UE19CS334 - NLP Project Approval

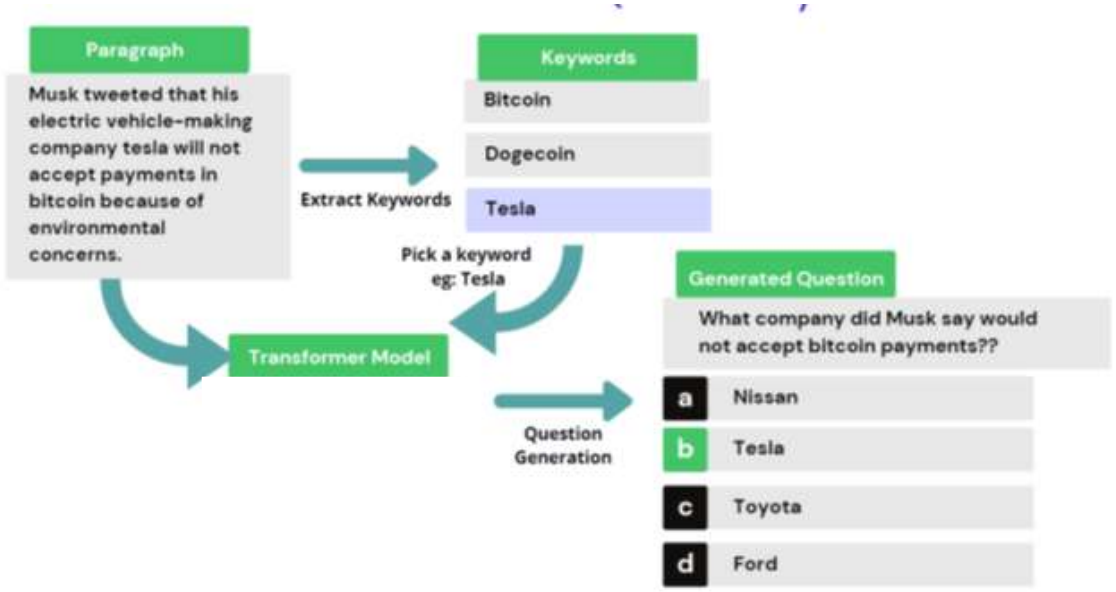
Project Title : Automatic MCQs generator from text
Project ID : 8
Project Team Students Name : B Pravena, Spoorthi R, Varna S

Outline

- Problem Statement
- Scope and Feasibility study
- Applications/Use cases
- Expected Deliverables
- Any other information

Problem Statement

Automatic generation of multiple choice questions based on a given piece of text.



Scope and Application of the Problem

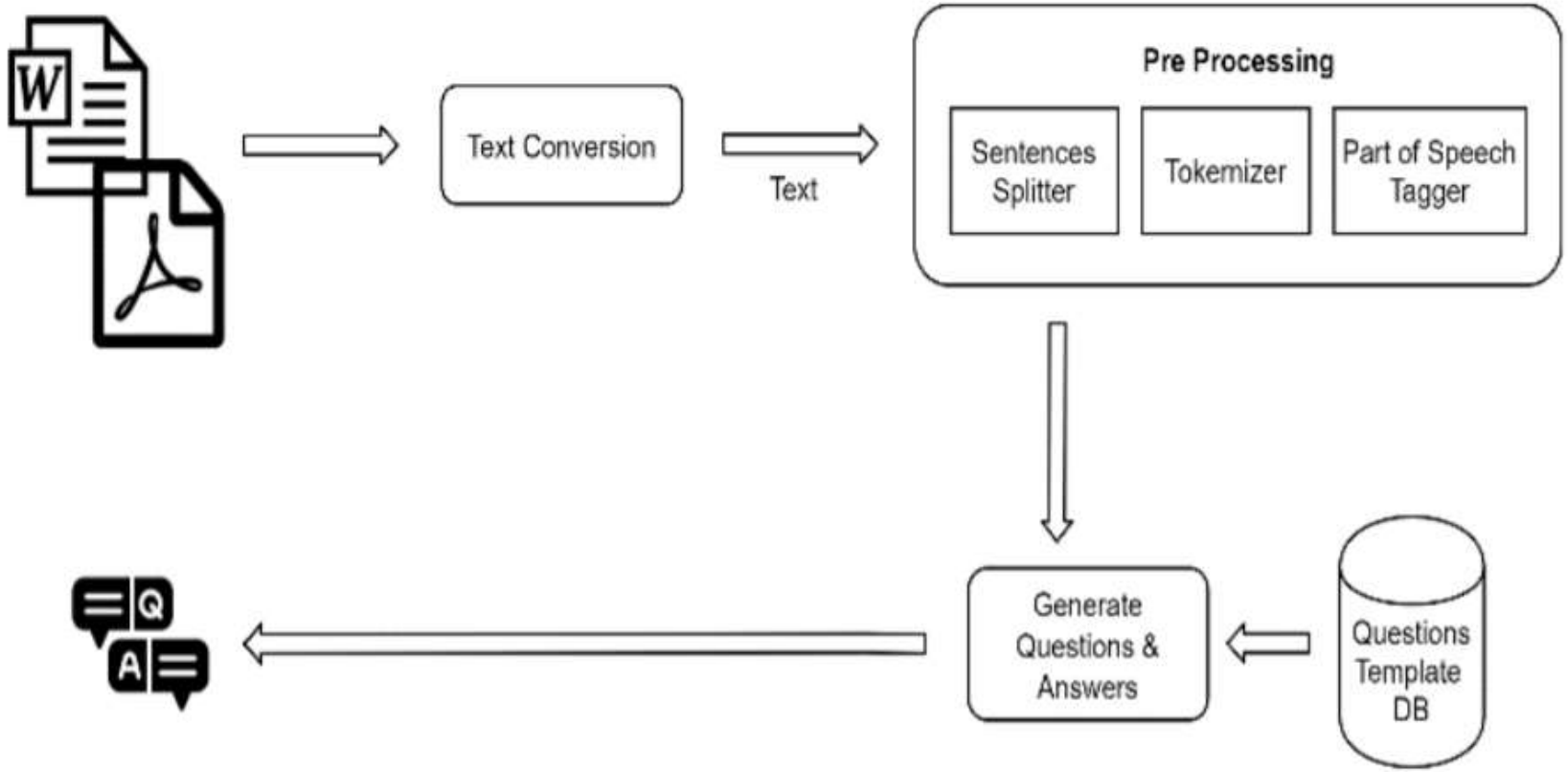
- The use of automated systems in second-language learning could substantially reduce the workload of human teachers and test creators. Automatic multiple choice question (MCQ) generation from a text is a popular research area. MCQs are widely accepted for large-scale assessment in various domains and applications. However, manual generation of MCQs is expensive and time-consuming. Therefore, researchers have been attracted toward automatic MCQ generation. Many systems have been developed for MCQ generation.

Scope and Application of the Problem

- Multiple-choice questions are the most popular assessment questions created whether it is for a school test or a graduate competitive exam. Given the increased volume of workload on teachers/assessment creators due to online learning, it would be very helpful if we can create an automated system to create MCQs.

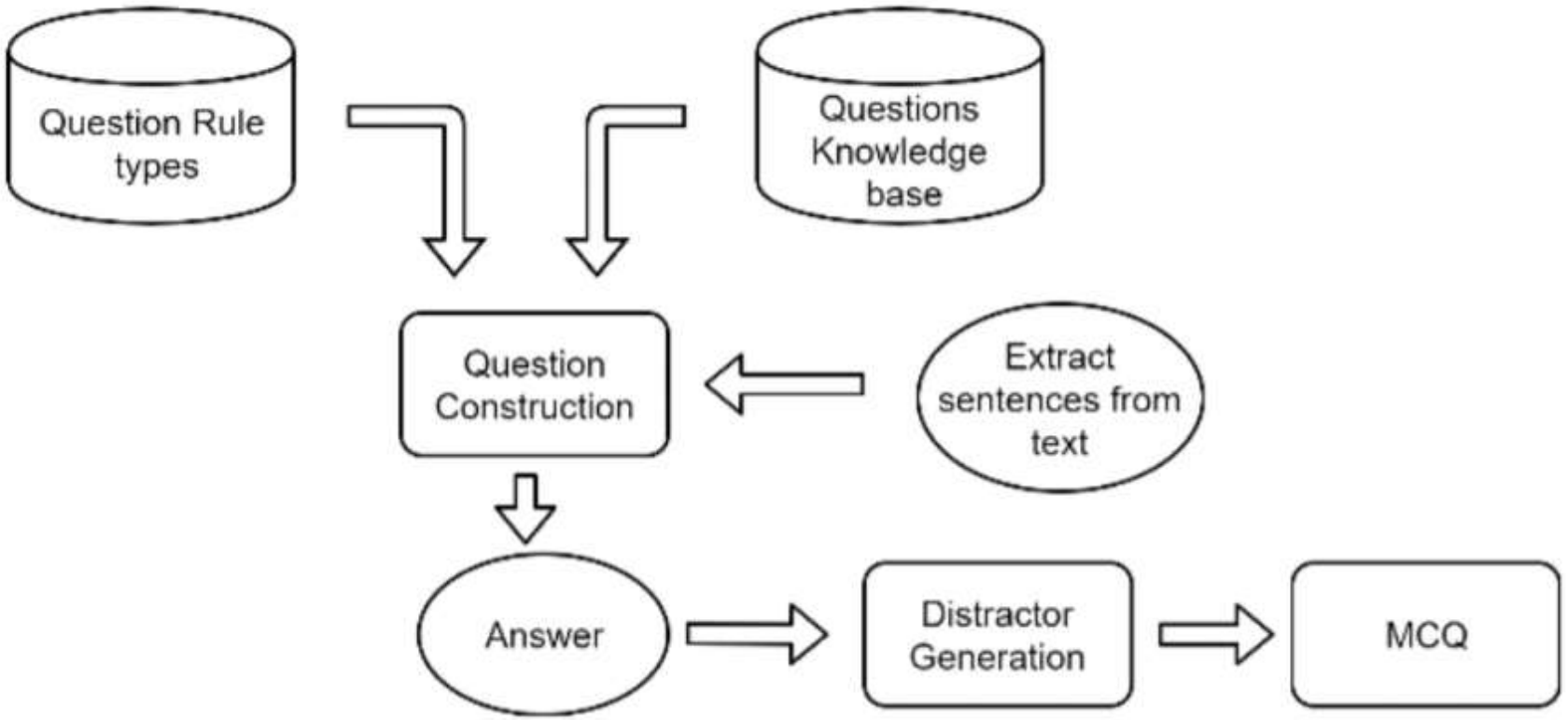
High level design

Text Extraction & Question Generation Process



High level design

MCQ Generation Process



NLP Concept used in this Project

- WordNet:
  - Wordnet labels the semantic relations among words and also captures the different senses of the word. We plan to use this for distractor generation.

- ConceptNet:
  - Conceptnet is good for generating distractors for things like locations, items, etc. which have a “Partof” relationship.

NLP Concept used in this Project

- Sense2Vec:
  - A neural network algorithm is trained with millions of sentences to predict a focus word given other words or predict surrounding words given a focus word. Through this, we generate a fixed size vector or array representation for each word. We call these word vectors.

- POS tagging to find focus word

Any other information

- We plan on first identifying focus words based on which we will produce apt questions.
- We also plan to produce good distractors/wrong choices for each generated MCQ.

Thank You