Automatic Question Generation from Sentences

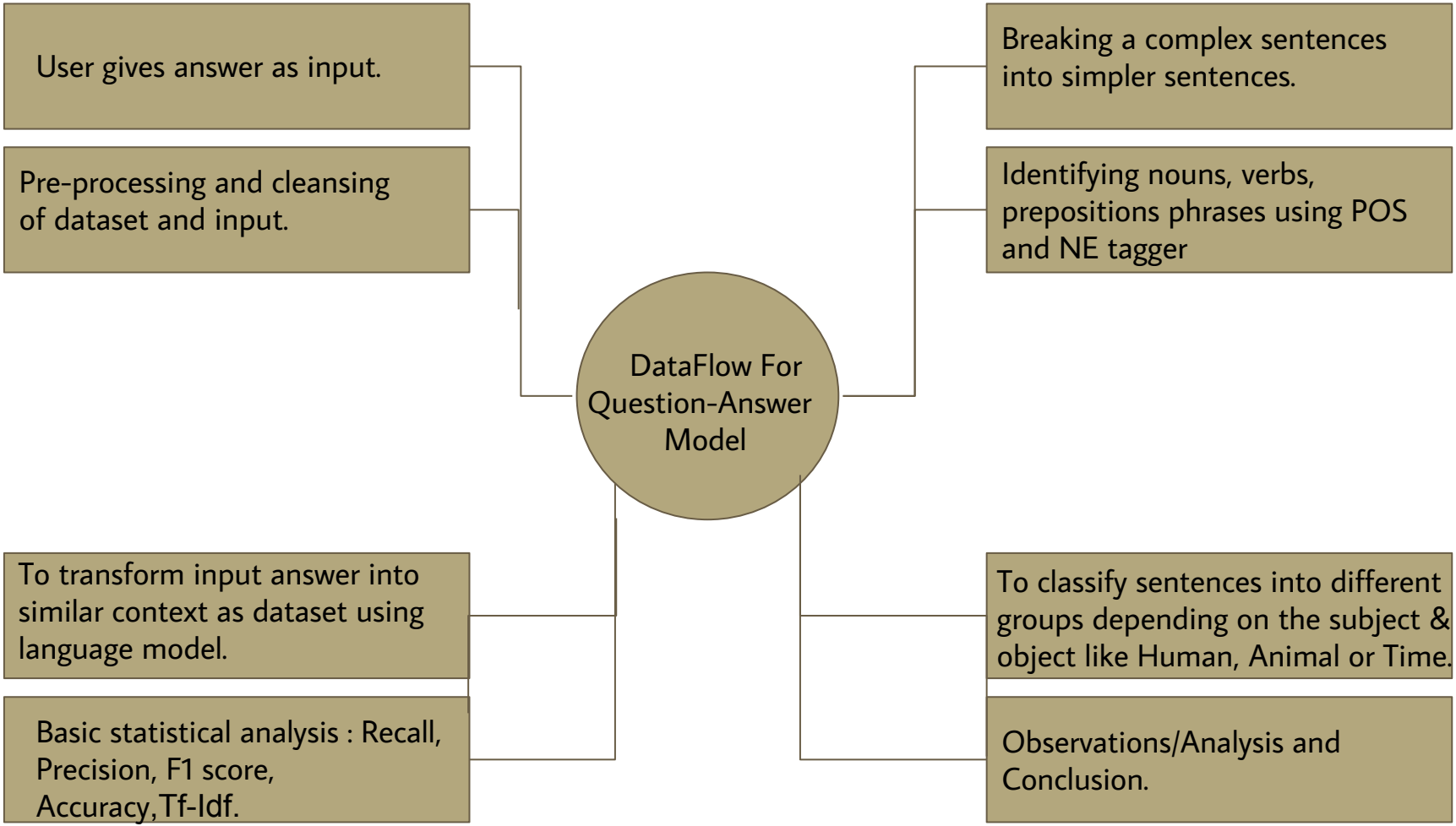
Yashika Goel.
Rhythm Nagpal.
Rashmi Nagpal.

Problem Statement.

- To develop statistical model for deriving multiple similar questions from question-answer archive; given answers as query.

Motivation.

- To generate quizzes on a particular topic; it's related terms needed to be identified to frame questions.
- To authenticate authorship of a person when they claim their copyright.
- To detect loopholes in an article, documentary or some literature. <3
- Enquiry systems for generating the perfect query.


User gives answer as input.
Pre-processing and cleansing of dataset and input.
DataFlow For Question-Answer Model
Breaking a complex sentences into simpler sentences.
Identifying nouns, verbs, prepositions phrases using POS and NE tagger
To transform input answer into similar context as dataset using language model.
Basic statistical analysis : Recall, Precision, F1 score, Accuracy,Tf-Idf.
To classify sentences into different groups depending on the subject & object like Human, Animal or Time.
Observations/Analysis and Conclusion.

Precision, Recall & F1 Measure.

Question Type (Number of Questions)	Questions Generated	Actual Questions	Recall Value
Who	91	71	0.282
Whom	12	7	0.714
What	92	118	0.22
Where	3	22	0.864
When	1	18	0.333
Whose	0	9	0.111

Recall value on QGSTEC Dataset.

Question Type (Number of Questions)	Questions Generated	Annotated Questions	Precision Value
Who	91	42	0.167
Whom	12	7	0.714
What	92	54	0.704
Where	3	20	0.85
When	1	18	0.33
Whose	0	4	1.0

F1 Measure score is 0.5075

Jaccard Similarity & Cosine Similarity.

Jaccard Similarity = 0.3952011

Cosine Similarity = 0.5398

System Module.

Segment Identifier

Divide the complex sentence into simpler.

Tokenizer

Tokenize it using NLTK Tokenizer.

NLTK POS Tagger

Identify TAGS using POS Tagger.

Clause Identification

Find clause, which is defined by noun phrase followed by verb phrase.

NER Tagger/ QG Module/ VB

Feed in the sentence in our rules, defined to generate questions.

Question Sense Disambiguation.

Truncate disambiguation in questions.