Embibe Problem Statement

1. About Embibe

Embibe is India's leading edtech platform powered by AI. We deliver personalised learning and predictable learning outcomes for each aspiring individual. We have built world class products that enable, better education delivery and personalised guidance, across each stakeholder - the student, the educator/educational institutions and the parent. We are on a journey to enable a quarter of India's population to excel in education: 300 million learners across 1.5 M institutions, including vernacular users. In our quest of discovering 'what matters in education', we have touched the lives of over 18 million students across different strata of society. Each Embiber is passionate towards making a meaningful impact on the way we learn, process and educate ourselves. We are highly transparent in our goals, our communication, feedback and results. Aditi Avasthi, the founder CEO, leads the company's vision of building the future of learning and education, personalised to each and available to all.

2. Statement

Scraping the question from any other site is the outdated job now. It is more interesting if we are able to generate the questions from text-book or any other content.

Definition: You are given a set of books, you need to generate the questions out of that for different grades and levels. Possible types of question can be any of

- Fill in the blanks
- Question that starts with *W / H*.
- Generic questions(genericity depends on the subjects).

3. FAQs

1. Will there be Mentors to guide during the hackathon?

   Yes, company officials will be there to help throughout megathon.

2. Will the training dataset be provided ?

   No, you are expected to find (or synthesise) data sets from your end. Apart from using abstract/full text of research papers, other technical resources can also be used.

3. Will you provide additional computing power (external GPUs, Server access, etc)?

   No you are expected to bring and work on your own computers only. No additional computing resources will be provided from our end.

4. Submission Format

- Presentation conveying your core idea

- Analysis of your idea's performance
- Code / Github repository link

5. Evaluation Criteria

Quality of the generated questions. Ratio of similarity of questions. Semantically same question with different constants and units and dimensions. Total number of questions generated with respect to the input text.

6. Resources

The 4 segments of QG(question generation)

- Sentence selection
- Coreference resolution
- Blank selection
- Distractor selection

There are several methods for coref. Resolution like: simple rule based, dependency parser based, neural net based, etc. Please find below some papers.

Automatic Gap-fill Question Generation from Text Books

Automatic Generation of Context-Based Fill-in-the-Blank Exercises

Learning to Automatically Generate Fill-In-The-

Generating Questions and Multiple-Choice Answers using SemanticAnalysis of Texts