AUTOMATIC MCQ GENERATOR & CONTENT SUMMIZER USING NATURAL LANGUAGE PROCESSING

By GROUP 15

Outline

- What is Automatic MCQs Generator?
- Walkthrough
- Usage
- Django

Keywords

- Wordnet
- NLTK
- Concceptnet
- Vector word
- PKE
- BERT Extractive Summarizer
- Hypernyms
- Hyponyms

What is Automatic MCQs Generator?

Multiple-choice questions (MCQs) generation using Natural Language Processing (NLP) is a system that set multi-choice questions from a document in a text file and thereafter provide answers to the questions generated. That is program will take an input (example a text file) then it will generate an output which will be a set of automatically generated multiple choice questions with efficient distractors (wrong answer choices).

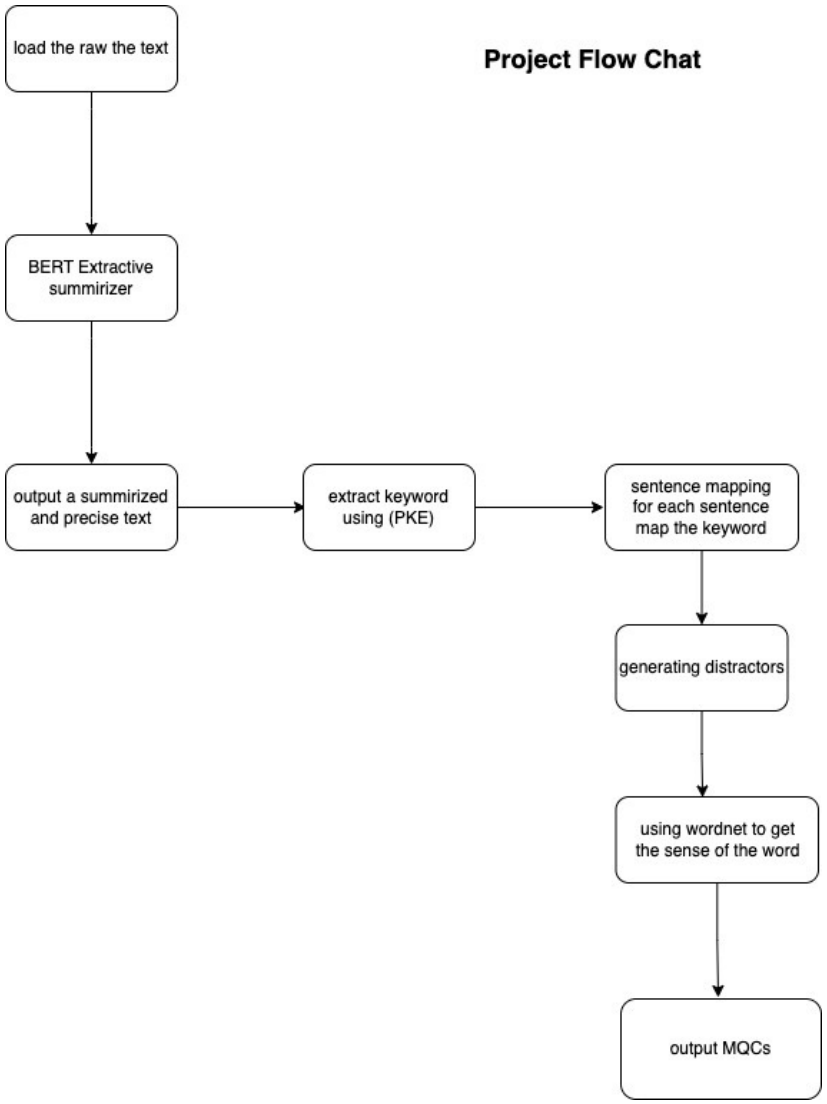
Approach

This is how we plan to achieve our goal.

- **First, we install NLTK**

NLTK is a python library used for natural language processing tasks.

Project Flowchart



Load the raw text

The first step is to load raw text i.e. input text of any domain for which the questions to be generated

Summarizer

Each sentence is not capable of generating questions. Only the sentences that contain a questionable fact can act as a candidate for creating MCQs. Therefore, sentence selection plays a crucial role in the automatic MCQ generation task. Hence for summarizing the text, BERT Algorithm is used.

Extract Keywords

After summarizing the text, keywords are selected from the sentence. This keyword will be the answer to the question. Since all the words in an informative sentence cannot serve as the key proper keyword selection is required. The extraction of keyword is done by using Python Keyword Extractor (PKE) library.

Sentence Mapping

For each of the keyword we will extract corresponding sentences that has the word from the text.

Distractors

This is the most crucial step in the generation of automated MCQs. The difficulty of MCQs highly relies on the quality of distractors produced. A good distractor is one that is very similar to the key but not the key itself. So, for generating distractors Wordnet approach is used. We get the distractors with the help of hypernyms and hyponyms of the key.

Output

After processing all the steps Multiple Choice Questions (MCQs) are generated successfully. Efficient questions are produced with good quality distractors by wordnet.

Usage

The problem of manually creating questions is solved with the proposed system. The proposed system creates automated questions with the help of NLP that reduces human intervention, and it is a cost and time effective system. And the accuracy of the distractor generated is reasonably high. This system not only helps teachers with E-assessments but also helps students who are preparing for competitive exams. Students can test their ability to solve the questions and can also check their understanding of the concepts.

Consequently, it can be used in every area of education sector to test the knowledge or level of the students.

Django

After we done with everything locally and it is working perfectly as expected we will then use Django framework to make it a web base application.

By using Xhtml2pdf, generated MCQs can be exported and download as pdf file.

Constraints

- Getting distractors for phrases and proper nouns