SCSE20113 - Machine Reading Comprehension for Question Answering

Presented by Hussain Khozema Kheriwala

Supervised by Dr Sunil Sivadas, Co-supervisor Dr Yan Shi Xing

Problem Statement

Build a question answering system which can help student to answer questions regarding GEM explorer program.

This will allow students to quickly query and efficiently get information without needing to read through various websites.

The information includes application deadlines, eligibility, cost, and other program specific details

Pipeline

1) Query Processing

Classifies questions according to the context. This module identifies the context and focus, classifies the type of question, and sets the answer type's expectations.

2) Document Retrieval

Information retrieval module that focuses on gathering relevant documents.

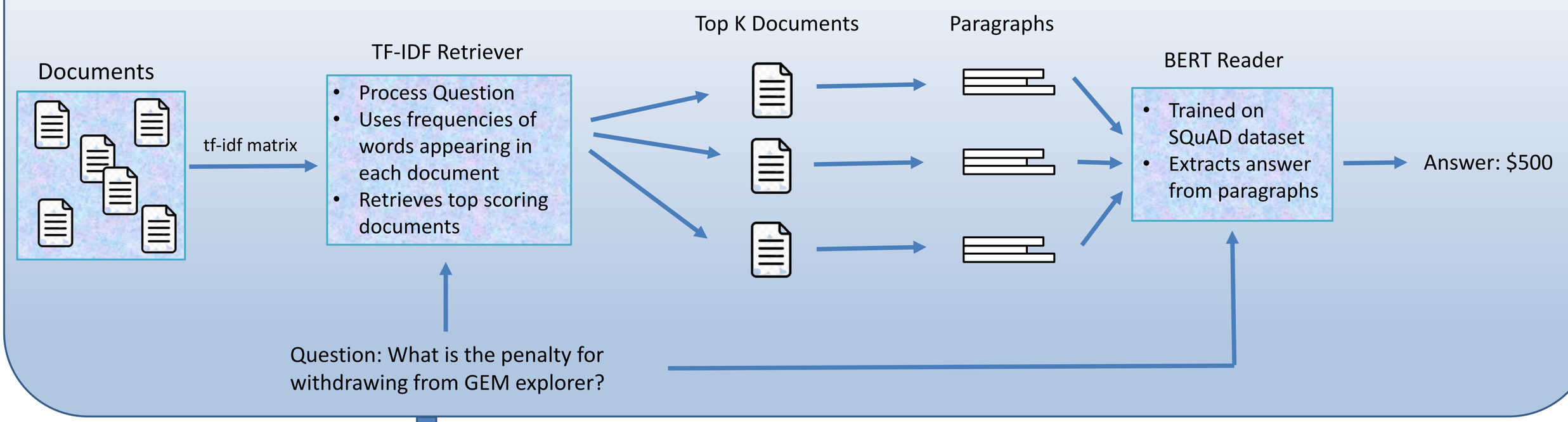
3) Document Processing

Filter the relevant paragraphs from the document

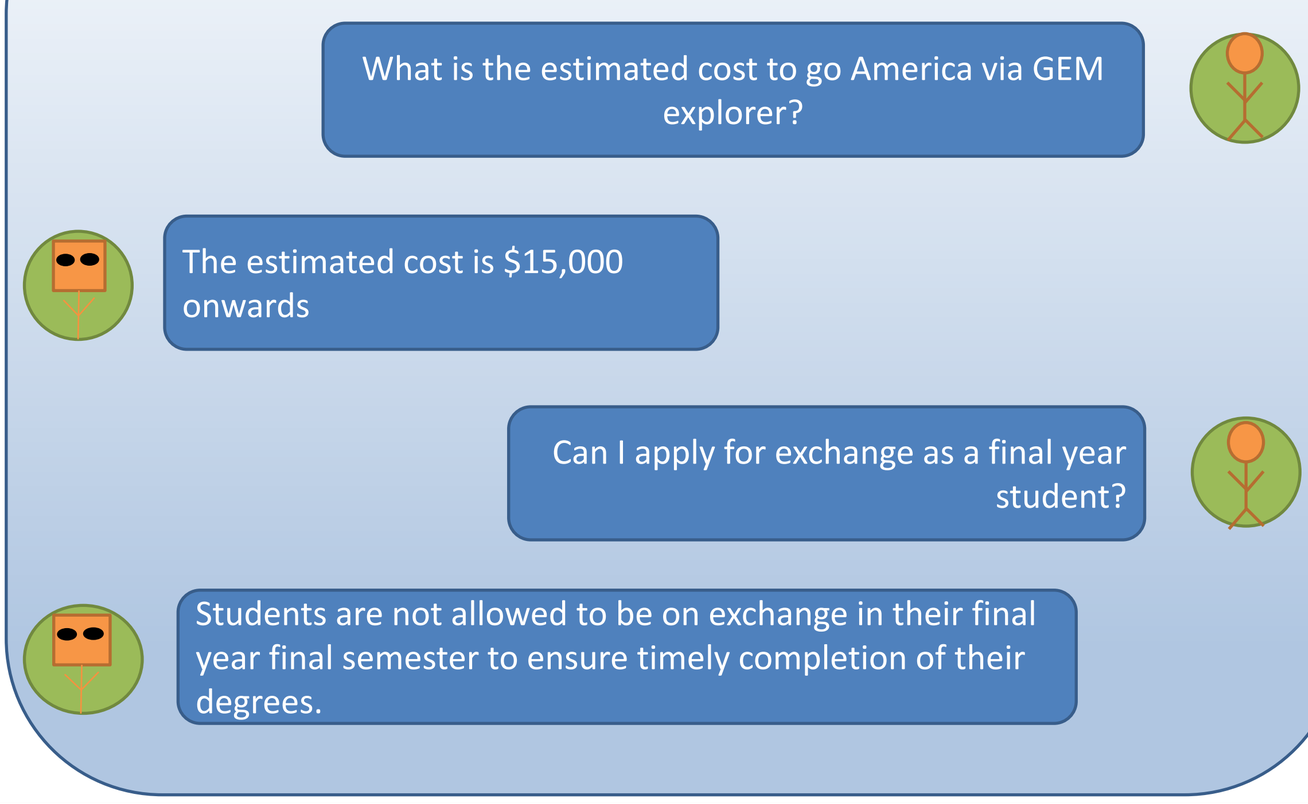
4) Answer Extraction

Parse through to obtain an accurate and appropriate answer.

Approach



Demo Questions



Acknowledgements

- This project is part of Undergraduate Research Experience on Campus (URECA) and was supported by Singtel Cognitive and Artificial Intelligence Lab for Enterprises at NTU
- Thank you to Dr Sunil Sivadas, Dr Yan Shi Xing Josey Matthew and Yi Ren for being part of this journey in furthering my research experience through feedback and guidance.

References

- Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, Percy Liang. (2016). SQuAD: 100,000+ Questions for Machine Comprehension of Text. https://arxiv.org/abs/1606.05250
- Pranav Rajpurkar (2016). The Stanford Question Answering Dataset (SQuAD) . https://rajpurkar.github.io/SQuAD-explorer/
- Nanyang Technological University GEM explorer (https://global.ntu.edu.sg/GMP/gemexplorer/Pages/index.aspx)
- CS224n :Natural Language Processing with Deep Learning (https://web.stanford.edu/class/cs224n/)
- How to create your own Question-Answering system easily with python (https://towardsdatascience.com/how-to-create-your-own-question-answering-system-easily-with-python-2ef8abc8eb5)
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, Veselin Stoyanov. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. https://paperswithcode.com/paper/roberta-a-robustly-optimized-bert-pretraining