Grenze International Journal of Engineering and Technology, June Issue

Enhancing Career Readiness: An Interview Grooming Model for College Students using Machine Learning

Anurag Patil[1], Neha Pol[2], Pratik Dhane[3], Yash Sarda[4] and Ravindra Murumkar[5]
[1-5]SCTR's Pune Institute of Computer Technology, Pune, India
Email: anuragdpatil@gmail.com, {nehapol40, 21pratikdhane, yashsarda2002} @gmail.com, rbmurumkar@pict.edu

***Abstract*— In today's competitive job market, the transition from college to a successful career hinge on a student's ability to navigate technical interviews with confidence and competence. Traditional interview preparation methods, although invaluable, often lack the adaptability required to address the unique needs and experiences of individual students.**
This paper solves this problem by presenting a comprehensive overview of an Automated Interview System that utilizes NLP and deep learning techniques to mimic the process of job interviews. The system comprises 3 modules, the first one being the resume analysis system, which is built on state-of-the-art NLP architecture. The subsequent module utilizes the output from the resume model to generate a set of relevant interview questions through semantic analysis by web-scraping. One of the most innovative parts of this system is its ability to generate dynamic questions based on the candidate's previous response using LSTM. The last module assesses the candidate's responses and provides a summary of the complete interview process, providing expected answers to asked questions, and highlighting the areas of improvement required. This research outlines the platform's development and its impact on students' readiness, highlighting the potential to transform career readiness for college students.

***Index Terms*— Interview Grooming, Machine Learning, Web Scraping, Dynamic Question Generation, Resume Analysis and Long Short-Term Memory.**

I. INTRODUCTION

The transition from college to a successful career involves mastering the art of technical interviews. However, the traditional methods of interview preparation often lack the adaptability and personalization required for each student's unique journey. Candidates often receive limited feedback on their applications, which can be frustrating and demotivating. To bridge this gap, we present an AI-enhanced interview Grooming Platform tailored for college students.

This paper explores the issue of interview readiness and how our platform addresses it. By leveraging machine learning and deep learning, our platform generates personalized interview questions based on a student's resume and job description. As students respond to these questions, the system adapts, offering tailored feedback and guidance, transforming interview preparation into a dynamic and engaging experience.

The project is divided into 3 modules. The resume module analyzes the candidate by extracting the candidate's resume and the provided Job Description, with the help of Natural Language Processing techniques. The output of the Resume Module is the list of the relevant skills/experience from the candidate's Resume. In the Interview

Grenze ID: 01.GIJET.10.2.533_1

module, questions are asked on the skills retrieved from the resume report. The initial questions are generated by web-scraping, followed by the dynamic question generation module, which generates questions in real-time based on the candidate's response to the previous question. The responses provided by the candidate are evaluated and a summary of the responses is provided with expected answers to these questions, with points highlighting the areas of improvement required.

In the sections that follow, we'll discuss the platform's de- sign, development, and evaluation, showcasing its potential to significantly enhance students' interview skills and confidence. Our platform offers a promising solution to equip college students with the skills they need to excel in the competitive job market

II. Literature Survey

"Ref [1]" suggests employing a range of NLPtechniques, including bigram, trigram, n-gram, and text classification, to effectively select top candidates from a candidate pool. The model leverages machine learning algorithms for the classification task. An advantageous aspect emphasized in the paper is the conversion of the textual corpus into a list of individual words. This procedure entails associating word variations with their lemmas, guided by concordance. The objective of transforming the text corpus into a word list is likely to streamline subsequent analysisand manipulation of the textual data.

"Ref [2]" discusses the use of Natural Language Processing (NLP) techniques and machine learning models to streamline the recruitment process in the growing Indian job market. The proposed system aims to extract valuable information from resumes, rank candidates, and provide personalized feedback for job applicants. The research outlines a system architecture, including a Resume Ranking System, and demonstrates promising parsing and machine learning model accuracy, making it a potentially valuable tool for both job seekers and hiring managers. Overall, the paper suggests that this intelligent recruitment system has the potential to revolutionize the hiring process in India, makingit more efficient and effective.

"Ref [3]", they developed interview bot uses machine learning to determine the competence level of a person. The research uses an Indonesian language dataset.To converse with human participants, the software used NLP technology. The results of the study can be used asthe basis for developing an interview bot that is closer toprofessional interviews. The interview bot is a valuable and reliable tool for conducting online interviews and supporting the assessment center process, especially with conditions of physical and social distancing constraints. It provides flexibility in terms of time and place for participants, and its process is delivered in Indonesia's Language.

"Ref [4]" addresses common issues that candidates have while participating in group interviews. The technology will have features like automated questioning systems and resume analysis. A Natural Language Processing model, which is particularly helpful in this process, would be used by the program to search for information based on the student's previous responses. The software would examinethe data collected after the evaluation to choose the best applicant for the open post. The project's goal is to streamline the hiring process as a result.

"Ref [5]" mainly focuses on the interview practice for the students rather than conducting company recruitments. This paper aims to deploy an interview practice chatbot for students to test their interview skills and get their scores analyzed at any given time of the day without the need for a physical interviewer. By conducting interviews using voice-based chatbots, assessing responses and emotions, and creating evaluations based on the responses, the purpose is to provide students with a long-term solution for practicing interviews online. The chatbot system can analyze variousaspects during the interview like emotions, correct answers, and whether the candidate used invalid means.

"Ref [6]" proposes a system that takes the candidate's personal, academic details in a database. Then the system will assign a unique ID to each candidate. When a candidate entershis unique ID, he will be authenticated and an interview session will be started. Chatbot will begin an interviewby greeting the candidate just like a human interviewer. Now, based on the candidate's details, both will start asking questions. Questions can be multiple choice, finding the output of the given program, based on current technology,etc. All responses will be checked and results will be storedin the database. Based on results and statistics classificationof candidates will be done. In this way, suitable candidatesfor a particular post in an IT company can be selected.

"Ref [7]" proposes a keywords-to-questions (K2Q) technique to assist users in creating and refining their queries. K2Q creates interview questions andword-refinement phrases using a set of input keywords. They offer a Client Question Motivation model to represent thecollaborative process of word and question generation to assess questions, suggest phrase improvements, and generate questions that may not have previously existed. This model can also suggest elaboration words that boost the entropy of the distribution of the inferred intent to further

strengthenthe accuracy of that inference. Experiment results show that 1) using a template increases proposal coverage by 6.6% to 16.2%. Second, the CQM model performs better than conventional language models in terms of ranking suggestions, and it has a recall that is almost 16%. Thirdly,the words suggested by the UII model are preferable as refining possibilities in comparison to a frequency-based baseline method. By using these clarifying phrases, K2Q outperforms the baseline strategy by 12.7% and ranks65.7% of the target questions in the top ten. These results show that our proposed UII model performs better at making helpful recommendations than the two-part techniques in K2Q.

"Ref [8]" provides a summary of the research on the automatic generation of questions from a given text. The authors of this study suggest reviewing a technique for creatingquestions out of text documents. The automatic question- generating system is created using several algorithms and approaches, as mentioned in the article. NLP is used to analyzethe text, and NER and SRL are used to find the significant relationship. The majority of the job is completed in English, and questions in the MCQ format are created. The automatic question generation system is a vast topic with room for research that can provide methods by identifying challenges and the kinds of questions that must be produced, like one- word answers and True/False questions.

III. METHODOLOGIES

A. Resume Analysis

The resume and job description for the candidate are inputs for this module. According to the model, the resume should include the following sections: Profile, Education, Experience, Skills, and Extracurriculars. For additional processing, each of these portions is extracted separately. The raw text that is retrieved from the resumeis cleaned using a regular expression library. This helps to clean up the raw text by removing the punctuation and weird words. The preprocessing is done in the second stage using NLP.

Tokenization is used for the resume's textual data. These tokenized data no longer contain stop words. These tokenized data are cleansed of stop words. After that, the tokens arelemmatized and stemmed. Only relevant tokens are taken into account for the analysis out of all the tokens.

The spacy NER model is equipped to recognize and extract items from resumes, such as names, firms, and talents, making it a useful tool for automating resume parsing chores and streamliningthe extraction of structured data. Using the same method, the requirements section of the job description is extracted. Finally, all of the resume's information—including experience, skills, profile, etc.—is entered intoa report and then forwarded to the following modules.

B. Web Scrapping Relevant Skills

The project then enters the skill-based web scraping module after the skill extraction phase. Customized web scraping tools now take center stage. These are configured using widely used frameworks like Scrapy and Beautiful Soup. These technologies' main objective is to automatically and methodically retrieve the questions that precisely match the previously determined skills from the resumes. This is significant since it signals the start of the project's more interactive phase. The algorithm will ask candidates skill- based questions based on the compiled job advertisements.The use of voice-based interactivity is a novel aspect of this. speech commands will be used to prompt candidates withquestions, and speech recognition technology will record their answers in real-time. These speech responses will thenbe converted into text for easy analysis and archiving. The voice-to-text feature improves user interaction and makes dataadministration easier. The conversational experience can be improved while also supporting effective data processing withthe help of the organized, analyzed, and stored transcriptions of the responses.

C. Conducting the Interview

The system then moves on to the phase of dynamic question generation after gathering and transcribing the candidates' responses. The objective of this stage is to generate context-aware questions based on the candidate's prior responses in order to engage them in a lively and interesting dialogue. Candidates will answer verbally, and their responses will then be transcribed from speech to text.If the accuracy of their answer surpasses 50%, the subsequent question will be generated based on the previous response. If the answer accuracy falls below 50%, the next question willbe drawn from the originally scraped set of questions. This approach aims to ensure a dynamic and adaptive interview process. Here is where LSTM, a kind of recurrent neuralnetwork (RNN), really shines. The LSTM model is frequently employed in natural language processing applications like textsynthesis because it is effective at handling data sequences. It does a fantastic job at identifying dependencies and patterns in sequential data, which makes it the best tool for comingup with questions that are contextually pertinent

to the current discussion. Sequence-to-sequence (seq2seq) problems can be used to model the dynamic question-generating process.

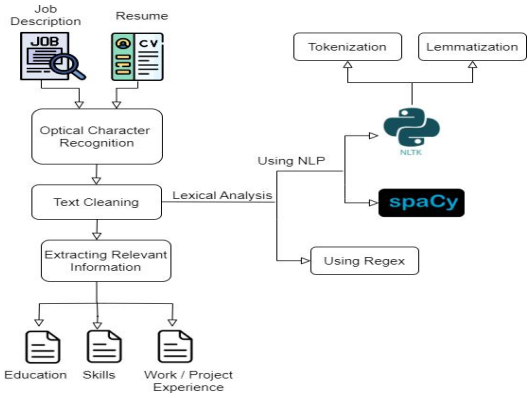

Figure 1. Resume Analysis Module

The model's objective is to predict the following question, which is the output sequence, using the candidate's prior responses as the input sequence, and to create queriesthat match the context of the conversation. A dataset thatcomprises potential answers and the related questions thatwere posed in earlier sessions, is used to train the LSTMmodel for dynamic question generation. This dataset servesas the model's learning environment as it discovers the connections between answers and queries.

Furthermore, the quality of question creation is improvedby using pre-trained models like BERT (Bidirectional EncoderRepresentations from Transformers), which are used to obtain rich contextual embeddings. The LSTM model uses the candidate's most recent response as input during the interviewto predict the following question in real time. The model's comprehension of the situation enables it to come up with inquiries that are not only pertinent but also consistent withthe conversation being held. In order to ensure that the questions are appropriate for the particular interview, it also considers the candidate's skill set and job needs.

The practice of creating dynamic questions is essential for improving the user experience. Candidates are more likelyto stay interested and believe that the dialogue is tailored to their experiences and talents. As a result, the system is better equipped to glean insightful information from interviews. The LSTM model can be further tuned and the quality of thequestions produced over time by gathering feedback fromusers, including interviewers and candidates can be improved.

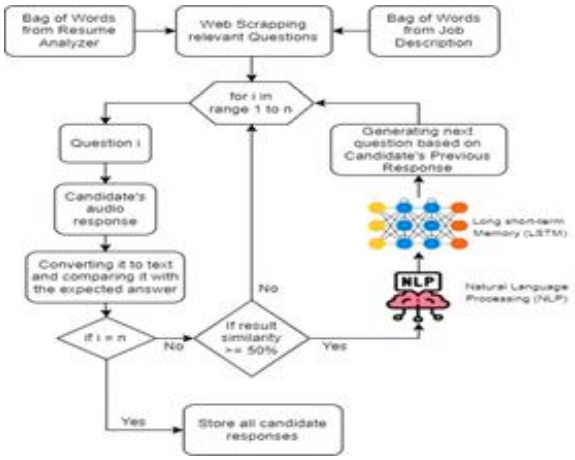

Figure 2. Interview Module with Dynamic Question Generation

D. Response Evaluation and Continuous Feedback

In an interview bot system, analyzing each candidate'sresponse is crucial to determining whether or not they are a good fit for the job. This research paper explores a cutting-edge approach that harnesses the power of the ChatGPT 3.5 API to evaluate student responses. The system first uses ChatGPT 3.5 API to generate an ideal answer and then compares and evaluates the effectiveness of the candidate's response, determining whether it

conforms to the anticipated response or falls short, using text similarity, and semantic analysis. This innovative approach not only assesses students' understanding but also presents them with ideal answers. The system actively helps a candidate improve when it recognizes that their response is inaccurate or not their best. Feedback is crucial to this process because it offers thorough justifications for why the answer was erroneous. It points out any places where the response might have lacked precision or thoroughness and makes helpful recommendations for improvement.

The method also acknowledges the value of practice-based learning. It recommends a collection of practice questions pertaining to the particular subject or idea that the candidate had difficulty with in order to make this easier. These practice questions offer the candidate a learning opportunity to solidify their understanding and efficiently address their areas of weakness.

The system can get better and better over time as applicants use it and learn from their errors. Incorporating machine learning models may improve the comprehension of applicant responses, resulting in more precise feedback and training resources. This method of conducting interviews, especially for technical or knowledge-intensive positions, not only helps candidates advance their knowledge and interview performance but also gives interviewers a thorough understanding of candidates' strengths and areas that require further development, making the candidate evaluation and selection process more effective.

IV. System Architecture

At its core, the ML model forms the backbone of the system. It includes multiple sub-modules: the Resume Analyzer model, which extracts relevant information from the student's resume and job description, enabling a tailored approach; the Speech Recognition and Synthesis model, facilitating the processing of spoken responses; and the Interview Question Generator, which is further divided intothe Base Question Generator and the Follow-up Question Generator. This dynamic generation component is the heartof the project, as it generates interview questions based onthe student's responses to prior questions. It creates a realistic interview experience by adapting the questioning process as the interview progresses.

Additionally, the Response Evaluation model assesses and evaluates the student's answers, offering feedback and suggestions for improvement. The system interfaces with a database (DB) to store and retrieve user data efficiently. The NLP Engine is integrated to enhance natural language understanding and processing. External APIs are leveraged to access externalresources, augmenting and enhancing the system's capabilities.

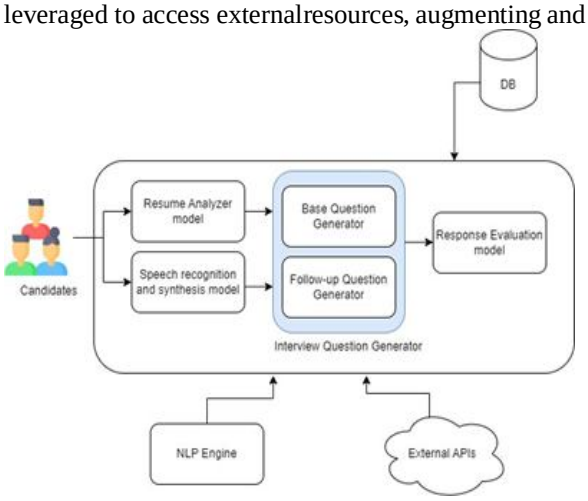

Fig. 3. System Architecture

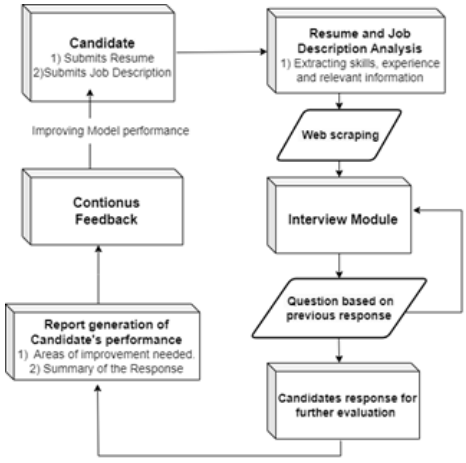

Fig. 4. Data Flow Diagram of the System

V. Conclusion

Automated systems are a noteworthy resource for student interview preparation. They provide students with realisticpractice settings that boost confidence and give them actual interview experiences. The system helps students prepare for such a process, right from resume analysis to technical rounds.The system also generates a report that has all the details ofthe candidate. It also has the details of each question askedin the interview and the expected answer. These systems offer individualized learning opportunities and prompt feedback to

encourage skill and confidence growth. They also save studentstime and money because they are effective and efficient. The impartial evaluation they provide guarantees fair evaluation and aids in students' understanding of their areas of strength and need for development. These systems are crucial in forming the future workforce because they give students the tools, they need to ace interviews and successfully enter the workforce.

VI. Future Work

Emerging technologies like virtual reality (VR) and augmented reality (AR) may be incorporated into interview systems, enabling candidates to participate in immersive, inter- active interviews that replicate real-world job scenarios. More sophisticated assessment techniques will be incorporated into interview methods for assessing technical skills. To evaluate problem-solving skills, these technologies might include code assessment, virtual coding environments, and machine learning models. Systems for conducting interviews will improve accessibility for candidates and employers globally by supporting numerous languages and accommodating various cultural norms.

References

[1] Abdul Wahab, Dr. M N. Nachappa. "Resume Parser with Natural Language Processing." International Research Journal of Engineering and Technology (IRJET) Volume: 09 Issue: 03, March 2022.

[2] Alkeshwar Jivtode, Kisan Jadhav, Dipali Kandhare. "RESUME ANAL- YSIS USING MACHINE LEARNING AND NATURAL LANGUAGE PROCESSING" International Research Journal of Modernization in Engineering Technology and Science, May 2023.

[3] Joko Siswanto, Sinung Suakanto, Made Andriani, Margareta Hardiyanti, Tien Fabrianti Kusumasari "Interview Bot Development with Natural Language Processing and Machine Learning",2022 International Journal of Technology 2022.

[4] J. Purohit, A. Bagwe, R. Mehta, O. Mangaonkar and E. George, "Natural Language Processing based Jaro-The Interviewing Chatbot," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), pp. 134-136, doi: 10.1109/ICCMC.2019.8819708, 2019.

[5] Siddhant Dharmatti, Sumit Jain, Sunny Khandekar, Prem Vora, Prof. Mukesh More "INTERVIEW PRACTICE – VOICE-BASED CHAT- BOT", 2022 International Journal of Advances in Engineering Research.

[6] Shriniket Yakkundi, Amey Vanjare, Vinay Wavhal, Shreya Patankar "Interactive Interview Chatbot", International Research Journal of En- gineering and Technology (IRJET) 2019. Z. Zheng, X. Si, E. Chang, and X. Zhu, 'K2Q: Generating Natural Lan- guage Questions from Keywords with User Refinements', in Proceedings of 5th International Joint Conference on Natural Language Processing, pp. 947–955, 2011.

[7] S. Rakangor and D. Y. R. Ghodasara, 'Literature Review of Automatic Question Generation Systems', 2015.

[8] Pimpalkar, A. Lalwani, R. Chaudhari, M. Inshall, M. Dalwani and T. Saluja, "Job Applications Selection and Identification: Study of Resumes with Natural Language Processing and Machine Learning," 2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, 2023, pp. 1-5, doi: 10.1109/SCEECS57921.2023.10063010.

[9] R. C. Tripathi and Chandramma, "An Approach of Intelligent Auto- mated Resume Analysis and Recommendations," 2022 Fourth Inter- national Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT), Mandya, India, 2022, pp. 1-5, doi: 10.1109/ICERECT56837.2022.10059603.

[10] M. Sharma, G. Choudhary and S. Susan, "Resume Classification using Elite Bag-of-Words Approach," 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2023, pp. 1409-1413, doi: 10.1109/ICSSIT55814.2023.10061036.

[11] P. Swami and V. Pratap, "Resume Classifier and Summarizer," 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), Faridabad, India, 2022, pp. 220- 224, doi: 10.1109/COM-IT-CON54601.2022.9850527.

[12] S. Bharadwaj, R. Varun, P. S. Aditya, M. Nikhil and G. C. Babu, "Resume Screening using NLP and LSTM," 2022 International Conference on Inventive Computation Technologies (ICICT), Nepal, 2022, pp. 238- 241, doi: 10.1109/ICICT54344.2022.9850889.

[13] R. Shaikh, N. Phulkar, H. Bhute, S. K. Shaikh and P. Bhapkar, "An Intelligent framework for E-Recruitment System Based on Text Categorization and Semantic Analysis," 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2021, pp. 1076-1080, doi: 10.1109/ICIRCA51532.2021.9544102.

[14] S. Kul and A. Sayar, "Entity Name Recognition in Job Postings and Resumes," 2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Ankara, Turkey, 2021, pp. 1-6, doi: 10.1109/HORA52670.2021.9461326.

[15] X. Li, H. Shu, Y. Zhai and Z. Lin, "A Method for Resume Information Extraction Using BERT-BiLSTM-CRF," 2021 IEEE 21st International Conference on Communication Technology (ICCT), Tianjin, China, 2021, pp. 1437-1442, doi: 10.1109/ICCT52962.2021.9657937.

[16] Q. Xu, J. Zhang, Y. Zhu, B. Li, D. Guan and X. Wang, “A Block-Level RNN Model for Resume Block Classification,” 2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA, 2020, pp. 5855- 5857, doi: 10.1109/BigData50022.2020.9377771.
[17] M. V. Belova and A. B. Zhernakov, “Modern Methods of Resume Processing in Recruiting Information Systems,” 2018 XVII Russian Scientific and Practical Conference on Planning and Teaching Engineering Staff for the Industrial and Economic Complex of the Region (PTES), St. Petersburg, Russia, 2018, pp. 19-22, doi: 10.1109/PTES.2018.8604263.
[18] R. Pandey, D. Chaudhari, S. Bhawani, O. Pawar and S. Barve, “Interview Bot with Automatic Question Generation and Answer Evaluation,” 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2023, pp. 1279-1286, doi: 10.1109/ICACCS57279.2023.10112918.
[19] D. V. Paul and J. D. Pawar, “Dynamic Question Paper Template Generation Using Bi-proportional Scaling Method,” 2013 IEEE Fifth International Conference on Technology for Education (t4e 2013), Kharagpur, India, 2013, pp. 80-83, doi: 10.1109/T4E.2013.28.
[20] D. V. Paul, S. B. Naik, P. Rane and J. D. Pawar, “Use of an Evolutionary Approach for Question Paper Template Generation,” 2012 IEEE Fourth International Conference on Technology for Education, Hyderabad, India, 2012, pp. 144-148, doi: 10.1109/T4E.2012.24.
[21] G. Deepak, N. Kumar, G. V. S. Y. Bharadwaj and A. Santhanavijayan, “OntoQuest: An Ontological Strategy for Automatic Question Generation for e-assessment using Static and Dynamic Knowledge,” 2019 Fifteenth International Conference on Information Processing (ICINPRO), Bengaluru, India, 2019, pp. 1-6, doi: 10.1109/ICInPro47689.2019.9092034.
[22] S. Jena, A. Raina, H. Raina and C. Nehete, “AI based Dynamic Model for Question Paper Generation and Performance Monitoring,” 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2022, pp. 16-23, doi: 10.1109/ICAAIC53929.2022.9792657.
[23] D. S. Rozali, M. F. Hassan and N. Zamin, “A survey on adaptive qualitative assessment and dynamic questions generation approaches,” 2010 International Symposium on Information Technology, Kuala Lumpur, Malaysia, 2010, pp. 1479-1484, doi: 10.1109/ITSIM.2010.5561484.
[24] J. McGough, J. Mortensen, J. Johnson and S. Fadali, “A web-based testing system with dynamic question generation,” 31st Annual Frontiers in Education Conference. Impact on Engineering and Science Education. Conference Proceedings (Cat. No.01CH37193), Reno, NV, USA, 2001, pp. S3C-23, doi: 10.1109/FIE.2001.964056.
[25] H. Xue, Z. Zhao and D. Cai, “Unifying the Video and Question Attentions for Open-Ended Video Question Answering,” in IEEE Transactions on Image Processing, vol. 26, no. 12, pp. 5656-5666, Dec. 2017, doi: 10.1109/TIP.2017.2746267.
[26] Shu-Chen Cheng, Yen-Ting Lin, Yueh-Min Huang, Dynamic question generation system for web-based testing using particle swarm optimization, Expert Systems with Applications, Volume 36, Issue 1, 2009, Pages 616-624, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2007.09.064.