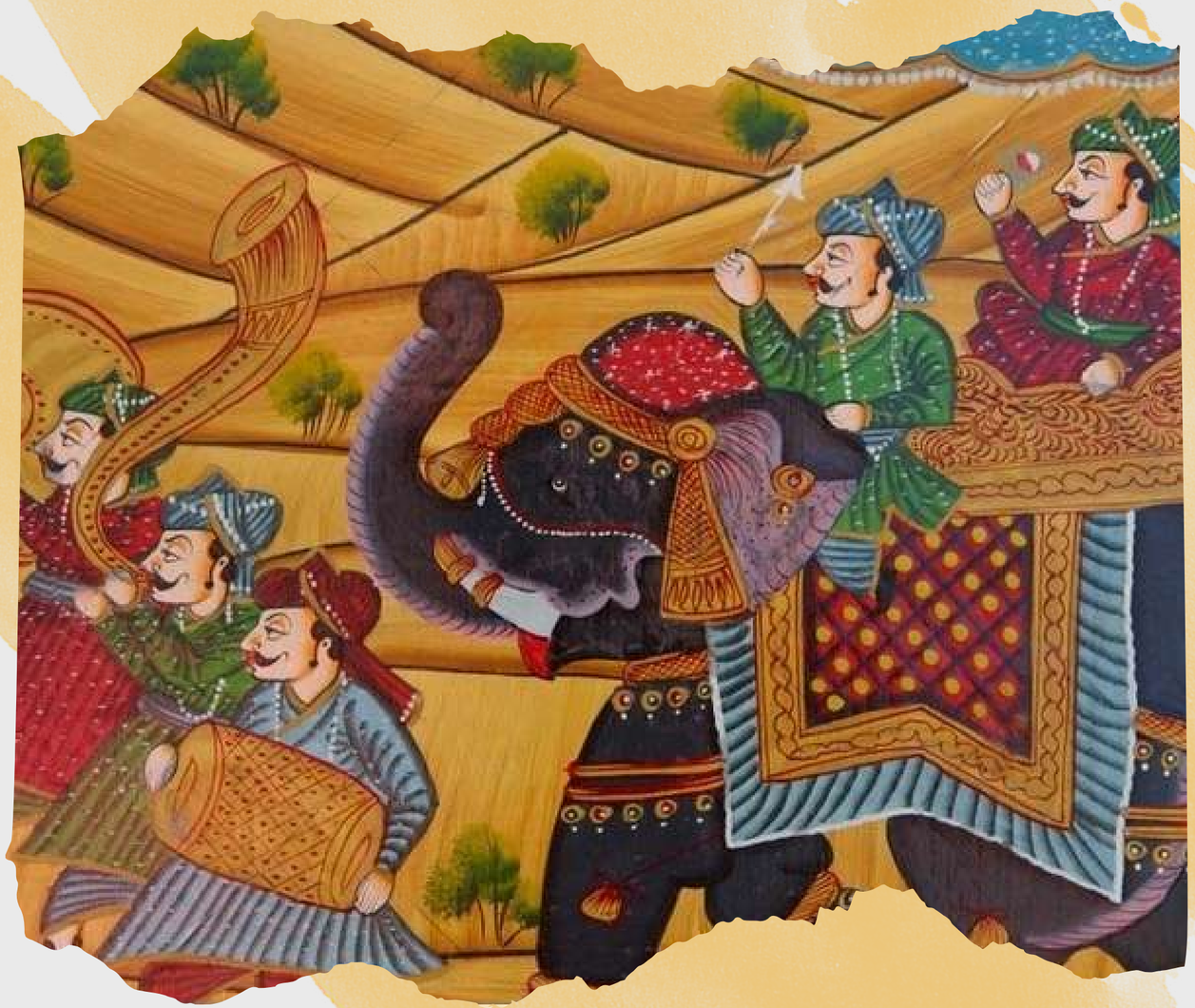
Question Generation from Fable Stories



PHASE 1

Team - DoNut Give Up

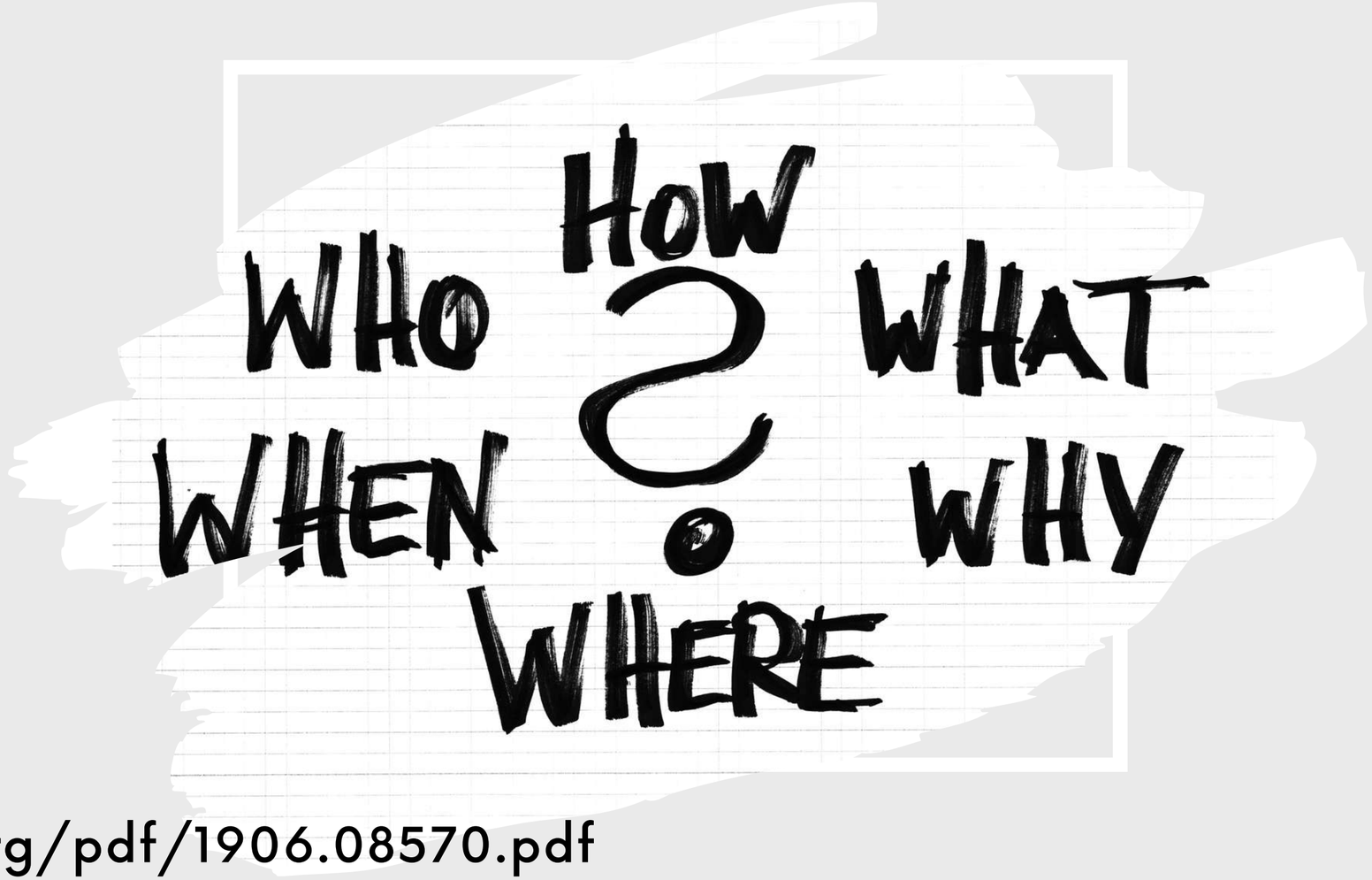
Gargi Shroff

Swarang Joshi

Ketaki Shetye

Introduction to Fable Stories

Fables are short stories that typically feature animals as characters and convey moral lessons. They often end with a moral or lesson.



Resource link: **https://arxiv.org/pdf/1906.08570.pdf**

Project Timeline

PHASE 1

Presenting the general outline of the project and procedure to be followed.

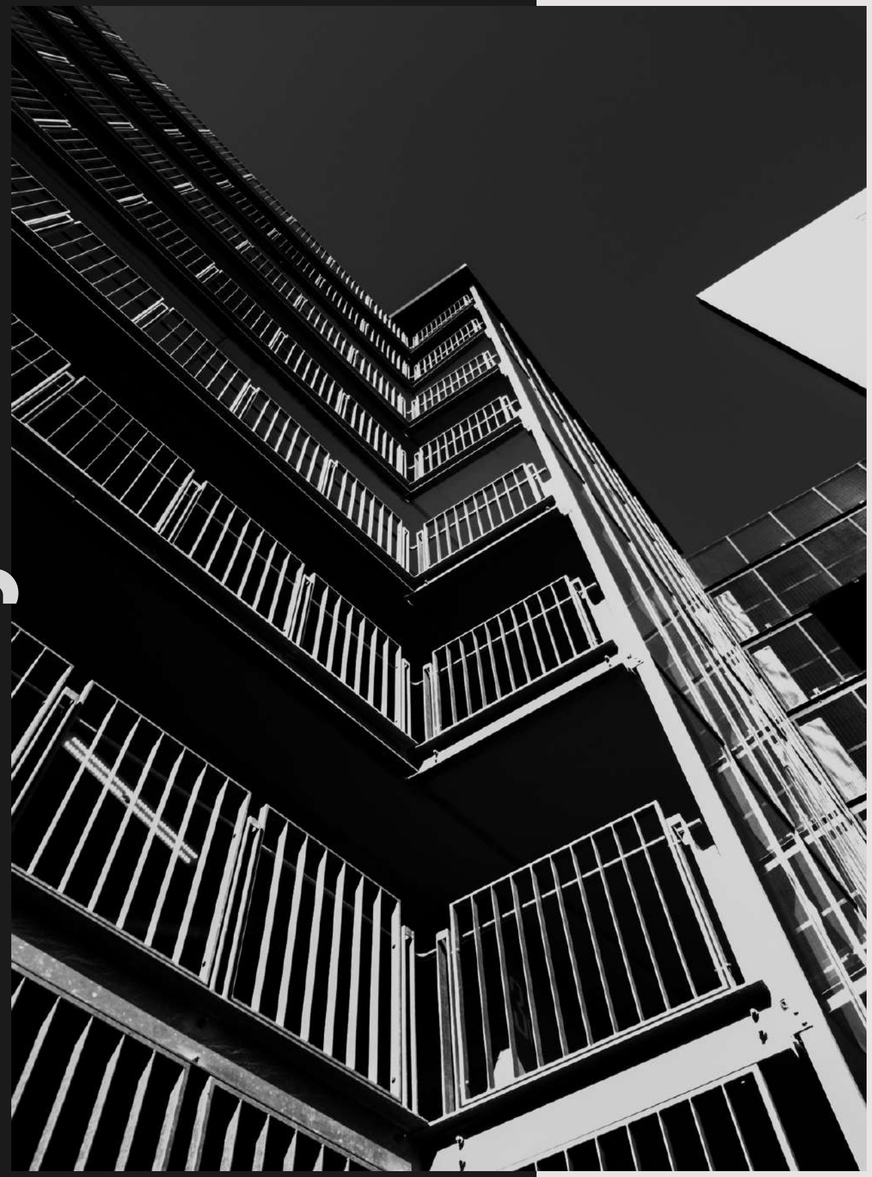
PHASE 2

T.B.D.

PHASE 3

T.B.D.

Flow of The Project



Collecting Dataset

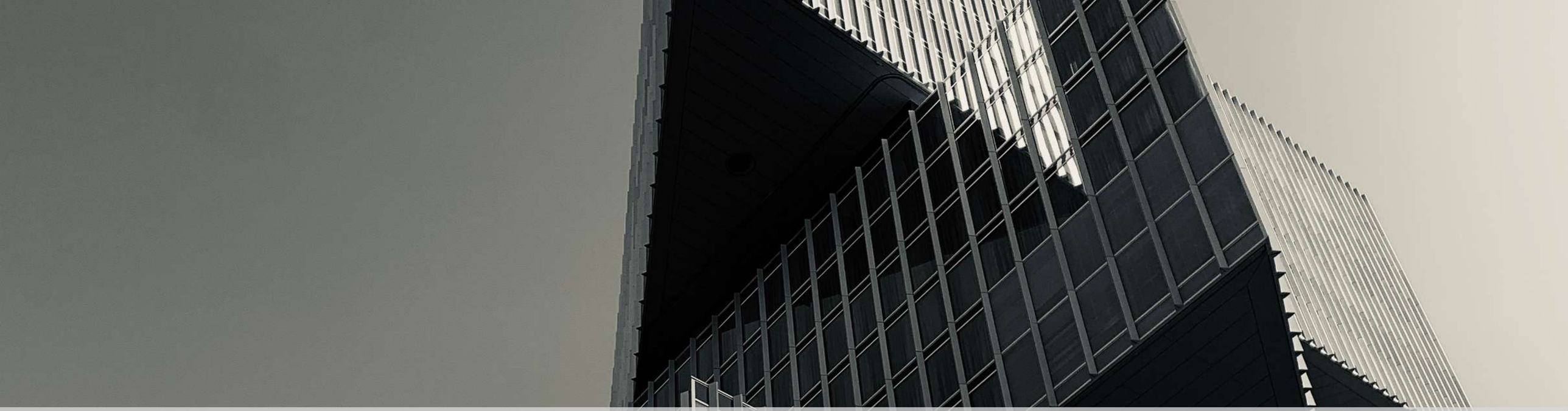
Preprocessing

Text Analysis

Rules Formation on the Basis of Dataset

Creation of Question Generator using Python

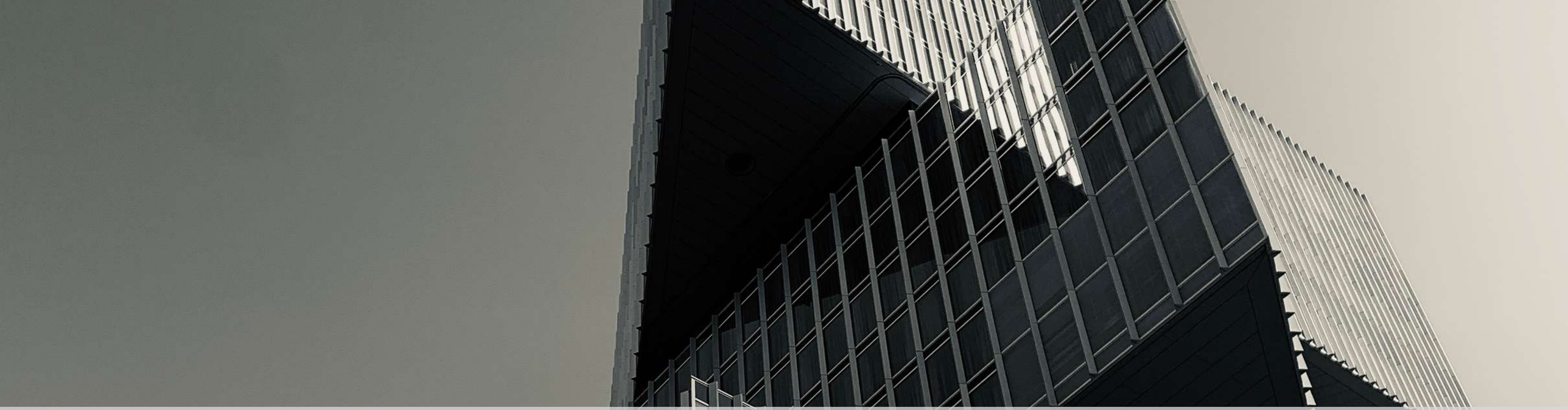
Analyzing the Results and also formulating the exceptions encountered



Collecting Dataset

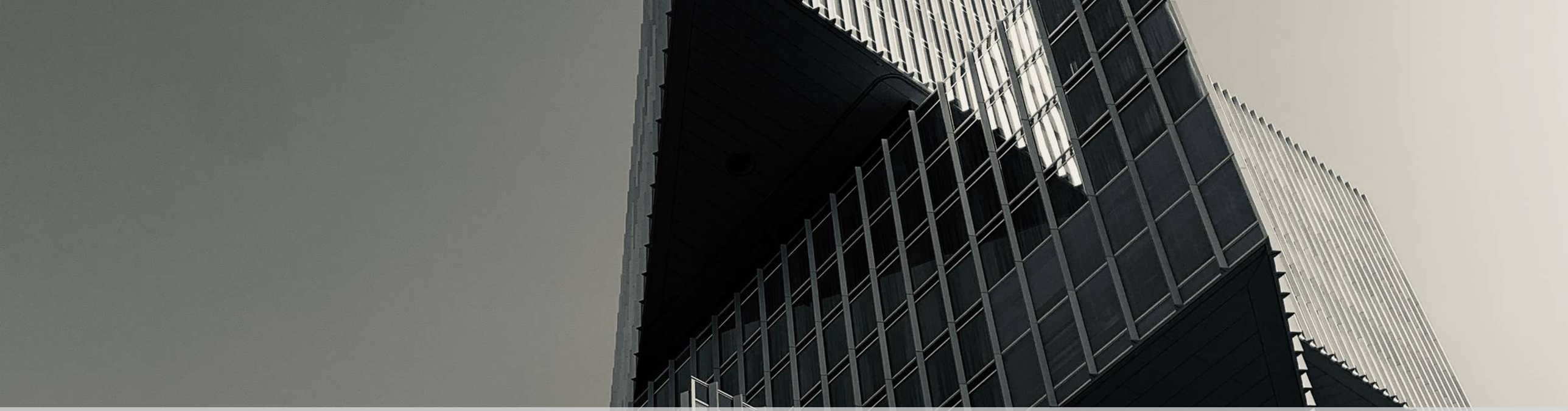
We collected around 30-35 Hindi Fable stories from the resources stated below.

Hindi Fable Stories Website - https://www.hindisahityadarpan.in/2016/06/panchatantra-complete-stories-hindi.html



Pre-processing

We will start with filtering the text by removing the punctuation marks,stop words,converting the text to lower-case and do text cleaning for ease of analysis and function.

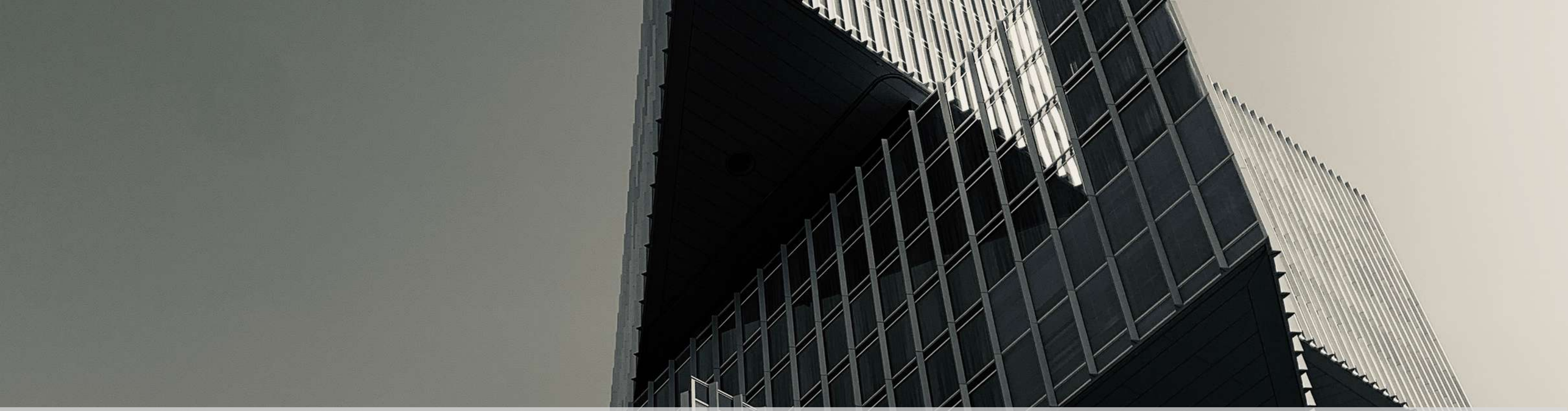


POS Tagging/Parser

Next task upon the list is POS Tagging of the considered dataset. For this we referred to 2 resources online.

LTRC Shallow Parser - http://ltrc.iiit.ac.in/analyzer/hindi/index.cgi

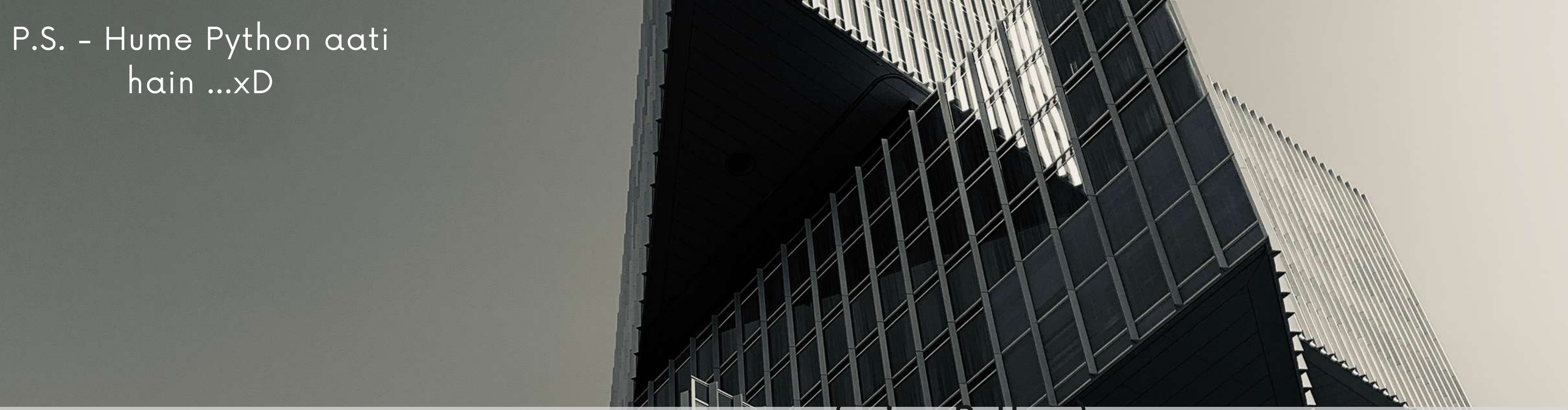
POS Tagger for Hindi Language - https://github.com/JayeshSuryavanshi/POS-Tagger-for-Hindi-Language.git



Rule Formation

Rules formulation is the most important task in a rule-based question generation project because it determines the quality and effectiveness of the generated questions.
Think of rules as the building blocks of the system - they are the instructions that tell the system what to do and how to do it. If the rules are poorly defined or inaccurate, the system will generate low-quality questions that may not be relevant or useful.

P.S. - Hume Python aati hain ...xD



(using Python)

Creation of QuesGen -

When it comes to rule-based question generation in Python, there are several ways to implement the rules:

- If-else statements: For the structure the structure "X verb Y," may lead to"What did X do to Y?"
- Template-based: Da Vinci painted Mona Lisa.question = "Who painted the " + subject + "?"
- Syntactic-based:Relies on analyzing the syntax.
- Keyword-based: Involves identifying keywords. eg-people, dates, and locations

In this example, we have a pre-context "राम बाजार में है।" (Ram is in the market.) and a text "वह घर पर है।" (He is at home.). We want to generate a question based on the text and include the pre-context in the question.

Using if-else statements, we check if the word "है" (is) is present in the text. If it is, we generate the question "कौन घर पर है?" (Who is at home?). Otherwise, we generate the question "वह किधर है वह?" (Where is he?) using the subject.

QuesGen Lite®

```
# Approach 1 : If-Else Statements

# Define the rules using if-else statements in Hindi
# Detect the subject through parsing

precontext = "राम बाजार में है।"
text = "वह घर पर है।"
subject = "राम","वह"

# Generate a question based on the text using if-else statements
if "है" in text:
    question = "कौन घर पर है?"
else:
    question = "वह किधर है " + subject + "?"

# Add precontext to the question
question = precontext + " " + question

# Output the generated question in Hindi
print(question)
```

The code snippet provided performs question generation using a template-based approach. The code defines a question template in Hindi, which includes placeholders for the verb and subject. The verb and subject are then filled in using specific information extracted from the text data. Finally, the generated question is outputted to the console.

QuesGen Lite®

```
# Approach 2 : Template Based

# Preprocessing and text analysis steps go here

# Define the question template in Hindi
question_template = "कौन/क्या [past-tense verb] [subject]?"

# Fill in the template with specific information from the text data in Hindi
question = question_template.replace("[past-tense verb]", "किसने बनाया").replace("
[subject]", "मोना लिसा")

# Output the generated question
print(question)
```

This code snippet demonstrates a simple example of generating a question in Hindi based on syntactic analysis. The sentence "जॉन ने सेब खाया" is analyzed to extract the subject ("जॉन"), verb ("खाया"), and object ("सेब"). The code then uses this information to generate a question in the format of "क्या [subject] ने [verb]?", which in this case becomes "क्या जॉन ने खाया?" Finally, the generated question is printed to the console.

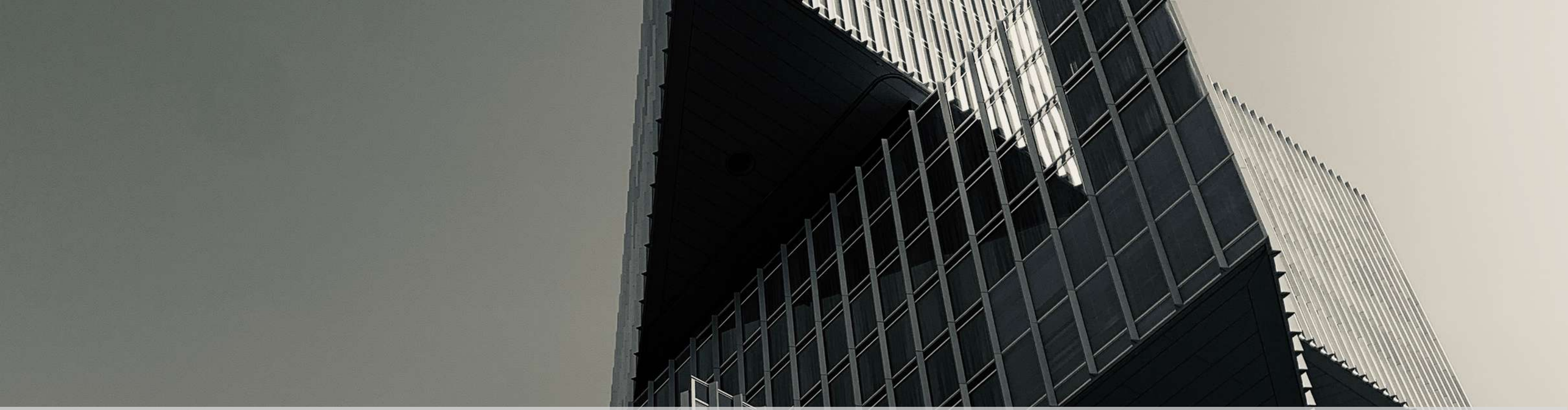
QuesGen Lite®

```
# Approach 3 : Syntactic Based

# Preprocessing and text analysis steps go here

# Define the sentence in Hindi
sentence = "जॉन ने सेब खाया"

# Extract the subject, verb, and object from the sentence in Hindi
subject = "जॉन"
verb = "खाया"
object = "सेब"

# Generate the question based on the syntax of the sentence in Hindi
question = "क्या " + subject + " ने " + verb + "?"

# Output the generated question in Hindi
print(question)
```

In this example, the script looks for the keywords "बनाया," "लिखा," "क्या," and "कौन" in the given Hindi text. If it finds a match, it generates a question based on the keyword and the subject of the sentence.

QuesGen Lite®

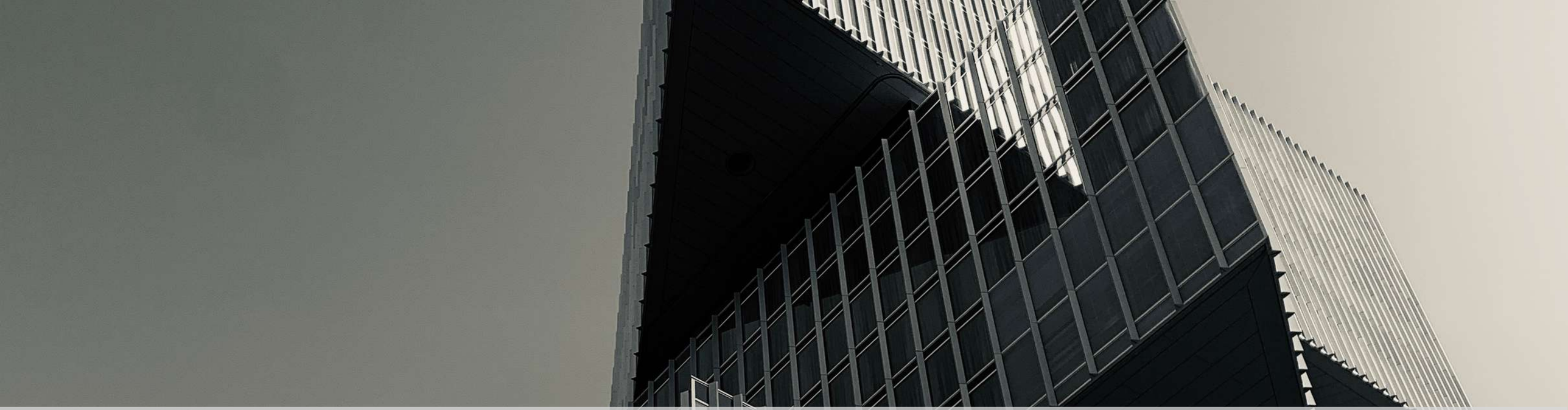
```
# Approach 4 : Keyword Based

# Preprocessing and text analysis steps go here

# Define the keywords to look for in the text data in Hindi
keywords = ["बनाया", "लिखा", "क्या", "कौन"]

# Loop over the keywords and generate a question
for keyword in keywords:
  subject = extract_subject(text)
    if keyword in text:
        if keyword == "बनाया"
            question = "कौन ने " + subject + " बनाया?"
            break
        elif keyword == "लिखा":
            question = "कौन ने " + subject + " लिखा?"
            break
        elif keyword == "क्या":
            question = "वह क्या है " + subject + "?"
            break
        elif keyword == "कौन":
            question = "कौन " + keyword_action(text) + " " + subject + "?"
            break
        else
            question = default_question_forSubject

# Output the generated question in Hindi
print(question)
```



Analysing Results

We will do the post-processing for filtering out irrelevant questions and phrases manually.We will aso take help of automated machines like ROGUE,BLEU for the same.

We will analyse the obtained results and also look for exceptions to our model. We will further try to compare the results between the different languages chosen and document the proficiency of the model.



Languages to be used

We will start with the dataset for Hindi and implement it based on the rules formulated by us on Python.

We will also try to include more languages like Marathi, Gujarati after the effectiveness of model is determined.

THANK YOU FOR EXPLORING THE WORLD OF QUESTION GENERATION USING FABLES WITH US!