
SMART LEARNING ASSISTANT SYSTEM

Prof. Priyanka Pujari^{*1}, Amrutha S^{*2}

^{*1,2}Department of Computer Science and Engineering, Angadi Institute of Technology and Management,
Belagavi, Karnataka, India

ABSTRACT

This survey paper explores how the technology can be used in improving the smart learning techniques that enhances better learning revising of the concepts by the students. The paper ensures the students the faster learning and efficient performance in the academics. Use of smart learning concepts has been increased day by days as better teaching and learning methods. Smart learning concepts are majorly helping the students to learn better in classrooms. It's been a difficult task during the class hours to listen, understand and take notes. In order to overcome these problems. The goal of this research is to help the students in learning. This research proposes a solution that contains the following components They are the smart noter, smart slide matcher, Smart reference finder, and smart question presenter which enables the students to get a summary of the concepts, easily follow the concepts during class hours and refer to the related resources for extra references and revise during the exam times and increases their academic performances.

Keywords: smart note taker, smart reference finder, smart question presenter, smart slide matcher.

I. INTRODUCTION

The application of smart learning system in the development of smart learning environments has gained momentum in educational firms and learning environments. Various issues associated with the learning in classrooms can be solved with help of different components such as note taker, reference finder, question presenter, smart question presenter to overcome the difficulties in learning [1]. There were many challenges faced by the students in the note taking process that can be avoided and can assist the student in efficiently taking the notes. Typically note taking is assumed as the common-sense skill that students either have it or learn by the trial-and-error method. This research helps to overcome these problems [2]. The different barriers and challenges to smart learning are considered and discusses the pillars of smart learning. The pillars of learning include accessibility, flexibility and affordability. The smart learning is the best and suitable path for the improvement of the student academic performance [3]. This study examines the current research on Smart learning systems, including their benefits, challenges, and potential applications. Furthermore, it aims to investigate the current challenges in the learning system. By comprehending the current state of research and its implications, this study intends to offer insights for future development and implementation of Smart learning environments in the schools and the colleges.

II. METHODOLOGY

Proposed System

The smart learning system is making the learning process easy to assist in faster and effective learning with help of algorithms. The system has four main components: Smart Note taker, Smart Reference Finder, Smart Question presenter and smart slide matcher. These components integrated to make the process of learning easy and effective. The system is assisting the students in smartly learning the concepts effectively and consistently by saving the time. Smart note taker helps the students in making the effective notes. Smart reference finder assists the students in finding the references in the web that are related to the lecture note. Smart question presenter generates and presents the question to the students. Smart slide matcher also improves effective learning of students.

System Architecture

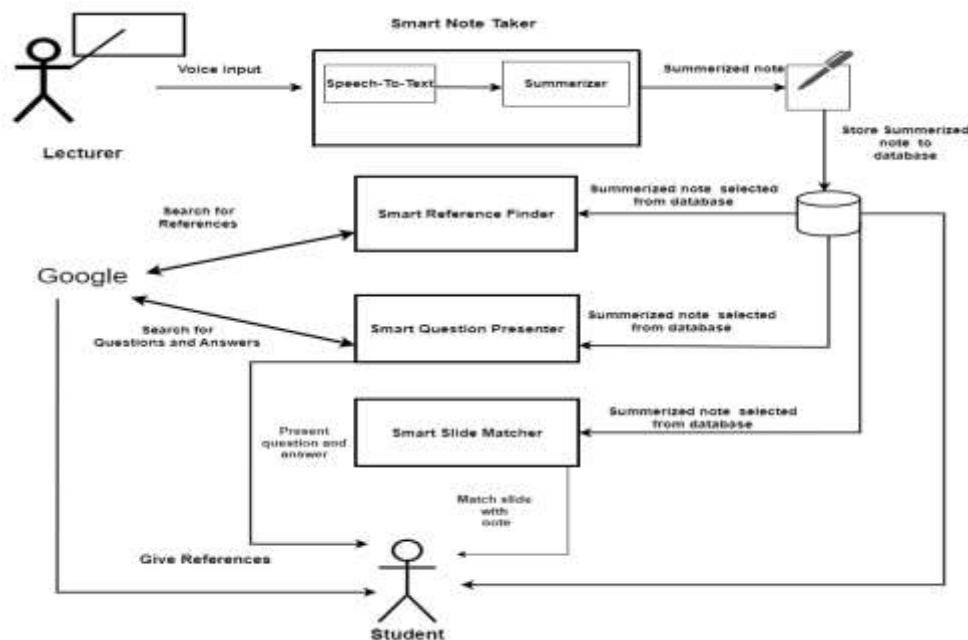


Figure 1: System Architecture

The Smart Learning system architecture is shown in Figure 1. The architecture comprises four main components: Smart Note taker, Smart Reference Finder, Smart Question presenter and smart slide matcher. Smart Note taker is used to convert the voice to the notes [4]. Smart Reference Finder finds the references related to the summarized notes [9]. Similarity between the web results and the lecturer notes can be calculated [12].

Smart Question presenter helps in forming the questions forming the questions for the selected notes from database. The answers can also be generated for the previously presented questions [15] Smart slide matcher also helps in finding the references related to the concepts. The table I below shows a summary of the components of the architecture.

Table 1. Components Summary

| Components | Description |
|--------------------------|---|
| Smart Note Taker | Speech is converted to the text. Text will be summarized using the TF-IDF algorithm and then summary is stored to database |
| Smart Reference Finder | Topics are selected from the summarized notes and it is given as input to the reference finder and it will related references over the web |
| Smart Question Presenter | Topics are selected from the summarized notes and it is given as input to the question presenter. Questions generated related to the concepts |
| Smart Slide Matcher | Topics are selected from the summarized notes and it is given as input to the slide matcher and it will give related references over the web |

Smart Note taker

The note taker component starts by recording the lecturer's voice to the computer. Then using a transcription algorithm audio is converted into text format. This ends the audio transcription part of the component. After this the summarizing part begins. The transcribed text is now converted into a summarized note by the text summarizer developed using Python 3.6 & TF-IDF text summarization algorithm. Reason to choose that particular algorithm, an online survey was conducted and the majority of people preferred the TF-IDF algorithm's summary over the Sentence Scoring algorithm's summary of the same text. Then the summarized note can be opened in the editor it helps the lecturer to edit the note manually if they desire. Then the summarized note is uploaded into the cloud database, so students can refer the concepts. Figure 2 represents the NLP process of the smart Note taker.

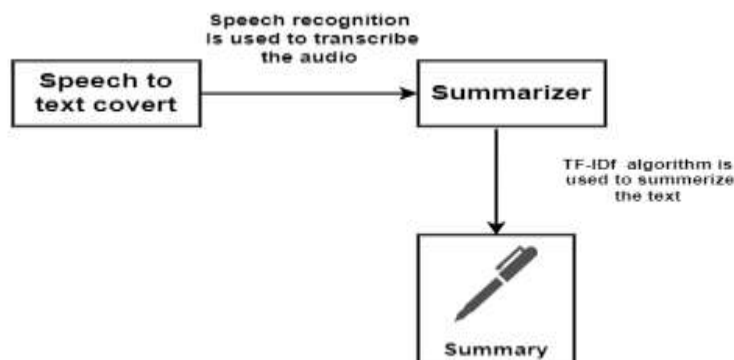


Figure 2: Smart Note taker process

Smart Reference Finder

Referring the relevant reference plays a very important role for the students when their goal is for best performance in any subject. It assists to understand difficult lessons, formulas, theories. The reference finding process is started when the student selects a summarized note that which is stored in database. Then the topic is extracted from note and google automatically searches the references based on the topic. The separate google tabs contain the references. The results include the YouTube videos, web articles, e-books, pdfs. Then with one of the references are in google, the system uses the similarity Algorithm to find the similarity between web results and summarized note. the next part of the process is with the helps of web scrapping content of the references is extracted. Before the process of comparing the content similarity the content is re-processed using the Tokenization and stop word removing techniques. Finally, the the google generated references and the summarized note are compared using the similarity algorithm and it generates a similarity value. That helps the student in getting the content that is similar to the summarized note. Figure 3 represents the NLP process of the Smart Reference Finder.

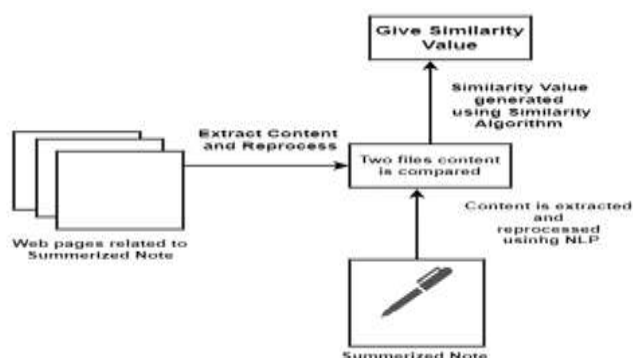
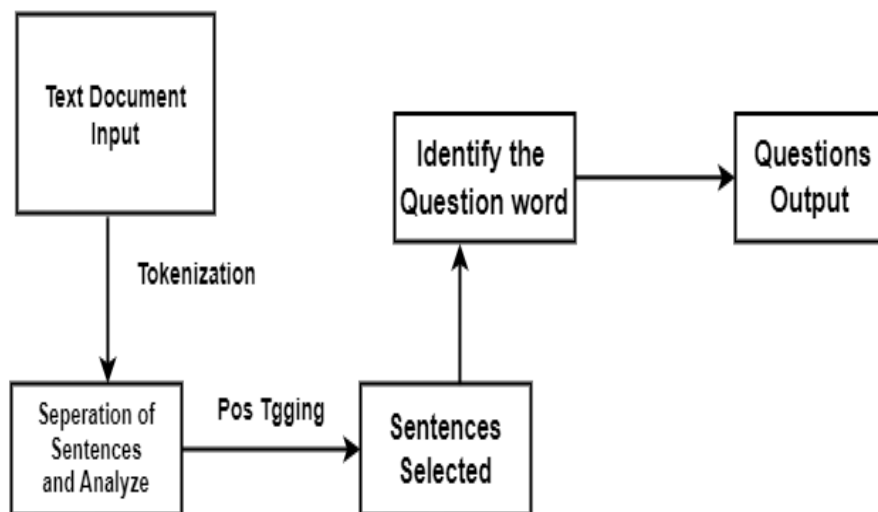


Figure 3: Reference Finding Process

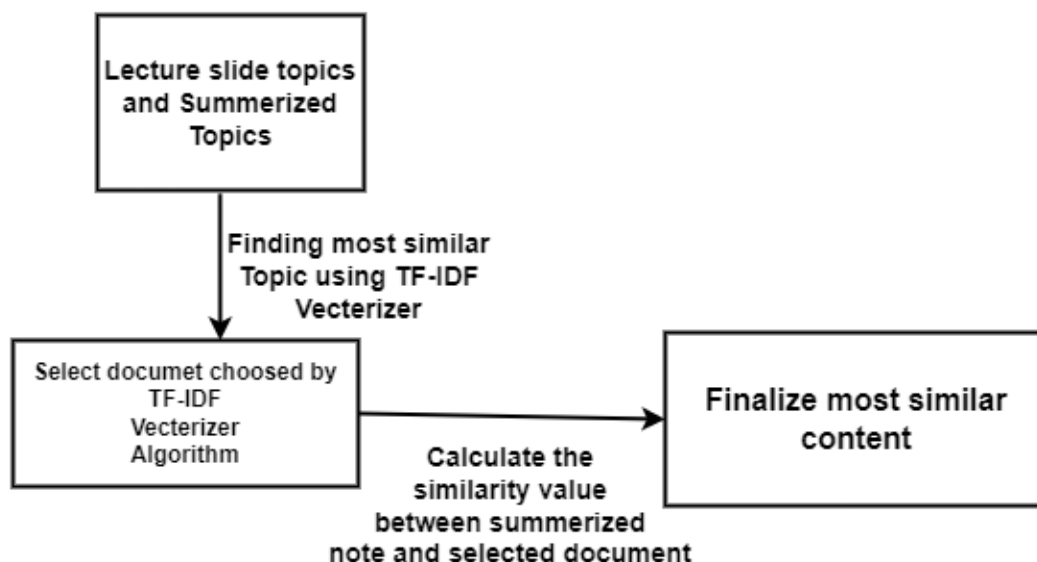
Smart Question Presenter

When the students are referring to the summarized note, they may want to refer to some questions and answers that are related to some topics in the note. So, the students need to find sample questions by searching in search engines manually. It is time consuming process for students and sometimes the sample questions or answers are not found. The Smart learning System is having a method that can assist the students in directly finding the questions and answers. Smart Question Presenter component is used to give sample questions and correct short answers regarding the relevant lecture part. The input to this component is the summarized text file and the output is questions that are generated by analysing each sentence. The next main goal is to find for online short answers using keywords by searching. First from the cloud database the student will select a summarized note which was stored before. After that student can select the questions which are generated using the summarized note before and synchronism for given keywords is checked by the system. An automatic Google search made that uses machine learning algorithms. In the end, the most suitable answers for given questions are selected by student. Figure 4 represents the NLP processes of the Smart Question Presenter.



Smart Slide Matcher

Smart Slide Matcher can assist the students in finding the related lecture note and also navigate to the relevant content directly. It is time saving process to the students and can study smartly by comparing summarized notes and lecture notes. students can choose the relevant note from the summarized note is available in the database. There are some steps in the process of data pre-processing. they include Tokenization, removing stop words, Stemming & Lemmatization. Then the next step is using an algorithm to find the similarity of summarized note's topic and topics of lecture note which generates the percentage value of similarity. If the value generated by the algorithm is greater than 75% and it has only one output, it can be taken as a result. If the value generated by the algorithm is less than 75%, another process will start to find the similarity of whole content related to the selected topic. This content will be again compared with all the contents of generated test results which has been scored less than 75%. The output that is best will be chosen according to these two results. If the calculated output value is less than 45% after all the above processes, another process that uses the Jaccard similarity algorithm, Count Vectorizer method, and Word Embeddings method, will be started. This occurs when compared sentences have the same meaning but different words. Figure. 5 represents NLP processes of the Smart Slide Matcher.



III. MODELLING AND ANALYSIS

One of the key benefits of smart learning assistant system is that it reduces the challenges of students and helps in providing the effective learning methods. By analysing current difficulties of the students in the learning process the technologies and algorithms can be used to model the system to overcome all the difficulties. The challenges of note taking can also solved [2]. This section the different technologies and algorithms that are used for making the smart learning system effective to the students.

Tokenization-Tokenization is a simple process which takes the raw data as the input and converts it into a data string that is useful. also, an important part of the NLP process. Tokenization is used natural language processing to split paragraphs and sentences into smaller units that can be more easily assigned meaning in natural language processing. Different tools of tokenization include the following white space tokenization, NLTK word tokenize, word and sentence tokenizer.

Summarization-The voice converted [4]to text should be summarized for making the notes more effective and easier for learning. This can be done by using the TF-IDF Algorithm. This algorithm finds the product of term frequency and inverse term frequency of a term and if the value is less than then the term is occurring many times else is not occurring multiple times. Based on this the sentence score is also calculated the score is greater than the Threshold it is included in the summary else it is discarded. This helps in summarization or the notes for that helps the students in revising concepts easily and effectively.

Cosine Similarity Algorithm- Students finds the different references for the selected note but it difficult to choose best from them to overcome this problem the similarity algorithm is used to find the best similar content [11]. The Cosine similarity algorithm generates similarity percentage between the web results and the selected note. By this the student can have an idea of the similarity between them and choose the best reference to study.

Question and Answering System-Question and Answer system helps in finding the relevant answers for the questions generated by the Smart Question Presenter. Question and answer system provides the answers for the questions that can assist the studying in learning the concepts easily. This System has steps such as Classifying the query and Extraction of relevant answer to the query. The classifying the query classifies the query based on the concept and the then relevant answers will be extracted to the student.

IV. LITERATURE REVIEW

A literature review is a summarization of previous researches on a topic. Literature review allows us in building the understanding and familiarity of current research in a particular field before creating a new invention. Literature review assist us to choose the methods that are simpler and efficient methods by avoiding the methods which were complex and difficult for getting the desired results. Use of smart learning concepts are increasing as improving the teaching and learning methods. During the class time some student's feel it difficult to listen, understand and take notes at a time to solve this problem a system was developed that has a note taker component that assist in making the notes easily [1]. This system also helps the students to find related references from the web for studying during the exam times. It also presents the questions related to the concepts that improves the learning method of the student Note taking typically assumed as a componential skill that students may or either may not have and if they do not have this skill they should through trial and error. The students face co challenges and in making the notes. Research addresses the various challenges faced by the students in making the notes in twenty-first century [2]. Another research discusses the barriers in the smart learning and also about the pillars of smart learning. The pillars of smart learning include accessibility, flexibility, affordability [3]. System was made for transcription of lecture speech [4]. A large cuprous of lecture talks are collected and initial baseline models were defined. As the speaking rate is generally faster and changes frequently therefore the sequential decoding and speaking rate decoding strategies were made. Phone models and decoding parameters are applied. These strategies achieved the improvement in the automatic transcription or real lecturer speech. Summarization is defined as a brief representation of input text such that the output is covering the most important concepts of the source in a compressed manner. Text Summarization is technique that is emerging for understanding the any type document's main purpose. For visualizing a large text document in a short duration and in small visible area like PDA screen, summarization is providing a greater flexibility. [5] This paper is presenting innovative unsupervised methods for the process of automatic sentence extraction they use graph-based ranking algorithms and shortest path algorithm. Automatic text summarization technology is

playing an important role in retrieval of information and classification of text, and it may provide a solution to the problem information overload. The process of reducing the size of a text while preserving its information content is known as text summarization. [6] The paper proposes an approach that is sentences clustering based summarization. The proposed approach includes of three steps: it clusters the sentences based on the sentences semantic distance in the document, and then on each cluster the accumulative sentence similarity is calculated based on the multi-features combination method, at the end with some extraction rules it chooses the topic sentences. In this paper [8], they study the problems of identifying the "Information Unit" of relevant pages that consists of all the input keywords as their answers. They model a set of most related web pages as a tree, where the nodes in the tree are the web pages and the edges are the links between the web pages. For Improvement in the search effectiveness, they propose an efficient and effective LCA-based algorithm that helps to identify those subtrees which are most related to the given input keywords. They have conducted a set of extensive experiments on the proposed algorithm. Their experimental results show that their method is achieving high search performance. The similarity between the Stings can be calculated with the help of cosine similarity algorithm [11]. The algorithm compares the two documents and finds a similarity percentage between them. This can assist the students in finding the best and similar references on the web pages related to the summarized notes. The questions can be presented based on the input note. [14] this paper, presents a system that helps in generation of questions automatically from natural language text using discourse connectives. The the usefulness of the discourse connectives is explored. The divides the Question Generation task is divided into content selection and question formation. The system will be evaluated manually for syntactic and semantic correctness.

V. RESULTS AND DISCUSSION

Smart Learning System is designed based on the four different components below. The lecture will be submitting the lecture recording to the Smart note taker component and it generates output as a summarized note. This summarized note can be used as the input to Smart reference finder, Smart question presenter, and Smart slide matcher components generate various results that can assist the students in effective learning

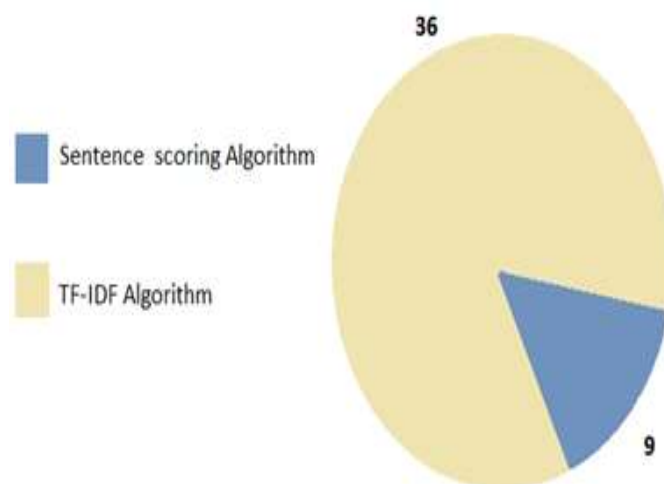


Figure 6: Results of survey for selection of algorithm

It was found that the real time converting method is best suitable because it is time efficient that helps in saving the time. In the speech-to text converter, the Word Error Rate of the transcriptions was reduced to 30-35% from 40-45% which is better than the normal rate. Then the Text Summarizer was used and a research survey has been conducted to select the best abstractive text summarizing algorithm for the text summarization process. As the result of that survey the TF-IDF algorithm was selected as most effective and consistent algorithm with the choice of 80% of respondents. The similarity algorithm is generating a similarity value and the student can assist the student to get an idea of the web results are how much similar to the selected summarized note that is stored in database. The similarity between the two contents can increased when the similarity percentage is higher. As shown in Figure. 7, By inputting the summarized note and the reference page the first output (63.072%) is generated. By inputting the same document as the two inputs the second output (99.999%) is generated.

```
In [1]: runfile('C:/Users/HP/Desktop/python/Extract Full site details.py', wdir='C:/Users/HP/Desktop/ python')
Similarity between two documents:
0.630726184889235
In [2]: runfile('C:/Users/HP/Desktop/python/Extract Full site details.py', wdir='C:/Users/HP/Desktop/ python')
Similarity between two documents :
0.9999999999999997
In [3]: runfile('C:/Users/HP/Desktop/python/Extract Full site details.py', wdir='C:/Users/HP/Desktop/ python')
Similarity between two documents :
0.031000014403072183
```

Figure 7: The similarity value generated by Smart Reference Finder

The question generation begins by the process of analyzing the selected summarized note from the database keywords are extracted. By analysing each sentence, system generates WH questions such as Who, what, When, Where & Why and answers based on these keywords, the system will give the correct answers automatically. Figure 8. Shows A sample question and answer system.

| |
|-------------|
| Question |
| What is css |
| Answer |

Figure 8: Question and answer System

The slide matching process will begin by the process of analysing the topics and the contents of the selected summarized note that was stored in database. Now the data will go under re-processing process and then the application is now compared the content of the summarized note and the lecture slides with the help of similarity algorithm.

Figure. 9 shows the calculated percentage of similarity between the summarized note and selected lecture.TF-IDF vectorization models and Gensim models are used to implement the smart slide matcher.

```
: print(f'Average similarity float: {float(sum_of_sims / len(file_docs))}')
print(f'Average similarity percentage: {float(sum_of_sims / len(file_docs)) * 100}')
print(f'Average similarity rounded percentage: {percentage_of_similarity}')

Average similarity float: 0.20402128994464874
Average similarity percentage: 20.402128994464874
Average similarity rounded percentage: 20
```

Figure 9: Content analyzing function output

VI. CONCLUSION

The Smart Learning Assistant System" an application that can be developed using a combination of various latest technologies which can assist the students to overcome regular problems and they can effectively learn and make the revision of the lectures done in the class rooms. The main goal of the application is to assist the students to making summarized notes using the smart note taker, find references online using the smart reference finder, generation of the questions related to lecture using the smart Reference Finder, and matching the notes and lecture slides making the learning process easy, effective and consistent

For achieving this objective, System has the four different components which are Smart Note Taker, Smart Reference Finder, Smart Question Presenter, Smart Slide Matcher. The future goal is in to deploying the application for active usage in the educational institutes. The system will be having the different levels of functionalities

VII. REFERENCES

- [1] Edu Easy - Smart Learning Assistant System A. Karunasena, P. Bandara, J.A.T.P. Jayasuriya, P.D. Gallage
- [2] J. M. S. D. Jaya Sundara, L. A. P. Y. P. Nuwan Jaya 2021 IEEE.
- [3] J. van der Meer, "Students' note-taking challenges in the twenty-first century: Considerations for teachers and academic staff developers," Researchgate.net, February 2012
- [4] M. A. Awar, "Pioneering smart learning". [Online]. Available: <https://www.ellucian.com/emea-ap/insights/pioneering-smart-learning>. [Accessed May 4, 2020].
- [5] T. Kawahara, H. Nanjo and S. Furui, "Automatic transcription of spontaneous lecture speech," IEEE Workshop on Automatic Speech Recognition and Understanding, 2001. ASRU '01., Madonna di Campiglio, Italy, 2001, pp. 186-189.
- [6] K. S. Thakkar, R. V. Dharaskar, and M. B. Chandak, "Graph-Based Algorithms for Text Summarization," 3rd International Conference on Emerging Trends in Engineering and Technology, Goa, 2010, pp. 516-519.
- [7] P. Zhang, and C. Li, "Automatic text summarization based on sentences clustering and extraction," 2009 2nd IEEE International Conference on Computer Science and Information Technology, Beijing, 2009, pp. 167- 170.
- [8] N. Madhuri, and R. Ganesh Kumar, "Extractive Text Summarization Using Sentence Ranking," 2019 International Conference on Data Science and Communication (IconDSC), Bangalore, India, 2019, pp.
- [9] X.Song, J. Feng, G. Li, and Q. Hong, "LCA-based Keyword Search for Effectively Retrieving Information Unit from Web Pages," Ninth International Conference on Web-Age Information Management, Zhangjiajie Hunan, China, July, 2008.
- [10] P. Shravanthi, and B. Srinivasu "Semantic similarity between sentences" MTech Scholar Dept. of Computer Science and Engineering, Stanley College of Engineering and Technology for Women, TelanganaHyderabad, India, International Research Journal of Engineering and Technology.
- [11] X. Yang, J. Liu, "Semi-supervised learning of dialogue acts using sentence similarity based on word embeddings" Department of Electronic Engineering, Tsinghua University Beijing100084, China.
- [12] Dario Radecic, "Calculating String Similarity in Python", towardsdatascience.com, Oct.30,2019. [Online]. Available: <https://towardsdatascience.com/calculating-string-similarity-in-python276e18a7d33a>. [Accessed: Aug.10, 2020].
- [13] Pandey, and K.C. Rajeswari, "Automatic Question Generation Using Software Agents for Technical Institutions", International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970) Volume-3 Number-4 Issue-13, December-2013.
- [14] M. Agarwal, R. Shah, and P. Mannem, "Automatic question generation using discourse cues", In Proceedings of the 6th Workshop on Innovative Use of NLP for Building Educational Applications, Association for Computational Linguistics, pp. 1-9, 2011.
- [15] S. Sarkar, V. R. Madasu, Baala Mithra Sm, Subramanya Vrk Rao "NLP Algorithm Based Question and Answering System", Seventh International Conference on Computational Intelligence, 2015, Modelling and Simulation, pp. 97-101.