



INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

PLACEMENT PREP RAG

Arya Ade, Prof Nilima Kulkarni, Sahil Vankudre, Sayan Paul, Ayush Kokil

Student, Professor, Student, Student, Student

School of Computing

MIT ADT University, PUNE

Abstract: PlacementPrepRAG is a tool that helps students to prepare for their placements and job interviews in a very seamless manner. In this an AI method is used namely Retrieval-Augmented Generation (RAG). It is like a single platform for multiple solutions - tips for interviews, aptitude tests, test questions. Whatever students would ask it will provide short simple answers. Some more features like quizzes which would be based on the student's performance and mock interviews that help boost student's confidence. So instead of spending more time on searching for study materials, students can focus more on gaining knowledge. Since AI is used, PlacementPrepRAG gives more personalized experience which helps students feel free and confident for job interviews.

Index Terms - Component, formatting, style, styling, insert.

I. INTRODUCTION

Today it has become very tough to land a job. It requires a lot of preparation and a strong understanding of the topics. Students have to struggle because their preparation materials are scattered and not present in a single place due to which a lot of time gets wasted and students get confused. PlacementPrepRAG serves as a solution to this problem by using Retrieval-Augmented Generation (RAG). It works as a single place for all placement preparation materials. Everything like- coding questions, interview tips and aptitude tests is served from a single place. Since it aims at providing personalized experience, students can get their personalized answers through chatting, reading summaries and practicing quizzes specially designed for them. This platform will help students to focus more on their weak areas and prepare accordingly. Hence it will save their crucial time at the time of preparation.

II. EASE OF USE

PlacementPrepRAG is a student helper which serves students by helping them study efficiently for their job placements. It is built in such a way that anyone, whether experienced or has just started can use it seamlessly. Students can study from their own notes in a very productive way by uploading it and asking only those questions which they don't know or want to recap. PlacementPrepRAG allows chatting through which students can ask questions in very simple everyday language. It replies quickly and correctly by analysing data from many trusted sources.

III. LITERATURE REVIEW

A. Overview of Placement Preparation Systems

Traditional methods of preparing for placements involve studying PDFs, documents or files. Since these materials are not in one place, students waste more time on searching rather than studying due to which students end up getting confused. Colleges and websites do provide structured courses but most of these courses take longer duration than required. PlacementPrepRAG not only saves student's time but also provides a personalized experience. Today there are a lot of tools that provide structured material but they are not user specific. PlacementPrepRAG is user specific and helps students progress personally. Therefore

PlacementPrepRAG is a tool that caters to student's problem of searching study materials during placement preparation in a very efficient way.

B. Information Retrieval in Educational Systems

Information Retrieval is a method that helps students get student study materials from different sources like websites, e-books and from many other online sources. Modern Information Retrieval tools use concepts like searching based on meaning, understanding natural language and indexing smartly to help users get what they need in the shortest time possible. Good IR tools are very much required for students as it allows them to get the exact study material they want at the earliest and also saves a lot of their crucial time. Still there are certain challenges like handling of many file types at once, arrangement of scattered data and verifying whether the collected data is actually correct and useful to serve the purpose.

C. Intelligent Tutoring System Frameworks

Intelligent Tutoring Systems (ITS) which are also called smart learning systems, use AI to teach students in such a way that best suits them. Like a private tutor these students best understand a student's progress and learning style. These systems analyze how fast a student learns, how are questions answered by a student and how a student approaches a solution. All these help the system understand about the student like what the student knows and what he doesn't. Then accordingly it provides the learning material.

D. Retrieval-Augmented Generation (RAG) Framework

This framework mostly referred to as RAG framework or pipeline works as a retrieval pipelines in the ai models which helps in retrieving the data from the vector databases. RAG framework works by retrieving the most relevant information of the domain needed. RAG models does this by working in a specific method. First the RAG model converts the datasets into the vector format and stores these vector values into the vector databases. While storing these values the most relatable vector values are stored close to each other in the database. Thus while retrieving data of a specific domain all the related data will be retrieved in a very short time with higher accuracy in the retrieved data. Thus in this way the RAG model does the work efficiently and in less time with more accuracy.

E. Use of Large Language Models (LLMs) in Education

Large Language Models also referred as LLMs are the brain of any AI model we work with. All the processing of input prompt from the user and the natural language processing of the out is done by the LLM. These LLM model works on the concept of tokenization of the dataset and based on the token while generating the out the next token will be predicted and the output will be generated in a natural language format. These LLMs play an important role into the education system as the generation of accurate answers for the questions is very important. The retrieved models retrieves the most suitable and related data but to generate an accurate answers without any helucination is very necessary. Thus LLMs play an important role in models output accuracy. As well the LLM model is trained so enhanced will be the accuracy of its output.

F. Knowledge Base Construction and Embeddings

The knowledge based construction and embedding of the data is necessary and crucial. As based on these embeddings the data will be stored and retrieved from the database. More efficient the data embedding done, more efficient will be the model output and the accuracy of the output as per the questions asked to the model. This will also ensure the relevance of the answers given by the model for the questions asked to it.

G. AI-based Personalized Learning and Recommendation Systems

In today's world of education the AI has been integrated as a part of education system and to make the training and learning preparation better the AI based personalized learning and recommendation system are necessary and should be developed with more precision. So which it would highly benefit students in their future tenure of learning. Personalized learning can make the ai education system to concentrate on each and every individual in a personalized way with working on individual student's strength and weaknesses. For preparation of exams and other learning activities the specific topics for the students will be created and according to the students state of preparation the further planning of the learning plan will be executed. Thus it will greatly benefit to each and every student of a personalized AI based educational model will be provided.

H. Conversational AI and Interactive Learning Platforms

Conversational AI and interactive learning platforms refers to the models which are specially trained of educational usage. These models will be trained in such a way that any domain specific questions the students have will be solved with proper reasoning as well as the model will be interactive for any doubts the students have. These models could also have knowledge of other further related topics through which all the necessary links regarding the topics will be provided accurately. This also gives students and immersive experience of 1 v 1 learning session through which students can have better understanding and clearance of the topics without any remaining doubts.

I. Evaluation Metrics and Effectiveness of AI-Assisted Preparation Tools

The Evaluation metrics of this AI assisted preparation tool will be done by evaluating this model with company specific questions as output. All the in depth concept of system design and software. This the model should be able to retrieve all the domain specific data as well as the reasoning and the answers of the asked questions should be proper and without any fault. The Effectiveness of this can be tested by deploying the prototype for students where they can do the preparation and as well as learning from the same model. Thus based on the test results of the students the efficiency and effectiveness of the model can be tested and verified.

IV. METHODOLOGY

The PlacementPrepRAG system makes the use of RetrievalAugmented Generation (RAG) to make it easy for students to access the scattered resources, evaluate their content , and offer customized company wise placement material. By utilizing intelligent Retrieval-Augmented Generation (RAG), summarization and adaptive generation with the help of LLM .The model converts scattered educational materials into a combined repository which can be used as a mentor..

4.1 Data Collection and Aggregation

The first phase of creating PlacementPrepRAG consists of data gathering and integration from various resources such as Telegram groups, college drives and student notes. Placementrelated resources are scattered across different digital platforms and formats such as PDFs, presentation slides, cloud storage, and telegram groups.

To develop a effective platform which solves the student problems , these scattered resources must be collected and organized for analysis and retrieval. This process starts with finding all the trustworthy data sources that contain important placement topics such as quantitative aptitude, logical reasoning, programming, communication skills, and core subjects such as OS,DBMS,JAVA etc.

We can make use of Automated web crawlers and APIs for large-scale extraction from public repositories. Each document that is collected from the online resources is validated to eliminate duplicates , irrelevant information, and outdated content.

4.2 Preprocessing of PDFs and Documents

The first step in building PlacementPrepRAG is to collect and combine data from different sources. Information related to placements is usually spread out across many formats like PDFs, PowerPoint slides, online forums, cloud drives, and chat apps. To make a useful and efficient platform, all these scattered resources need to be gathered and organized properly so that the system can easily access them whenever needed.

The process begins with text extraction where we made the use of tools such as PyPDF2 or PDFMiner to convert the content of PDFs into plain text. After that data cleaning is performed to eliminate unnecessary symbols and words, extra spaces. After that we perform tokenization and stop-word removal to help break down the sentence and get the semantic meaning. Using lemmatization procedure the collected words are converted into their base forms which improves semantics. Hence using NLP which stands for Natural Language Processing the text extraction is carried out. This extracted data is still the raw data which needs to be processed in order to store this inside a vector database, hence metadata such as file name, topic category is stored for better structing. The raw data is then cleaned and preprocessed which helps the

PlacementPrepRAG to generate accurate, context rich and meaningful answers as per user searched requirements.

4.3 Data Storage and Indexing

We have used RAG i.e. the retrieval augmented generation pipeline for the retrieving the stored data from the vector database. The retrieval layer processes the all the data and pdf documents stored in the main vector database where all the similar or related keywords are stored closely as vector embeddings. With proper encoding the document data is converted into vector embeddings using transformer models, helping the system context aware and understand meaning of the data.

4.4 Embedding Generation for Semantic Search

Generating Embeddings is crucial for the search and output feature of PlacementPrepRAG system. This is the backbone of the system where the targeted search is done based on the keyword search. This method does not follow the regular search based on just the keywords, within the current workflow our system search the keywords and the document data which is converted into embeddings are stored near to closely related data. This helps our search feature extract all the matches based on what user is exactly asking for and needs to access.

Explaining the actual workflow of the system, first all the text data is collected from the input source provided to the RAG pipeline. This text data is then cleaned and structured by removing all the unwanted data and duplication. This cleaned data is then converted into numerical codes through embedding models like BERT. These created vectors are then stored into vector databases like Pinecone or pgVector which is designed to find the similar numbers at low latency. All the user asked questions are also converted into embedding tokens and matched with the database data and closest match is given as output.

Feature Engineering plays a crucial role in our Placement Prep-RAG Software. It transforms unstructured and unordered data into a structured format that helps students prepare effectively for placements.

The process begins with gathering placement-related data from various sources such as PDFs, Telegram channels, and company websites. Once the data is collected, it is cleaned and transformed to ensure accuracy and consistency. After processing, the system enables faster and more efficient preparation. The refined data is then converted into relevant questions and answers, which assist in technical interviews, assessments, HR rounds, and coding assignments. These are further tailored based on different company requirements and job roles.

The text data is then converted into vectors (embeddings) using embedding models, which allows the system to understand the meaning of words and context, rather than just matching exact keywords. This helps PlacementPrepRAG retrieve the most relevant results from the query provided by the user.

4.5 Question-Answering Module

The Question-Answer section of the PlacementPrepRAG system helps students fetch accurate, and useful answers to their placement-related problems. It works using a method called RAG which stands for Retrieval-Augmented Generation, which combines a system with an LLM which we use for AI answer generator. This ensures that every answer is correct and easily understandable.

When a student asks a question, the system first searches through its collection of notes, PDFs, and web links in the vector database to find the most relevant information. It then sends this information to the LLM, which combines it into a simple and structured answer according to the users query.

This project can handle multiple types of questions, from simple to complex. It can manage aptitude, coding, finance, and management-related questions. We have also added a feedback option so we can improve the system. This feedback helps in increasing accuracy.

Placement Prep RAG Assistant (Gemini API)

Accepted format: Plain text (.txt) or Markdown (.md)

Upload Custom Placement Notes

Drag and drop file here
Limit: 200MB per file - TXT, MD

salesforce_repo.txt 2.9KB

Added 10 new chunks from salesforce_repo.txt!

Please reload chat to use the new data.

Placement Prep RAG Assistant (Gemini API)

HELLO! I am the Placement Prep RAG Assistant. Ask me about DSA, OOPS, CN, OS, or DBMS concepts from the interview dataset.

sorting

I found 5 relevant questions:

- Topic: DSA - Topological Sort**
Question: What is Topological Sort, and state the necessary condition for a graph to allow a topological ordering. Give a real-world application. Company: GlobalTech Solutions Year: 2025
- Topic: DSA - Merge Sort**
Question: Explain the working principle of Merge Sort and derive its time complexity. Company: TechMinds Consulting Year: 2024
- Topic: DSA - Quick Sort Worst Case**

4.6 System Evaluation and Testing

PlacementPrepRAG is tested multiple times to ensure it runs error-free and accurately with zero bugs. It is made easy for students and professionals to use and study from. We used both manual and automated testing to ensure proper working.

It is also tested on multiple networks and locations, under different conditions, with various types of data, load, and capacity. Performance is tested multiple times to ensure good speed. After deployment, it is tested in real-time and also through users, including students and professionals. We are continuously improving the system and adding new features on a daily basis.

V. ACKNOWLEDGMENT

We thank our guide, professor, and university for their support, help, and guidance during this project. Their important suggestions and insights helped our team in developing our project PlacementPrepRAG. They also guided us in adding innovative features. We are always thankful to them.

REFERENCES

- SHAO, Z., GONG, Y., SHEN, Y., HUANG, M., DUAN, N., & CHEN, W. (2023). *Enhancing Retrieval-Augmented Large Language Models Through Iterative Retrieval-Generation Synergy*. arXiv Preprint.
- CUCONASU, F., TRAPPOLINI, G., SICILIANO, F., FILICE, S., CAMPAGNANO, C., MAAREK, Y., TONELLOTO, N., & SILVESTRI, F. (2024). *The Influence of Noise on Retrieval in RAG Systems*. arXiv Preprint.
- CHENG, X., LUO, D., CHEN, X., LIU, L., ZHAO, D., & YAN, R. (2023). *Retrieval-Augmented Text Generation with Self-Memory Mechanisms*. arXiv Preprint.
- RAM, O., LEVINE, Y., DALMEDIGOS, I., MUHLGAY, D., SHASHUA, A., LEYTON-BROWN, K., & SHOHAM, Y. (2023). *In-Context Retrieval-Augmented Language Models: Techniques and Insights*. arXiv Preprint.
- RASHTCHIAN, C., & JUAN, D.-C. (2025). *Exploring Contextual Depth in Retrieval-Augmented Generation*. Google Research Blog.
- SANKARAN, S. (2024). *Retrieval-Augmented Generation (RAG): A Concise and Comprehensive Guide*. Medium.
- GUPTA, S., RANJAN, R., & SINGH, S. N. (2024). *Retrieval-Augmented Generation (RAG): Evolution, Trends, and Future Directions — A Survey*. arXiv Preprint.