14) In this phase, the appropriate part of the textbook that was to be covered in class, is passed as input to the application. The similarity score between the transcripts obtained from the speech recognition stage and the textbook material is computed. Document similarity (or distance between documents) is a one of the central themes in Information Retrieval. In general, documents are considered similar if they are semantically close and describe similar concepts. We will review several common approaches.

15) Cosine Similarity for Vector Space Models (VSM): In this approach, each document is considered as a bag of words. Each document is represented in the form of a sparse vector, which contains the number of occurrences of each word. The level of similarity between two documents is a measure of the angle between the two vectors representing each document.

16) Cosine Similarity between documents doc1 and doc2 is given as follows:

$$SIMILARITY(DOC1, DOC2) = \frac{DOC1 \,.\, DOC2}{|DOC1| \times |DOC2|}$$

17) This approach is not the best way to compute the similarity between the documents. This metric is a measurement of orientation and not magnitude, it can be seen as a comparison between documents on a normalized space because we're not taking into the consideration the magnitude of each word count (tf-idf) of each document, but the angle between the documents. This method ignores the higher term counts on documents. Suppose we have a document with the word "sky" appearing 200 times and another document with the word "sky" appearing 50, the Euclidean distance between them will be higher but the angle will still be small because they are pointing to the same direction, which is what matters when we are comparing documents.

18) TF-IDF Vectorization: Similarity between documents can be computed using the TF-IDF Vectorization method.

19) This method, although a bag-of-words approach, helps to filter out the helpful words (words that play an important role in distinguishing the documents) and words that contribute little towards distinguishing the documents by assigning weights to each of these words. It is a way to score the importance of words (or "terms") in a document based on how frequently they appear across multiple documents. The two measures are used to weight each term present in the documents are Term Frequency (TF) and Inverse Document Frequency (IDF).Term Frequency of a word, also known as TF, measures the number of times a term (word) occurs in a document. If a word appears frequently in a document, the word is important, and is given a high weight. Term frequency of a term t in a document d is given by the formula:

$$TF(t, d) = 0.5 + 0.5 \times \frac{freq(t, d)}{\max\{freq(t'd) : t' \in d\}}$$

20) Inverse Document Frequency of a word, also known as IDF, measures how common a word is among all documents. If a word appears in many documents, it's not a unique identifier of the document, and is given a low weight. A low document frequency of a word indicates that it is a unique identifier of a document. IDF of a term is given by the formula:

21) $IDF(t, D) = log \frac{N}{|\{d \in D : t \in D\}|}$

©IJRASET (UGC Approved Journal): All Rights are Reserved

22) The TfIdf value for a word is given by the following formula :

$$TF - IDF(t, d, D) = TF(t, d) \times IDF(t, D)$$

23) A high value of the TfIdf measure implies that the term is very important in determining the similarity measure between the documents.

24) There are major drawbacks with this method. Since this is a bag of words approach, it fails to capture position in text, semantics and co-occurrences in documents. It also fails to disambiguate polysemy(coexistence of many possible meanings for a single term) and synonymy (different terms conveying the same meaning).

25) Latent Semantic Analysis (LSA) and Singular Value Decomposition (SVD):Latent Semantic Analysis (LSA), also known as Latent Semantic Indexing (LSI) takes a step forward from the previous approaches, to find the underlying meaning or concepts of those documents. LSA attempts to solve this problem by mapping both words and documents into a "concept" space and doing the comparison in this space. Due to usage of synonyms, these concepts are obscured, leading to a noise. LSA attempts to find the smallest set of concepts that span all the documents. LSA also makes use of a bag of words approach, where order of words aren't considered. This algorithm also assumes that each word can have only one meaning (it ignores polysemy).

26) LSA is centered around computing a partial singular value decomposition (SVD) of the document term matrix (DTM). This decomposition reduces the text data into a manageable number of dimensions for analysis. Latent semantic analysis is similar to principal components analysis.

27) The singular value decomposition approximates the DTM using three matrices: **U**, **S**, and **V'**. The relationship between these matrices is defined as follows:

$$DTM \approx U * S * V'$$

28) The singular vectors capture connections among different words with similar meanings or topic areas. If three words tend to appear in the same documents, the SVD is likely to produce a singular vector in V' with large values for those three words. The U singular vectors represent the documents projected into this new term space.

29) Although LSA provides a good measure of the semantic similarity between documents, it has certain limitations. LSA cannot handle polysemy (words with multiple meanings) effectively. It assumes that the same word means the same concept which causes problems for words like bank that have multiple meanings depending on which contexts they appear in. LSA depends heavily on SVD which is computationally intensive and hard to update as new documents appear. However recent work has led to a new efficient algorithm which can update SVD based on new documents in a theoretically exact sense.

30) WordNet based Semantic Similarity: In the previous approaches examined, each document is represented as a vector of characteristic features (words/terms). This feature selection ignores the semantic information present in the document, resulting in an inaccurate similarity score. Such approaches don't take polysemy and synonymy into consideration. This application uses a WordNet based semantic similarity algorithm.

31) As described in [1], the WordNet based approach incorporates co reference resolution and examines semantic relationships among words by tackling polysemy and synonymy problems using WordNet and semantic similarity.

32) WordNet is a lexical English database that groups Nouns, Verbs, Adjectives and Adverbs into sets of synonyms called synsets. Synset forms a basic building block of the WordNet. Each synset consists of a set of synonyms expressing a particular concept. Different words having the same sense are grouped into same synset and different senses of the same word are separated into different synsets. This approach consists of the following phases:

33) Document Preprocessing is done to transform a document into a suitable form for measuring similarity. Fig 2 shows the sub modules of preprocessing module each of which is explained subsequently.

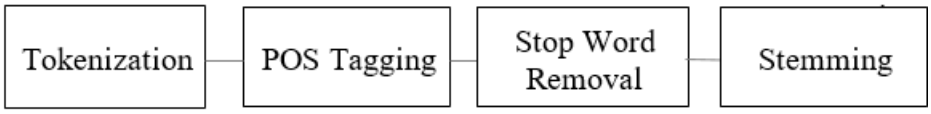


Fig 2: Steps in Preprocessing

34) *Tokenization:* Each sentence is partitioned into a list of words, and we remove the stop words. Stop words are frequently occurring, insignificant words that appear in a database record, article, or a web page, etc.

©IJRASET (UGC Approved Journal): All Rights are Reserved

35) *POS Tagging*: Parts of speech such as noun, verb, adjective, adverb etc. of each word in the document is identified and is tagged with it. Identifying Part of Speech of each word is important as it helps in exploiting the information from the WordNe

36) *Stop Word Removal:* A document contains thousands of words. Some words do not contribute to the meaning of the document. Such words are called stop words. Identifying and eliminating stop words helps in reducing the size of feature space for the document representation.

37) *Stemming:* Stemming is the process of reducing an inflected(derived) or a morphological form of a word to its root form. The most widely used algorithm for this is the Porter Stemming Algorithm. This can be thought of as a lexical final state machine with the following states

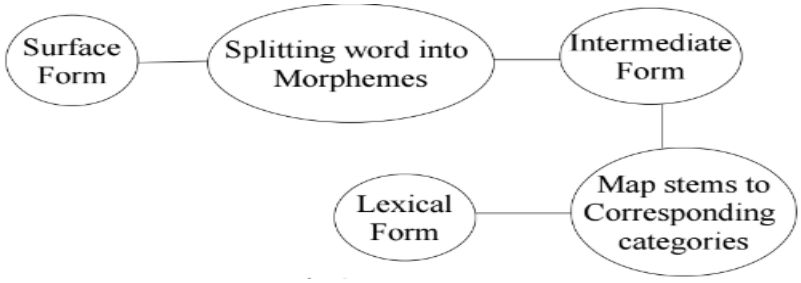


Fig 3: Steps in Stemming

D. Word Sense Disambiguation (WSD)

1) Word sense disambiguation is the process of finding out the most appropriate sense of a word based on the context in which it is used. Word Net performs this by assigning a synset ID to each of the words that are to be disambiguated, thus providing a solution for polysemy and synonymy identification.

2) A popular and efficient algorithm for carrying out WSD is the Micheal Lesk Algorithm. To disambiguate a word in a phrase, the gloss of each of its senses, which are taken from an English dictionary, is compared to the glosses of every other word in the phrase. A word is assigned to the sense whose gloss shares the largest number of words in common with the glosses of the other words.

E. Semantic Similarity

1) Similarity is measured at three levels

a) *Word Level Similarity:* Worded can be used to measure semantic similarity between two synsets. To compute the similarity between two words, we base the semantic similarity between word senses. We capture semantic similarity between two word senses based on the path length similarity. A simple way to measure the semantic similarity between two synsets is to treat taxonomy as an undirected graph and measure the distance between them in WordNet. P. Resnik quoted once, "The shorter the path from one node to another, the more similar they are". Note that the path length is measured in nodes/vertices rather than in links/edges. The length of the path between two members of the same synset is 1 (synonym relations). Fig 4 shows an example of the hyponym taxonomy in WordNet used for path length similarity measurement:

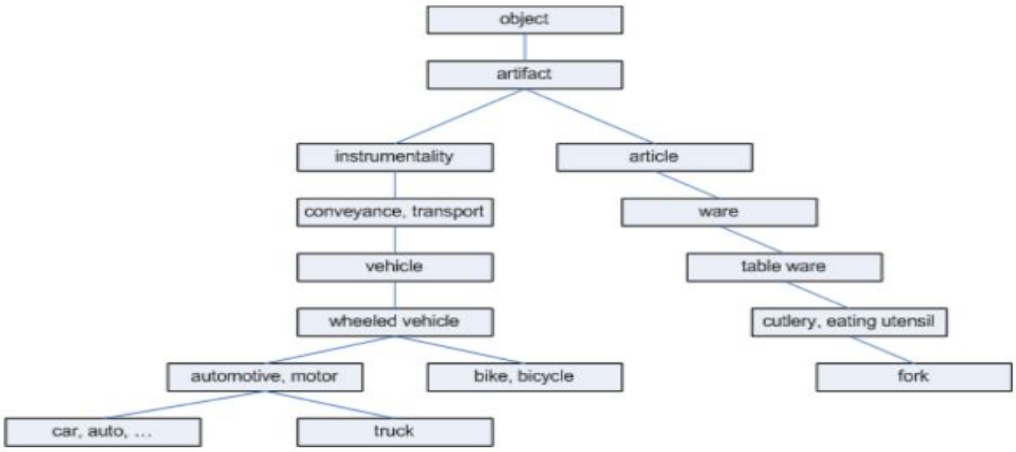


Fig 4 : Synsets and Word Similarity

©IJRASET (UGC Approved Journal): All Rights are Reserved

b) Sentence Level Similarity: To compute the similarity between two sentences, we build a semantic similarity relative matrix R[m, n] of each pair of word senses, where R[i, j] is the semantic similarity between the most appropriate sense of word at position i of sentence X and the most appropriate sense of word at position j of sentence Y.

Thus, R[i,j] is also the weight of the edge connecting from i to j. We formulate the problem of capturing semantic similarity between sentences as the problem of computing a maximum total matching weight of abipartite graph, where X and Y are two sets of disjoint nodes. To compute the similarity between these two sentences, the following formula is used:

$$SENTENCE\ SIMILARITY = \frac{2 \times MATCH(X,Y)}{|X|+|Y|}$$

c) Document Level Similarity: To compute document level similarity, sentence level similarity is computed for all possible pairs of sentences. The arithmetic mean of all such values will result in the semantic similarity between the documents.

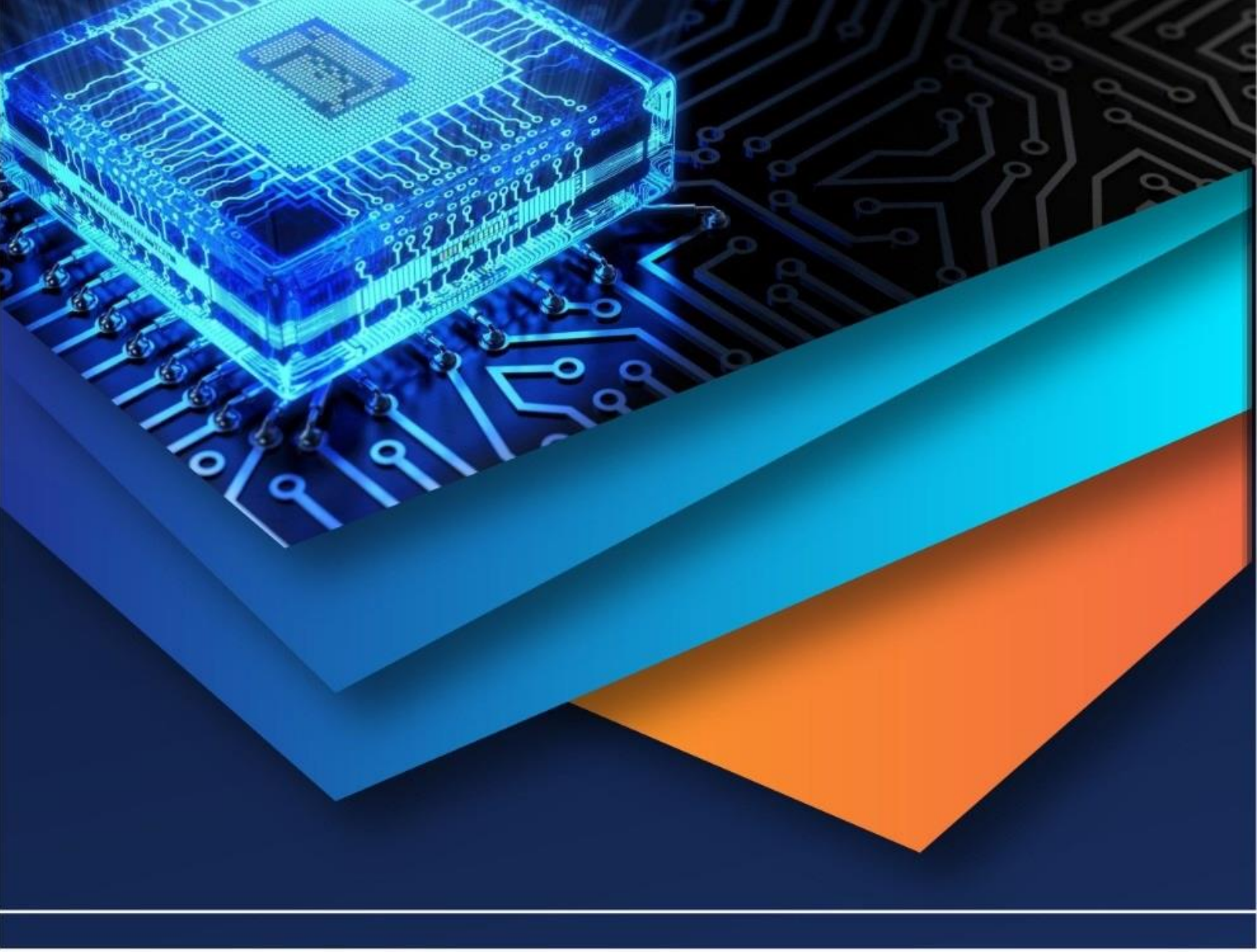
III.CONCLUSION

The application presented in this paper is an efficient and hassle-free technique to measure teacher performance. Choice of speech recognition engine is crucial to the performance of the application. The use of WordNet based semantic similarity algorithm has increased the accuracy of the similarity measure to around 77%, which is a stark improvement as compared to Cosine Similarity, TF-IDF and LSA. Further increases in accuracy can be achieved by using ontology-based deep learning and natural language processing (NLP) algorithms.

REFERENCES

[1] 2016 International Conference On Computational Systems and Information Systems for Sustainable Solutions – WordNet and Semantic Similarity based Approach for Document
[2] MIPRO 2017, May 22-26, 2017, Opatija, Croatia -The Struggle with Academic Plagiarism : Approaches based on Semantic Similarity
[3] SAI Computing Conference July 13-15, 2016, London, UK – Visualizing Document Similarity using N-Grams and LSA
[4] WordNet-based semantic similarity measurement- https://www.codeproject.com/Articles/11835/Wordnet-based-semantic-similarity-measurement
[5] Latent Semantic Analysis – https://technowiki.wordpress.com/2011/08/27/latent-semanticanalysis-lsa-tutorial
[6] TF-IDF and Cosine Similarity – https://janav.wordpress.com/2013/10/27/tf-idf-and-cosinesimilarity

©IJRASET (UGC Approved Journal): All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)