



DATA STRUCTURES SUBJECTIVE ANSWER EVALUATION SYSTEM

PRASHU VIMAL
244I9MCA005



Problem Statement & Solution Proposed

Subjective paper evaluation is a tricky and tiresome task to do by manual labor. Insufficient understanding and acceptance of data are crucial challenges while analyzing subjective papers using Artificial Intelligence (AI). Several attempts have been made to score students' answers using computer science. However, most of the work uses traditional counts or specific words to achieve this task. Furthermore, there is a lack of curated data sets as well.

The solution proposed will utilize various machine learning, natural language processing techniques, and tools such as **Keyword/TF-IDF, Semantic Embeddings, Concept Mapping, Concept Weighting, Transformer Fine-tuning** to evaluate descriptive answers. We use different evaluation measures such as **TF-IDF Cosine similarity+Keyword Coverage, BERTScore + Concept Overlap, Graph Similarity +, MAE+RMSE etc.** to evaluate the performance of various models performance.

Dataset Design

01

Context.json:

- Contains chapters, paragraphs, or topic-level content.
- Acts as the primary knowledge source for evaluating answers.
- Provides concepts, keywords, and explanations that are used for semantic comparison.

Structure:

```
{  
  "context_id": "",  
  "topic": "",  
  "text": ""  
}
```

02

Question.json

- Stores multiple ways each question can be framed.
- Linked to context_id to associate questions with relevant content.

Structure

```
{  
  "context_id": "",  
  "questions": [  
    {"id": "1", "question": "question_pattern_1", "max_marks": {}},  
    {"id": "1_1", "question": "question_pattern_2", "max_marks": {}},  
    {"id": "2", "question": "question_pattern_3", "max_marks": {}}, ...  
  ]  
}
```

03

Answer.json:

- Provides three reference answers per question for supervised training:
- Correct Answer: Taken directly from the source book or material.
- Partially Correct Answer: Common partial understanding from students or paraphrased version
- Incorrect Answer: Represents wrong or irrelevant responses.

Structure:

```
{
  "answers": [
    {
      "question_id": "1",
      "correct": "Correct answer text", "score": " ",
      "partial": "Partially correct answer text", "score": " ",
      "wrong": "Incorrect answer text", "score": " "
    },
    {
      "question_id": "1_1",
      "correct": "Correct answer text for variant", "score": " ",
      "partial": "Partially correct text for variant", "score": " ",
      "wrong": "Incorrect text for variant", "score": " "
    }
  ]
  ...
}
```

User Interface

Key components

1.) Chapter Selection Dropdown/List

Example: Recursion, Sorting, Graphs

2.) Question Textbox

3.) Student Answer Input

- Multi-line Textarea: “Enter Student Answer”

4.) Evaluate Button

- Sends the input to the backend via API call

5.) Output Panel

- Displays:
 - Score
 - Label (Correct / Partially Correct / Wrong)
 - Feedback (Missing points and suggestions)

Final marks = Model score * Max marks

○ ○ ○ ○ ○
TECH-STACK

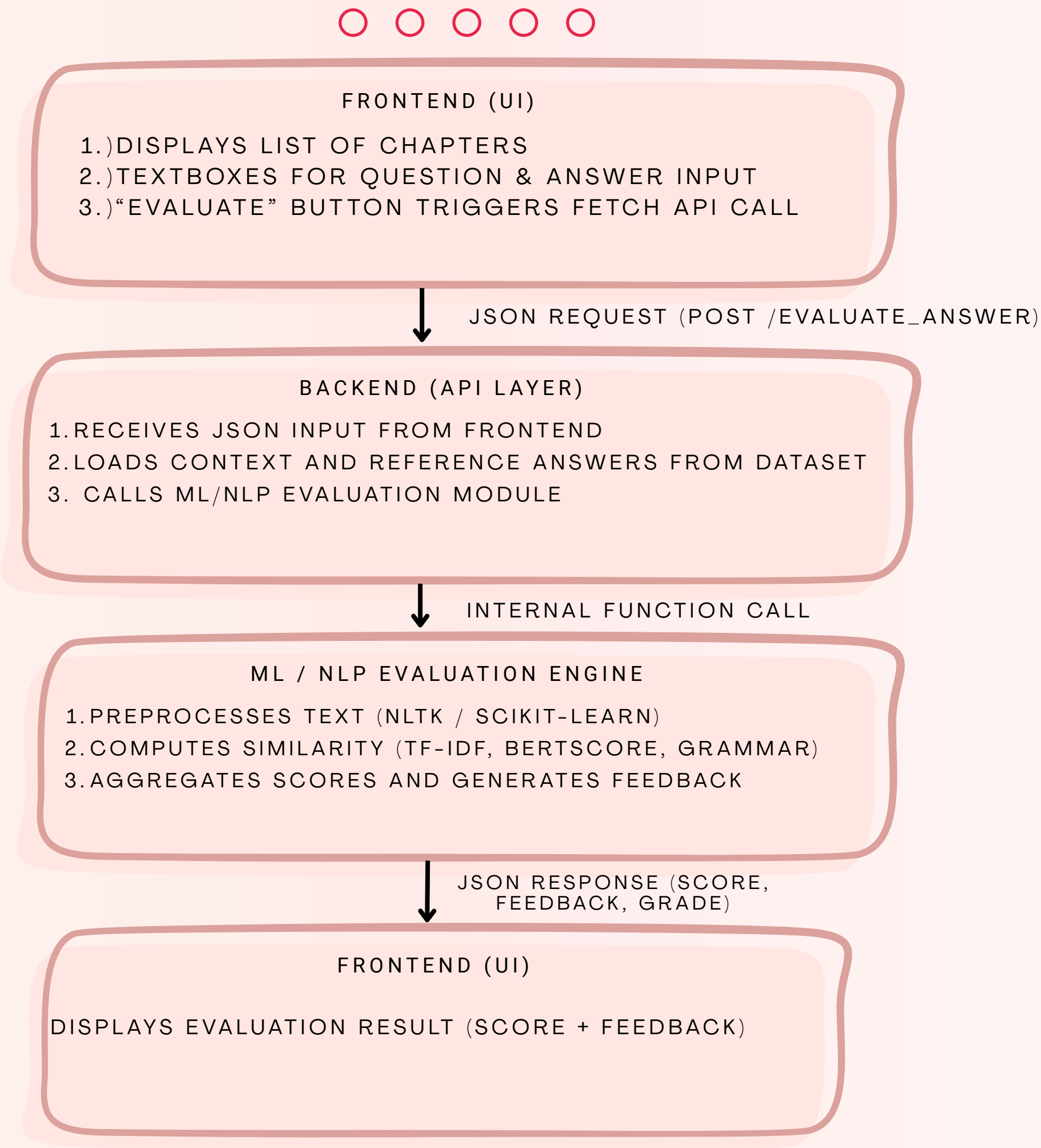
Frontend

- **UI Framework + Styling:**React.js + Tailwind CSS
- **API Calls:**Fetch API
- **Build Tool:**Vite

Backend

- **Programming Language:** Python
- **Web Framework:**FastAPI
- **ML/NLP Libraries:**Transformers / Sentence-BERT / Custom model
- **Text Processing:**NLTK/ scikit-learn
- **Data Storage:**JSON

WORK
FLOW



DATA COLLECTION & PREPROCESSING

- EXTRACT QUESTIONS AND REFERENCE ANSWERS FROM BOOKS, ONLINE DSA NOTES, AND TUTORIALS.
- GENERATE PARTIALLY CORRECT AND WRONG ANSWERS INDUCING MANUAL ERRORS IN THE ANSWERS.

PREPROCESSING:

- SENTENCE SEGMENTATION
- STOPWORD REMOVAL
- TOKENIZATION
- LEMMATIZATION
- POS TAGGING

FEATURE EXTRACTION:

- **TF-IDF**: MEASURES WORD IMPORTANCE IN STUDENT VS. REFERENCE ANSWERS
- **KEYWORD EXTRACTION**: IDENTIFIES DOMAIN KEYWORDS USING POS TAGGING
- **SEMANTIC EMBEDDINGS**: CAPTURES CONTEXTUAL SIMILARITY USING SENTENCE-BERT
- **GRAPH-BASED CONCEPT MAPPING**: MAPS CONCEPTUAL RELATIONSHIPS
- **CONCEPT WEIGHTING**: ASSIGNS WEIGHTS TO CRITICAL CONCEPTS

EVALUATION METRICS

- **COSINE SIMILARITY:** LEXICAL AND CONTEXTUAL SIMILARITY
- **BERT-SCORE:** SEMANTIC SIMILARITY
- **MAE / RMSE :** ERROR IN PREDICTED VS ACTUAL MARKS
- **BLEU / ROUGE-L:** OVERLAP BETWEEN REFERENCE AND STUDENT TEXT
- **PRECISION / RECALL / F1:** ACCURACY OF CLASSIFICATION OF ANSWER QUALITY

FUTURE SCOPE

- 1.DATASET EXPANSION
- 2.DATABASE & PERSISTENT STORAGE
- 3.MULTI-SUBJECT SCALABILITY
- 4.STUDENT PERFORMANCE INSIGHTS

CONCLUSION

THE SYSTEM WILL TRY TO COMBINE LINGUISTIC, SEMANTIC, AND CONCEPTUAL INTELLIGENCE TO MAKE SUBJECTIVE ANSWER EVALUATION RELIABLE, SCALABLE, AND MORE HUMAN-LIKE.

REFERENCES

BOOK REFERRED

- DATA STRUCTURES AND ALGORITHMIC PUZZLES

RESEARCH PAPER

- [HTTPS://IEEEEXPLORE.IEEE.ORG/DOCUMENT/9627669](https://ieeexplore.ieee.org/document/9627669)



Thank You!

