

# Creating a Multiple-Choice Question (MCQ) Generator: Methods and Techniques Report

Process:

Text birth pdfplumber -----> Text Tokenization-----> Keyword birth----->  
Named reality -----> Recognition( NER)

Steps:

## Step 1 Text birth

The first step in our MCQ generation process is to prize the textbook from educational content, generally stored in PDF format. We use the pdfplumber library to prize textbook from PDF lines. This step ensures that we've the content to work with.

## Step 2 Text Tokenization

To work with the textbook effectively, we tokenize it using the BERT tokenizer. Tokenization breaks the textbook into lower units, generally words or subwords, which is essential for posterior NLP tasks.

## Step 3 Keyword birth

Keyword birth is pivotal to identify essential generalities and terms in the textbook.

## Step 4 Named reality Recognition( NER)

Feting realities like names, dates, and locales in the textbook is pivotal for generating contextually applicable MCQs. We employ the Babelscape's NER model to identify realities within the textbook. This step helps us understand the environment more.

## Step 5 MCQ Generation

The core of our MCQ creator is the MCQ generation step. We use the T5 model, a textbook- to- textbook motor model, for this purpose. For each honored reality, we formulate a question advisement, similar as "What's the part of( reality)((entity\_type)) in the environment of the textbook?" Using this question, we induce MCQs.

For each reality, we produce two correct answer choices, icing a balance between the options. We also induce two incorrect answer choices, or distractors. These distractors are named from the textbook and should be contextually presumptive but incorrect.

## Results

Our automated MCQ creator successfully produces MCQs grounded on the honored realities within the educational textbook. The generated questions are contextually applicable and include both correct and incorrect answer choices. This tool significantly reduces the time and trouble needed to produce MCQs for educational assessments.

IF Time would have permitted:

I would have tried with different model of hugging face ner as it was required to be subject inclusive. The idea linking to topics and creating mcqs wich are dual correct would required more powerful ner model and better text sementics analytics and more use of gpt-4 if possible and better topic expert would have given more insights of type of question generated checking .