



Automating The Technical Interview Using Semantic Similarity Matching, Speech Recognition and BERT

Dr. G. Fathima ME., PhD¹., Arun Sundar P², Bhavya V³, Kabilan R⁴

Professor, Department of CSE, Adhiyamaan College of Engineering (Autonomous), Hosur, Tamil Nadu, India¹

UG Scholars, Department of CSE, Adhiyamaan College of Engineering (Autonomous), Hosur, Tamil Nadu, India^{2,3,4}

Abstract: Recruiting freshers when they are in final semesters through off-campus drive is a time-consuming process as it needs more recruiters to assess them. It requires more recruiters to filter the best candidate. The average time to assess a candidate is 10 to 20 minutes. Then it will take more time to shortlist the best candidates to the company. So with this paper, tried to automate this process of hiring using Natural Language Processing, Bidirectional Encoders, Speech Recognition and Semantic Matching and Supervised Similarity Measuring Techniques. The answers provided by the candidates can be recorded and pre-processed using speech recognition and Bidirectional Encoder Representations from Transformers (BERT). The recorded answers can be checked and matched through semantic sentence matching using the question and answer database designed by the recruiters. This web application would be the fastest and perfect way to hire candidates to the company.

Keywords: semantic similarity, cosine similarity, soft cosine similarity, word embedding, Euclidean distance, Cosine distance, BERT, Soft Cosine Similarity, TF-IDF Vectorizer.

I. INTRODUCTION

Automated Face to Face Technical interview is a method of interviewing candidates where the recruiter assesses the candidate without conducting the interview in real-time. In real-time interviews, the interviewers or candidates are required to travel or spend a significant amount of time in setting up and completing the process. While in an automated asynchronous technical interview[4], the recruiter can simply send questions to the candidate, which candidates can answer and send back whenever they can and at their own comfort.[12] Automated interview greatly reduces the workload of recruiting teams, allowing for greater flexibility, automation of administrative functions, and greater efficiency. This reduces some unnecessary works like managing employee details manually and phone call personally to get availability to attend the interview etc. An automated video interview or 'asynchronous video interview' is when a candidate answers questions asked by a recruiter in the form of a video. Whether scheduling phone screens, event interviews or complex interview days, automated interview scheduling ensures the process is optimized for all participants.

II. EXISTING SYSTEM

A. Face-to-Face Conventional Interview

The traditional way of recruiting the candidates is through the face-to-face interview. In recent days due to pandemic situation the world turns into remote, so the corporates need to find a way to hire the candidates through online so this is how online interviews are introduced. Before the online mode there are always been two conventional ways of hiring the professionals through on campus.[11]

B. On-Campus Placement Drive

These drives are conducted on campus, reducing your effort of having to visit companies for placement drives. Colleges often have placement committees which are responsible for the companies coming on-campus. These software and IT companies have their own processes of hiring, most of which have the basic steps of pre-placement talks, aptitude and technical Assessment, group discussions and personal Face to Face Final interview.

C. Off-Campus Placement Drive

Unlike the on-campus drive, the recruitment process of various companies is conducted off the bounds of your college campus. They most often take place in a common place which has space enough to accommodate a number of students. Also called Pool-campus drives, students from different colleges visit the location provided in the details of the drive. Further processes take place there. The level of competition is generally way higher in off-campus drives than in the



former because of the number of candidates who sit for the placement drive. Every company lists their eligibility criteria for the various profiles on offer. Along with that, the various steps of the placement process are declared beforehand too.

D. Interview

This is the last leg of the selection procedure. Interviews are conducted to judge the candidates in a number of aspects. These range from confidence and knowledge to skills and the value he/she will add to the company. Often divided into two, the interviews are of two major kinds: Technical and HR. The HR interviews are conducted to know the candidate better, their strengths and weaknesses, how enthusiastic they are to get hired and the value they'll add to the company. Whereas technical interviews judge the prospective candidate's mettle in their core subject.

E. Online Interview

Due to pandemic online interviews are introduced, the candidate can attend the interview from anywhere he belongs to, so the distance does not stop the candidate from attending the interview but the real deal starts with the time the candidate should be present at the schedule time and also internet connection plays a major role in the interview specially the candidates from rural places face the consequences. Even though the candidate is met with the skillsets required for the company these issues which leads to lose the opportunities.[1]

III. PROPOSED SYSTEM

A. Overview

The proposed system will evaluate candidates marks automatically by recognizing candidates recorded video and audio using speech recognition, semantic matching and BERT. It will accurately find the similarity between the stored actual answer and answer from the candidate. It will also reduce marks for grammatical errors by checking word formation. Marks for each question will be stored for future reference and it can be consolidated for further calculation of marks.

B. Flow and Architecture Diagram of Proposed System

The employer will create and schedule the interview to the candidates using Design Interview Portal in his logged session. As mentioned in the fig (1) all the candidates will be scheduled for the interview and then after they attended the virtual interview their recorded videos will be processed automatically using Representational Transformer. Here this system used BERT Encoder which is created by google to embed sentence into numerical value by extracting the core meaning. Extracted core meaning will get encoded as vectors and similarities between those two sentences will be calculated automatically using Supervised similarity Measuring Techniques. To check the similarity between these vectors cosine similarity, soft cosine similarity and TF-IDF Vectors, dot product and Euclidean distance measuring strategies are used. The answer recorded from the candidate will also be checked for grammar errors for proper word and sentence formation Because the candidate should not record only the keywords related to the answers.

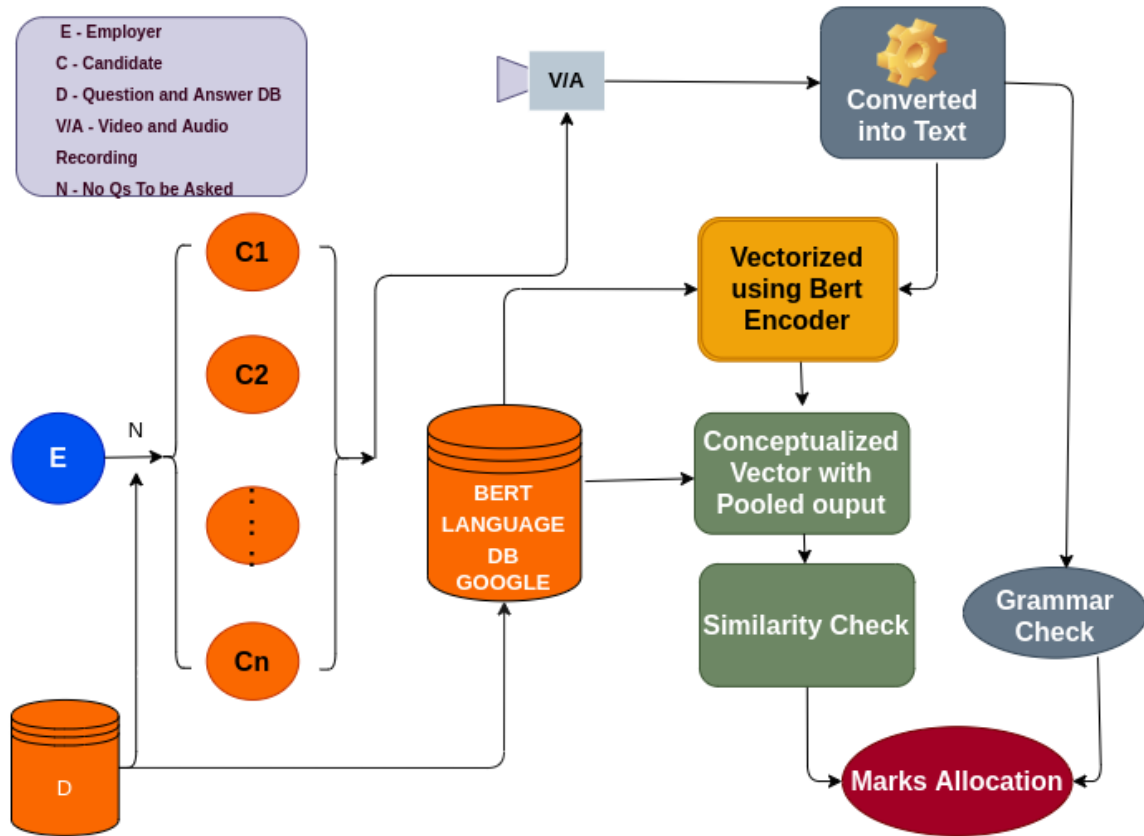


Fig. 1 System Flow Diagram

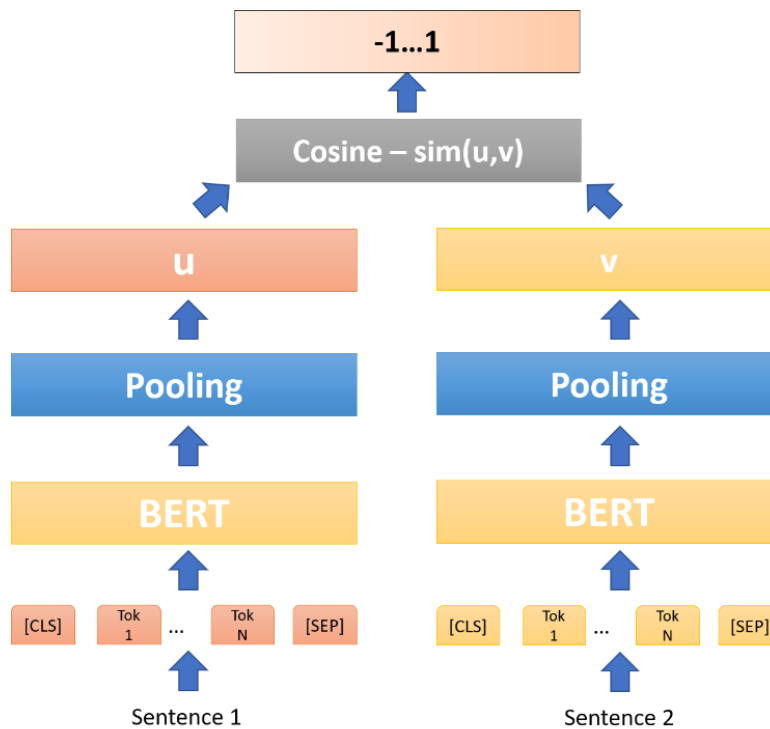


Fig 2. BERT Architecture

After the candidate has recorded his answers for all the questions, the recorded video will be processed step by step. At first the video will be converted into audio format and then the audio will be further converted into text format and it will



be stored in JSON file which has mapped to the candidate details. The JSON file consists of all the details related to his interview like questions to be asked, question's actual answer, the answer recorded by the candidate and marks for each questions. For sentence similarity match checking the actual answer and the answer recorded by the candidate will be compared using BERT. As seen in fig (2) Both the sentences will be tokenized using separators for processing each token as important key words. After tokenizing the sentences BERT will find the concept meaning of the sentence and then vectorize the sentence using Encoder Representation. After converting those sentences into single dimensional numerical value. Using those numerical value similarity can be measured using Supervised Similarity Measuring Algorithms.

C. The modules used in this proposed system are splitted up and explained individually bellow

1) Candidate and Employee Portal

The Home page of the website has two options to choose. The first one is the Employee Portal. Employee can able to login to his dashboard to view the candidate database, schedule the interview and evaluating marks. The View Candidate Database option has the table consists of all the candidate details about his Academic performance. The Interview Scheduling section has a form which will ask for the total of number of questions to be asked to the candidate and in which domain the question should be asked and the time to answer each question, CGPA, batch of candidate for eliminating candidates who didn't met the eligibility criteria. The Mark evaluation section will evaluate after the candidate attended his Interview. Once all the Candidates finished their interview, the employee will evaluate the marks by simply clicking a button. The second one is the Candidate Portal. Candidate can able to login to his dashboard and view his personal details and academic details. If the candidate met with the eligibility criteria, he will get the application status as scheduled. Once the interview is scheduled, the Start Interview button will appear in the candidate dashboard. Then he can start the interview. All the questions will be recorded one by one and it will be stored in the backend for further processing.

2) Questions and Answers Database

The List of questions to be asked to the candidate will be collected from a large question and answer database. This Question and Answer database is in the python dictionary format, so that it can be picked randomly using the dictionary's key. Questions and Answers are stored separately in the domain format, so for each domain or language there will be questions and answers dictionary. For Example: Python has two dictionaries. The first one is python_quest_dict and the second one is python_answer_dict. The question dictionary consists of question number as key and question as value to that key. Likewise the answer dictionary has answer number as key and answer as value. So every question in question dictionary has the answer in answer dictionary mapped with the question and answer number.

3) Interview Scheduling

The Employee will give the CGPA, batch, department and number of questions to be asked in the design interview section. Based on the inputs given by the employee, the system will generate n number of questions from the desired or predefined domain and store those questions and answers for each every candidate in a json file. The json file consists of the candidate's email id as key and their name, questions to be asked, answers for that questions, candidate answered answer and finally his marks after the evaluation.

4) Filtering Candidates

The filtration of candidates can be done using CGPA , batch, department and number of backlogs.

5) Preprocessing The Recorded Video

Questions will be displayed one by one to the candidate while he is attending the interview. Each question's video answer will be recorded in separate folder called answers_dir where each candidates answers will be recorded and stored in each folder named as email id of the candidate.[10]

6) Conceptualize Sentence and Transform using BERT

After the video has been recorded, those videos will be preprocessed using BERT, which is bidirectional encoder representational Transformer. It will find the concept meaning of the sentence and it will encode the concept meaning into a vector. This vector is output given by the BERT model.

7) Semantic Similarity Matching

a. Manual Similarity Measure

To Measure the similarity between two Sentences, you need to combine all the feature data for those two examples into a single numerical value. The smaller the numerical difference between sizes, the greater the similarity between input features. This kind of Manual similarity measure is called a manual similarity measure. As the data that needs to be processed



becomes more complex, creating a manual similarity measure becomes much harder. When the data becomes complex enough to process, it can't be able to create a manual measure. This is the time to change the manual measure with an automated method, Here Supervised Similarity Measure algorithm called BERT is used. It can handle complex data without any issue and can convert those features into useful numerical values.[3]

b. Supervised Similarity Measure

Instead of comparing manually-combined feature data, Reducing the feature data to Vector representations using embeddings and then compare those converted embeddings. Embeddings are generated by training a supervised deep neural network (DNN) on the feature data itself.[6] The embeddings space has less feature data to a vector in an embeddings. The embedding space has fewer dimensions than the feature data in a way that captures some latent structure of the feature data set. The embedded vectors for similar examples, such as You-tube and other social media videos watched by the same users, ends up close together in the embedding space.[2] The output from the BERT which is in the vector form will be compared with the actual answers vector using some similarity measuring formulas like cosine similarity, soft cosine similarity, Euclidean distance, dot product etc.[5] But in our project cosine similarity gives better results than the others so the system is fixed with this as a similarity measuring formula.

8) Grammar Check

For Grammar checking in the candidate's answer, python language tool had been used. It contains the English vocabularies and sentence rules for English. Using that number of grammatical errors will be found and marks will be reduced accordingly.

9) Consolidation of Marks

Marks will be evaluated using BERT and Similarity Matching Algorithm. Each question's mark will be stored in the candidate's profile which will be in the format of JSON. Which can be further consolidated for total results.

10) Updating marks in the Candidate Database

The consolidated marks will be updated to each and every candidate one by one in the database so that the employer will be able to view the performance of all the candidates.

11) Advantages

- The primary advantage of this technique is that it allows the candidate to attend the interview whenever he is free.
- Simple and Easy to use the application to record answer and checking the results.
- No need of large number of Talent acquisition team to testify the candidates.
- Can be used to assess more candidates in a constant time so the time required for assessing automatically is considerably less.
- For every video, answers are recorded for future references so that it can be re-analysed if there is a need.

IV. COMPARISON OF SIMILARITY MEASURING ALGORITHMS

For Testing the quality of similarity measurement, three different similarity measuring algorithms are used. All the three algorithms are based on the statistical calculation to identify the relation between those vectors. The three algorithms are Cosine Similarity, Soft Cosine Similarity and TF-IDF Vector Similarity Measuring Algorithms.

To Test these similarity measuring algorithm 10 sample questions are generated from the questions and answers database and compared with the recorded video by all these three algorithms and the results are as follows in Table (1) and Table (2).

Table 1 : Answer Comparison between Actual and Recorded Answer



S. No	Actual Answer	Recorded Answer
1	Decorators are used to add some design patterns to a function without changing its structure. Decorators generally are defined before the function they are enhancing.	decorators for you should to add some design patterns to a function without changing its structure iterators generally to a defined before the function they are in anything
2	It is an environment variable which is used when a module is imported.	environment variable which is used Immoral is imported
3	Slicing is used to access parts of sequences like lists, tuples, and strings.	slicing is used to access sequence like list tuple and sings
4	The .py files are the python source code files. While the .pyc files contain the bytecode of the python files. .pyc files are created when the code is imported from some other source.	the Python source code file file are created when the code is imported from some other source
5	An interpreted language is any programming language which is not in machine-level code before runtime. Therefore, Python is an interpreted language.	an interpreter language isn't any programming language please it is not mission level code before run time therefore the nation interpreted language
6	Yes. Python is a case sensitive language.	yes Python is a case sensitive language
7	PEP stands for Python Enhancement Proposal. It is a set of rules that specify how to format Python code for maximum readability.	I don't know about papeete
8	Memory management in python is managed by Python private heap space.	memory management in Python is managed by Python private as space
9	Keywords in python are reserved words that have special meaning.	keyboard in Python are required words that special meaning
10	The common built in data types in python are numbers,List,Tuple,String,Set,Dictionary,Boolean.	The common built in data types in python are numbers,List,Tuple,String,Set,Dictionary,Boolean.

*Note – recorded answers are automatically converted to text from audio so it may have spelling mistakes or wrong word formation.

Table 2 : Similarity Measurement of Different Algorithms

Similarity Matching Algorithms	Qs:1	Qs:2	Qs:3	Qs:4	Qs:5	Qs:6	Qs:7	Qs:8	Qs:9	Qs:10
Cosine Similarity	71	58	85	66	59	96	0	89	77	100
Soft Cosine Similarity	75	63	70	55	70	80	2	77	60	100
TF-IDF Vector Similarity	79	66	31	31	59	99	0	90	50	100

Marks calculated automatically based on the comparison between the actual answer and the recorded answer using BERT for Encoding it into vector and then similarity has been measured using the three algorithms namely cosine similarity, soft cosine similarity and TF-IDF Vector Similarity.[9]

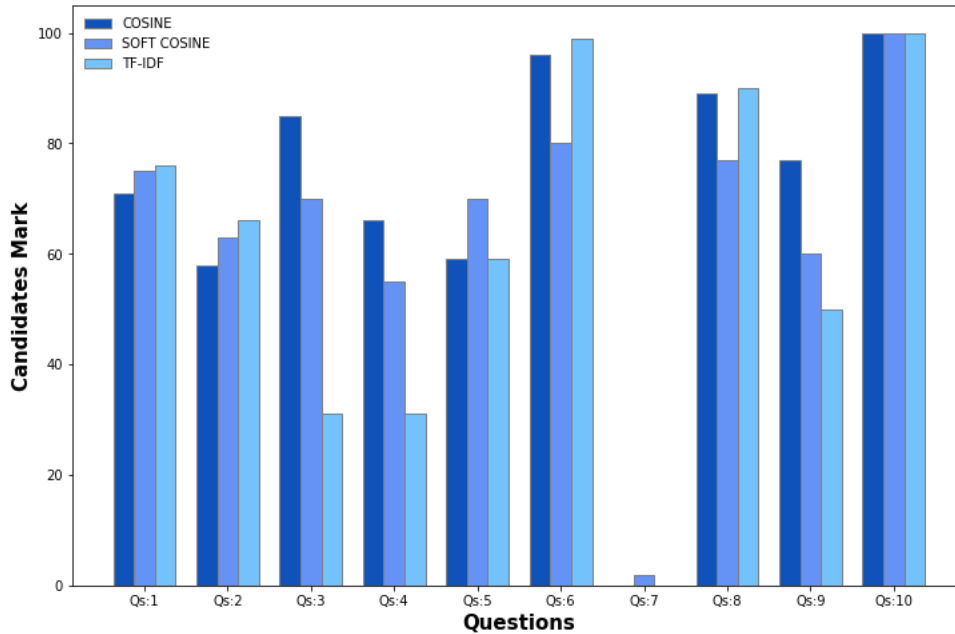


Fig 3. Marks evaluated using Three Similarity Measuring Algorithm

All the three similarity matching algorithms are performing almost the same but cosine similarity gives us slightly better results than the others. As mentioned in (Fig 3) all those marks are plotted in the bar graph for references of difference between the algorithms. For the question number 7, the candidate has recorded as “I don’t know ”. The cosine Similarity given mark as zero, the soft cosine given the mark as 2 and the TF-IDF given the mark as zero. Hence the Cosine Similarity and the TF-IDF has calculated the similarity better than the soft cosine similarity. And for the last question, answer is exactly same as the actual answer that’s why all the three algorithms calculated the similarity as 100 percentage match.

V. RESULTS OBTAINED

Using Bidirectional Encoder Representations from Transformers the core meaning of the sentence has been identified and encoded in single numerical value. Both the actual answer and the recorded answer are matched using three popular similarity matching algorithms and their results are as follows in Fig 4.

- Cosine Similarity – 70.1%
- Soft Cosine Similarity – 65.2%
- TF-IDF Vectorizer and Similarity – 60.5%

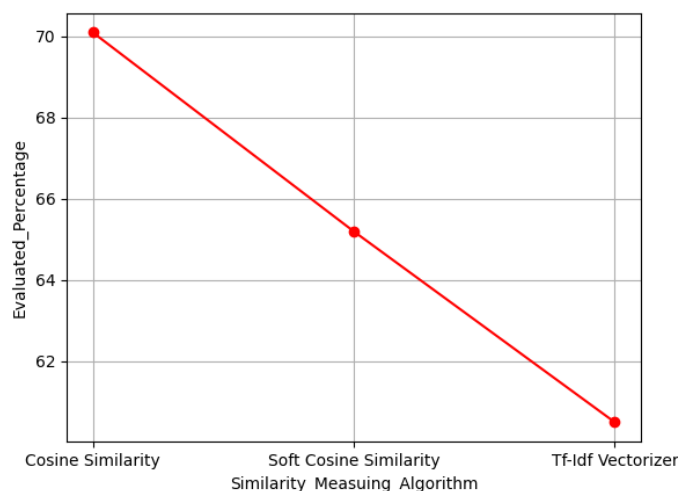


Fig 4 Performance Measure of Similarity Matching



VI. CONCLUSION

With the use of Web RTC, BERT and Supervised Similarity Measure has been used to find the similarity between the actual answer and the answer from the candidate. Using these technologies mentioned above the system successfully found the similarity between those sentences and for further evaluation the same has been updated to the candidate database with his mark details. This technology can be used in many real time applications to find similarities and many related problems can be automated using this similarity matching approach.

REFERENCES

- [1] Chen, J.; Ma, L.; Chen, X.; Jie, Z.; and Luo, J. 2019. Localizing Natural Language in Videos. In AAAI.
- [2] Priyadarshini Patil, Chandan Rao, Gokul Reddy, Riteesh Ram & S. M. Meena, " Extractive Text Summarization Using BERT", springer 10 Jan 2022.
- [3] HUANG Cheng-Hui¹, YIN Jian, HOU Fang², " A Text Similarity Measurement Combining Word Semantic Information with TF-IDF Method ", Chinese journal of Computers, 2011.
- [4] John Tang, Jennifer Marlow, Aaron Hoff, Asta Roseway, Kori Inkpen and Chen Zhao, " Time travel proxy: using lightweight video recordings to create asynchronous, interactive meetings ", CHI 2012, May 5–10, 2012.
- [5] Alfirna Rizqi Lahitani; Adhistya Erna Permanasari; Noor Akhmad Setiawan, " Cosine similarity to determine similarity measure: Study case in online essay assessment ", ieeexplore, 29th Sep 2016.
- [6] Yaokai Cheng, Ruoyu Chen, Xiaoguang Yuan, Yuting Yang, Shan Jiang and Bo Yang, " Overview of Long-form Document Matching: Survey of Existing Models and Their Challenges ", 12 Nov 2021.
- [7] Hao Zhang, Aixin Sun, Wei Jing, Liangli Zhen, Joey Tianyi Zhou and Rick Siow Mong Goh, " Natural Language Video Localization: A Revisit in Span-based Question Answering Framework ", IEEE Xplore, 16th June 2021.
- [8] I. Naim, M. I. Tanveer, D. Gildea and M. E. Hoque, "Automated Analysis and Prediction of Job Interview Performance," in IEEE Transactions on Affective Computing, 1 April-June 2018.
- [9] J. Xue, X. Tang and L. Zheng, "A Hierarchical BERT-Based Transfer Learning Approach for Multi-Dimensional Essay Scoring," in IEEE Access, 2021.
- [10] G. Szaszák, Á. M. Tündik and K. Vicsi, "Automatic speech to text transformation of spontaneous job interviews on the HuComTech database," 2011 2nd International Conference on Cognitive Infocommunications (CogInfoCom), 2011.
- [11] P. Senarathne, M. Silva, A. Methmini, D. Kavinda and S. Thelijjagoda, "Automate Traditional Interviewing Process Using Natural Language Processing and Machine Learning," 2021 6th International Conference for Convergence in Technology (I2CT), 2021.
- [12] L. Hemamou, G. Felhi, J. -C. Martin and C. Clavel, "Slices of Attention in Asynchronous Video Job Interviews," 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII), 2019,.