Crowd Sourcing Model for Automated Question Bank & Paper Generation

Presented by:
A-01 Devashish Bhake
A-18 Viren Joshi
A-22 Divija Kinger
A-66 Taaha Multani

SOMAIYA
VIDYAVIHAR UNIVERSITY

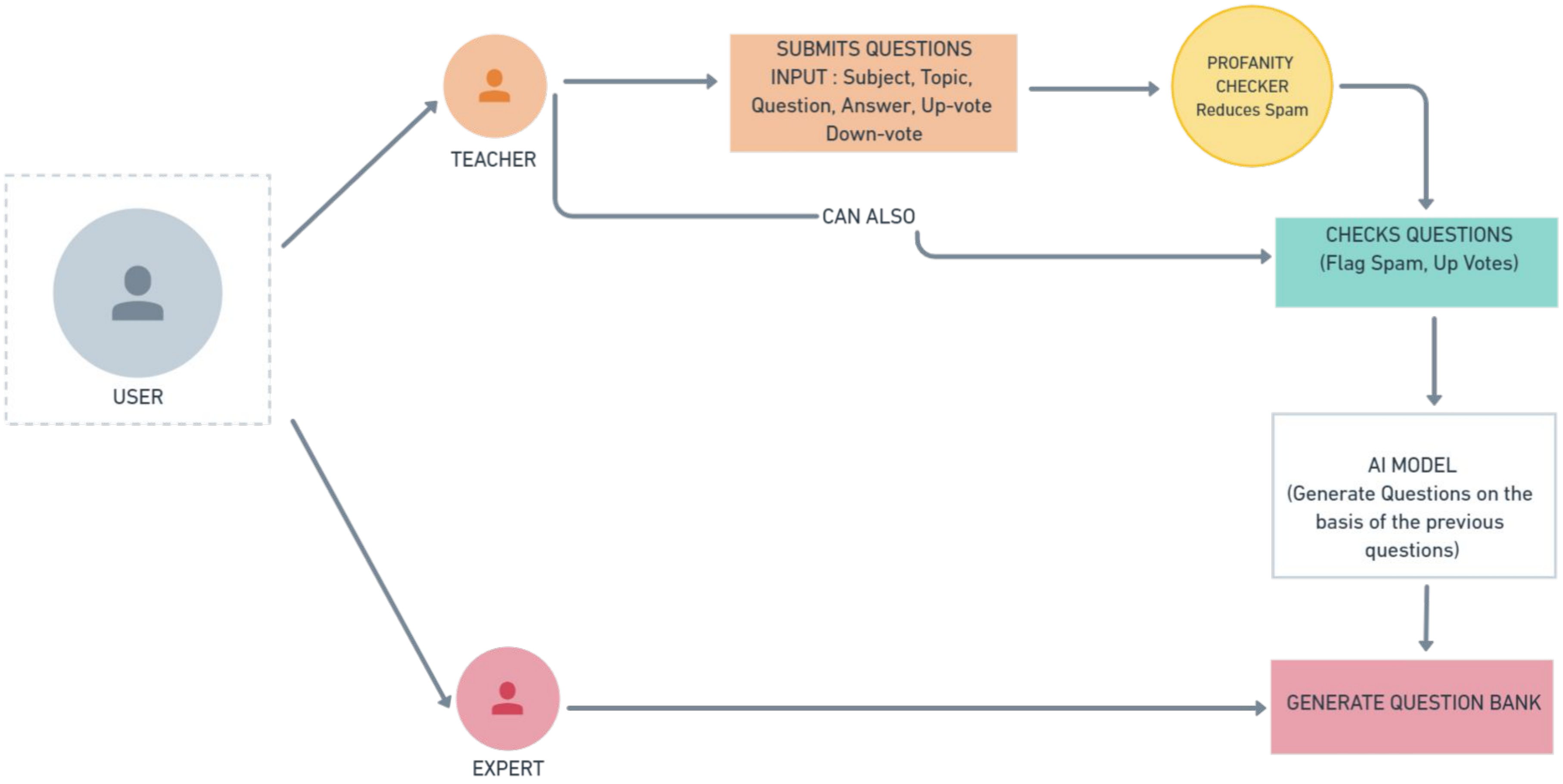
Introduction

The motive is to build the model to create a crowdsourcing platform where individual users can come together and collaborate on building a question bank while maintaining their anonymity and effectively helping in creation of question papers for all levels of education and fields.

Idea Description

- Aimed at helping teachers.
- Generate automated question banks with the help of staff all over the university.
- Deep Learning model that aims to simplify Question Bank generation process.
- Help staff improve quality of question papers and standardize the process.

Proposed Solution Structure



- Teachers: Can post questions and relevant answers to these questions. They can also validate previously posted questions.

- Subject-experts: Can generate question banks and question papers based on the marking criteria.

- Upvote System: Users have the ability to upvote if they think the question is relevant or good enough, making the question bank generating being accurate and adaptive.

- Flag System: Users also have an option to flag questions that seem irrelevant or incorrect.

Tech Stack

DATABASE
PostgreSQL

APIs & Server Side
Django

FRONTEND
HTML,CSS,JS,REACT

DATASET GENERATION

Dataset Generation Phase

Scrape Questions	Create Questions	Generate Paragraphs	Final Dataset
We scrape questions from various websites, question banks and other digital repositories for filling our dataset with questions.	We create our own questions for the dataset and assigned BT-level with the help of our project guide.	We generate paragraphs that contain the answers for the questions for training the model to create a relation between the paragraph and question and BT-level.	The final output is the dataset that contains 3 major parameters:- "Questions", "Paragraphs" and "BT-levels".

SOMAIYA
VIDYAVIHAR UNIVERSITY

QUESTION GENERATION

Model Process

Input-id Extraction	Output-id Generation	id Decoding	Question Generation
The input text that is given is the sent to the tokenizer that will extract all the ids from that input and send it for generating ids for for the question generation.	Generation of ids from the model that will be given for the further stages.	These ids that was generated in the previous stage will be decoded using the Tokenizer of the given model to generate the final Questions.	After the ids are decoded the questions are generated which are generated by taking into account the three major parameters which are Paragraphs and BT-levels.

Somaiya TRUST

QUESTION GENERATION

- Detailed / In-Depth Question Generating Model structure :

Keyword Extraction → BT-Level Relation → Creating Frame → Output Question

Keyword Extraction	BT-Level Relation	Creating Frame	Output Question
Keyword extraction, often known as keyword detection, is a text analysis technique that extracts keywords from the text. People usually use it to summarize enormous quantities of data to identify the vital points of discussion.	The required BT-Level of the given question is used to create a frame of reference on how the question is structured.	The keywords (extracted from the paragraph) and the keywords for the provided BT Level are used to create a question of the relevant structure.	The output question generated will be the crux of the input paragraph and holds the level of complexity of the given BT Level.

BT LEVEL

1 Knowledge	2 Comprehension	3 Application	4 Analysis	5 Synthesis	6 Evaluation
Define	Describe	Apply	Analyse	Construct	Appraise
List	Discuss	Demonstrate	Calculate	Deduce	Argue
State	Explain	Estimate	Categorise	Delineate	Assess
	Express	Extrapolate	Classify	Design	Choose
	Identify	Generalise	Compare	Develop	Consider
	Illustrate	Interpret	Conclude	Formulate	Decide
	Outline	Relate	Contrast	Modify	Judge
		Use	Detect	Plan	Justify
		Value	Differentiate	Prepare	Review
			Discriminate	Produce	Revise
			Distinguish	Propose	Validate
			Determine		
			Evaluate		
			Examine		
			Predict		
			Question		
			Recognise		
			Select		
			Solve		
			Test		

Literature Survey & Analysis

- Most papers use these two methods to generate questions from provided text -

Sentence Level QG

The model summarizes the paragraph of text into a few important sentences. The questions are generated then from these sentence of maximum found relevance.

SL-QG tends to oversimplify things and if the summarization of the text fails, the generation of questions is affected.

Paragraph Level QG

These models have a general hierarchical architecture for better paragraph representation at the level of words and sentences. This architecture is agnostic to the type of encoder, so the model is based on hierarchical architectures on BiLSTM and Transformers.

These tend to have a comparative good semantic accuracy but have comparatively average accuracy for contextual accuracy.

Literature Survey & Analysis

PAPER TITLE	AUTHORS	KEY TAKEAWAYS
Towards Automated Assessment Generation in e-Learning Systems Using Combinatorial Testing and Formal Concept Analysis	Frano Škopljanac-Mačina,Ivona Zakarija,Bruno Blašković	Compile a concise set of labels that describes the course material. Each question is described by a subset of those labels. Formal Concept Analysis (FCA) is used to automatically generate a corresponding formal description of the questions' set and to store it in a relational database.
Secure Automatic Question Paper Generation with the Subjective Answer Evaluation System	Ragasudha Ragasudha, M. Saravanan	Use of random algorithms and Bloom's Taxonomy. In the random algorithm, the question will not repeat in the same question paper again and again. Bloom's taxonomy is the one of methods to classify the learning purpose. It will classify the question based on remembering, analyzing, skilled, evaluating, justify and storing in the database by the user.

SOMAIYA VIDYAVIHAR UNIVERSITY

Literature Survey & Analysis

PAPER TITLE	AUTHORS	KEY TAKEAWAYS
Automatic Multiple Choice Question Generation From Text : A Survey	Dhawaleswar Rao CH, Sujan Kumar Saha	Text is preprocessed via lexical analysis, sentence-simplification, sentence-structure analysis, normalization, etc. The following stage is to choose sentences that can be used to spark queries. Only statements containing ambiguous facts will be chosen. To create MCQs, keywords are taken out of these sentences. Declarative sentences are transformed into questions. Finally, alternatives for the MCQs are created using distractors.
Automatic Chinese Multiple Choice Question Generation Using Mixed Similarity Strategy	Ming Liu , Vasile Rus, and Li Liu	Machine learning approach to the task of generating Chinese MCQ distractors. Measurement formula for Chinese character visual resemblance. This approach generated a similarity score between matched smaller components after iteratively breaking down a complex Chinese character into smaller components.

Somaiya TRUST

References

- https://arxiv.org/pdf/1911.03407.pdf
- https://ieeexplore.ieee.org.library.somaiya.edu/stamp/stamp.jsp?tp=&arnumber=8703745
- https://ieeexplore.ieee.org/document/9393337
- https://ieeexplore.ieee.org/document/9419410
- https://ieeexplore.ieee.org/document/9068704
- https://ieeexplore.ieee.org/document/6421938
- https://ieeexplore.ieee.org/abstract/document/7814813?reload=true

Somaiya TRUST

SOMAIYA
VIDYAVIHAR UNIVERSITY

SOMAIYA
VIDYAVIHAR UNIVERSITY

Somaiya
TRUST