



QUESTION GENERATION for DISCUSSION FACILITATION



Biman Gujral : DA-IICT (biman_gujral@daiict.ac.in)
Ashudeep Singh : IIT Kanpur (ashudeep@iitk.ac.in)
Mentor : Dr. Carolyn P. Rosé

Radhika Kedia : DA-IICT (radhika_kedia@daiict.ac.in)
Divyanshu Bhartiya : IIT Kanpur (divbhar@iitk.ac.in)
Mentor : Mr. David Adamson

MOTIVATION

In contrast to IRE (Initiate-Response-Evaluate) pattern, discussion facilitates understanding of the concepts by applying their learning. Discussion helps:

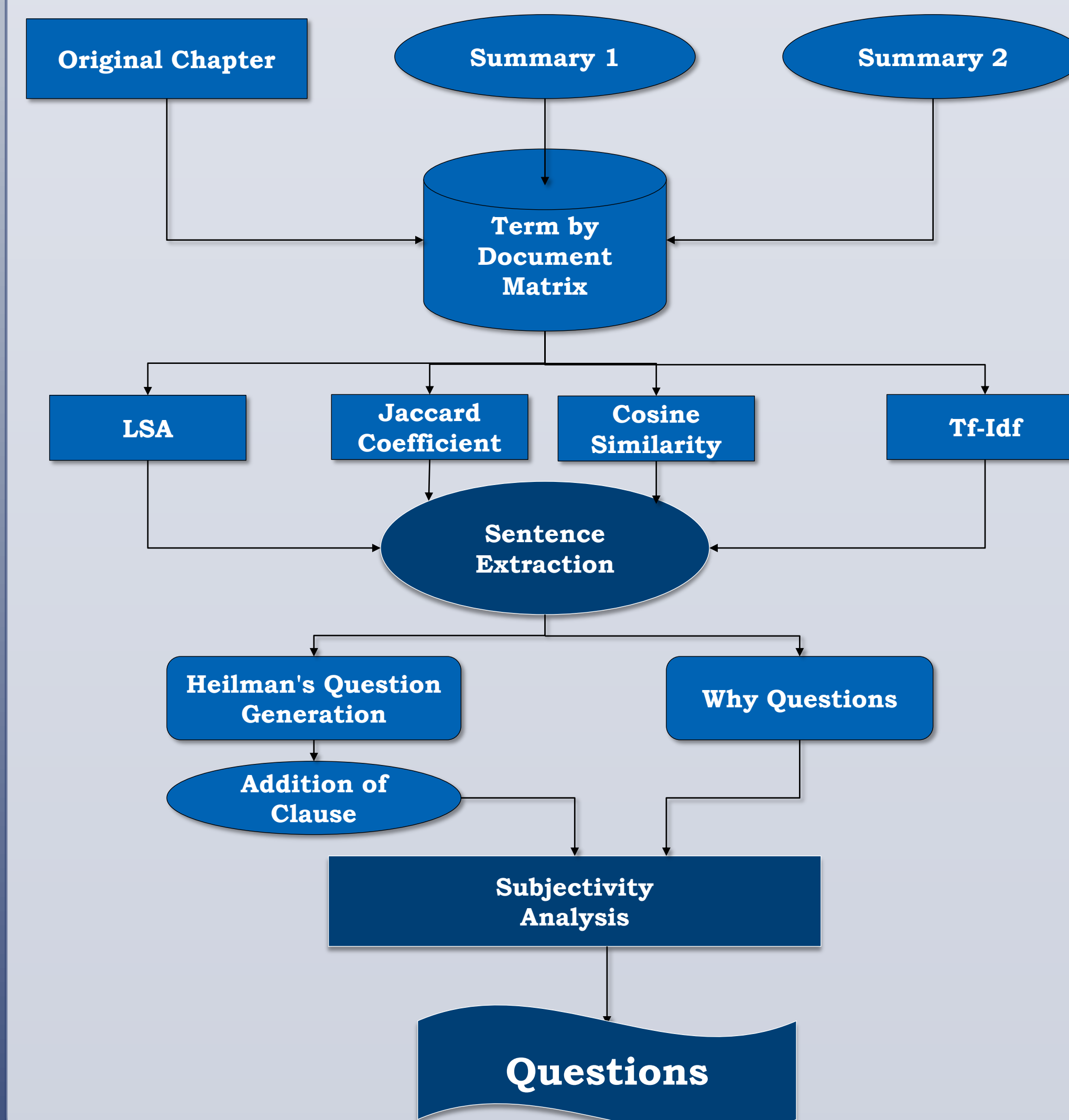
- Students construct explanations & reason effectively.
- Students learn how to communicate within a group.
- Cross questioning helps in imbibing knowledge and better retention.

VISION: To generate questions automatically from source texts that promote rich discussion.

PRIOR WORK

- Automatic Gap-Fill Question Generation from Text Books by *Manish Agarwal and Prashanth Mannem*
- Question Generation via Over-generating Transformations and Ranking by *Michael Heilman and Noah A. Smith*
- Question Generation from Paragraphs at Upenn : QGSTEC system description
- Goals and Strategies of a Problem-based Learning Facilitator by *Cindy E. Hmelo-Silver, Howard S. Barrows*

Our APPROACH

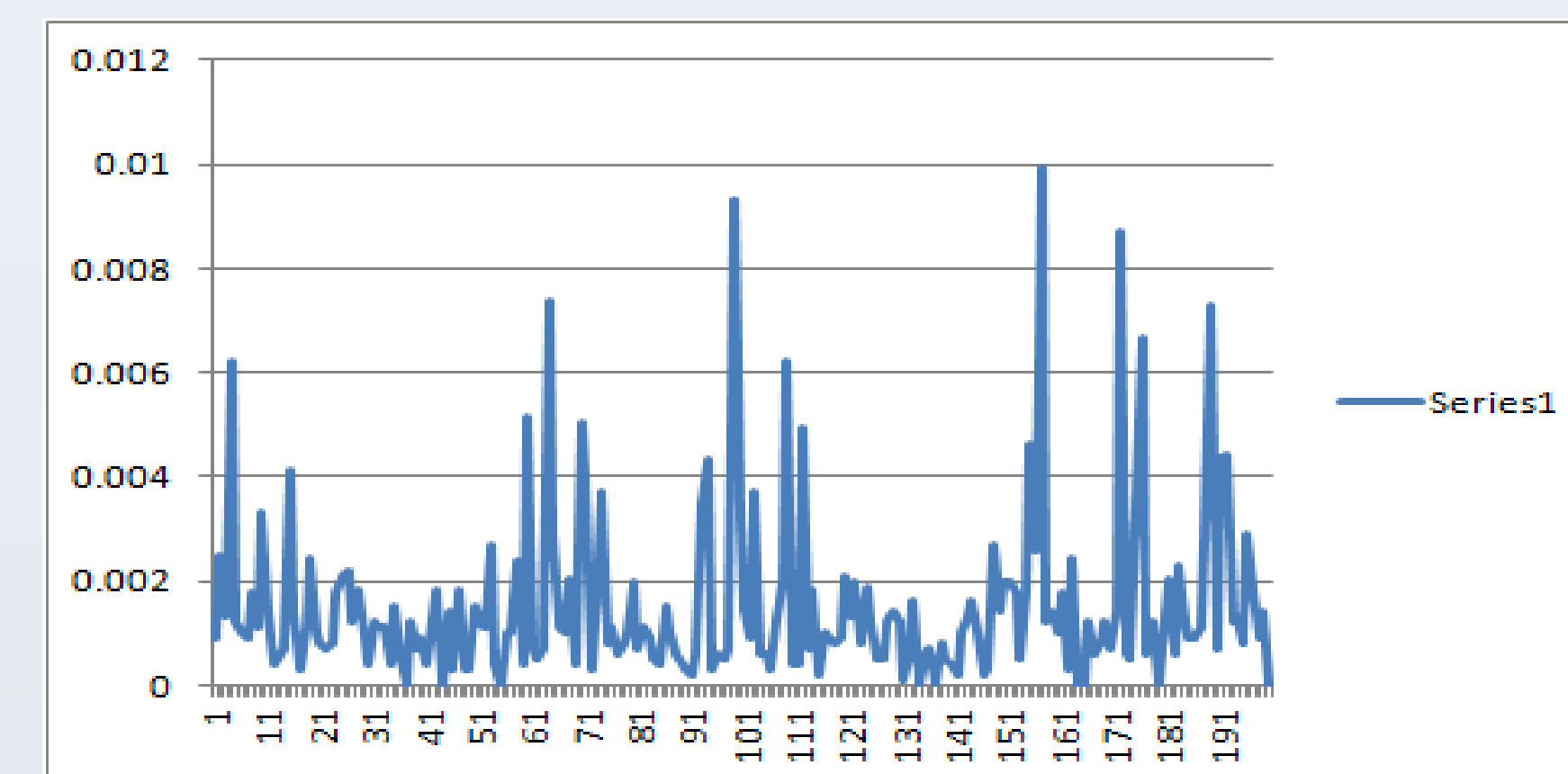


PHASE 1

SENTENCE EXTRACTION

Why Sentence Extraction ?

- Sentence Extraction is used to get those sentences that are the best abstraction of chapter.
- Sentence Extraction from summaries avoids the problem of redundancy and getting the best representation of chapter.

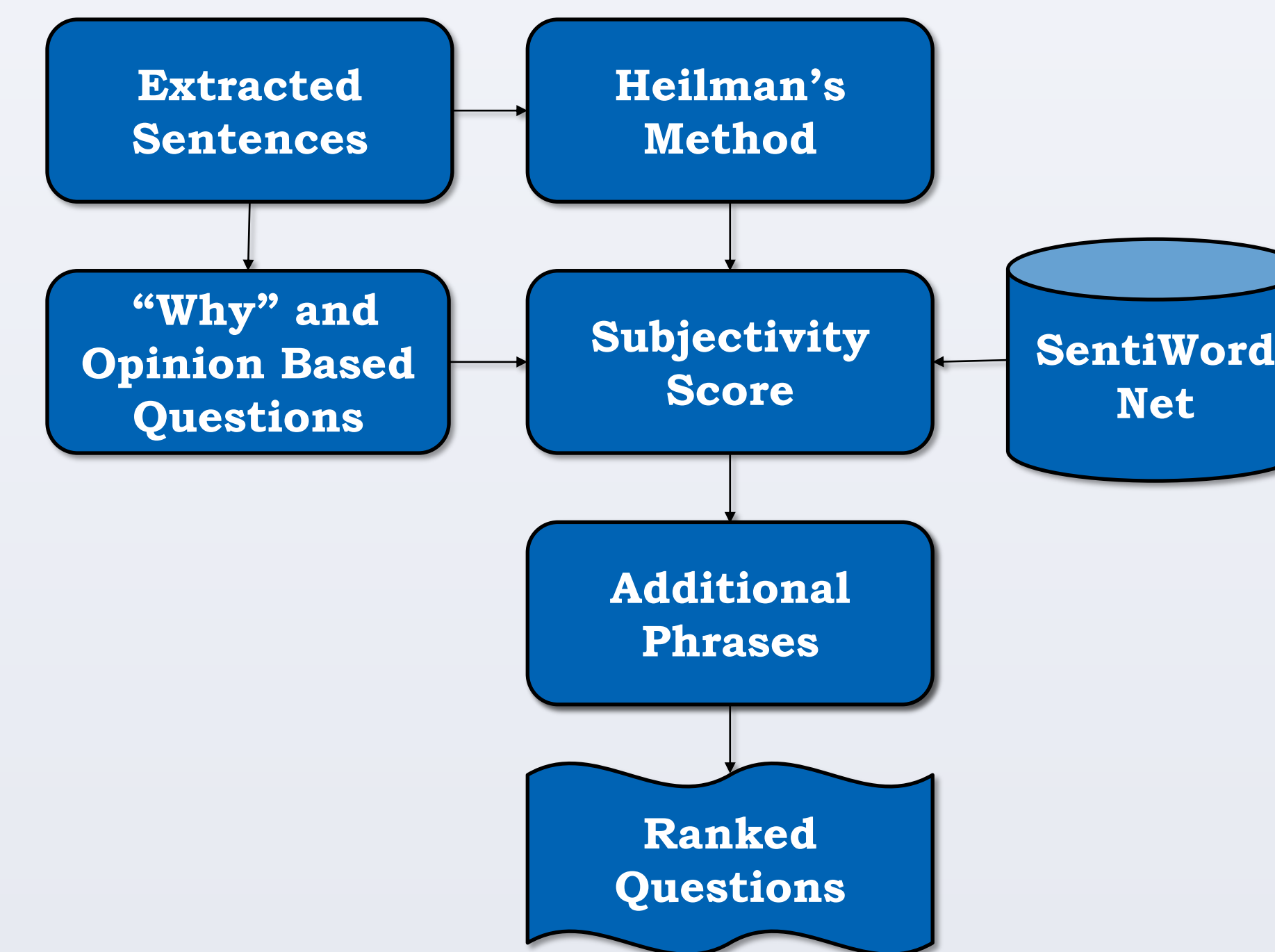


TECHNIQUES USED :

- **LSA (Latent Semantic Analysis)**
 - Analyzes relationship between terms and documents
 - Given a term by document matrix, it creates a n dimensional vector for each document
- **NORM of DOCUMENT VECTOR :**
 - The norm of document vector obtained after LSA represents the amount of total information contained in it.
 - $|d| = \sum (w_i^2)$
- **TFIDF (Term Frequency – Inverse Document Frequency)**
 - Tf-Idf weighs the frequency of a term t in a document d with a factor that discounts its with its appearances in whole corpus.
 - Measure of importance of words in a document
- **JACCARD COEFFICIENT :**
 - Measure similarity between documents
 - Similarity defined as the intersection divided by union
 - Measure of similarity between summary and document
- **COSINE DISTANCE :**
 - Measure similarity between documents as the angle between two document vectors
 - Measure of context in summary as used with the previous sentence in summary

PHASE 2

QUESTION GENERATION



Heilman's Method :

- Selects sentences from the passage as candidates for question generation
- Generates direct and objective questions, used in comprehension passages
- Doesn't generate conceptual questions which can trigger discussion
- No implementation of 'why' and 'opinion-based' questions

QUESTIONS GENERATED BY HEILMAN'S METHOD

Who play an inordinate role in helping new trends?
Do Connectors begin to tip?
How many rules of epidemics that Gladwell identifies are: the Law of the Few, the Stickiness Factor, and the Power of Context?
Do Connectors play an inordinate role in helping new trends?
Are the three rules of epidemics that Gladwell identifies: the Law of the Few, the Stickiness Factor, and the Power of Context?
What do Connectors play?
Where do Connectors begin?
What do Connectors begin to?
Who begin to tip?

Why Subjectivity ?

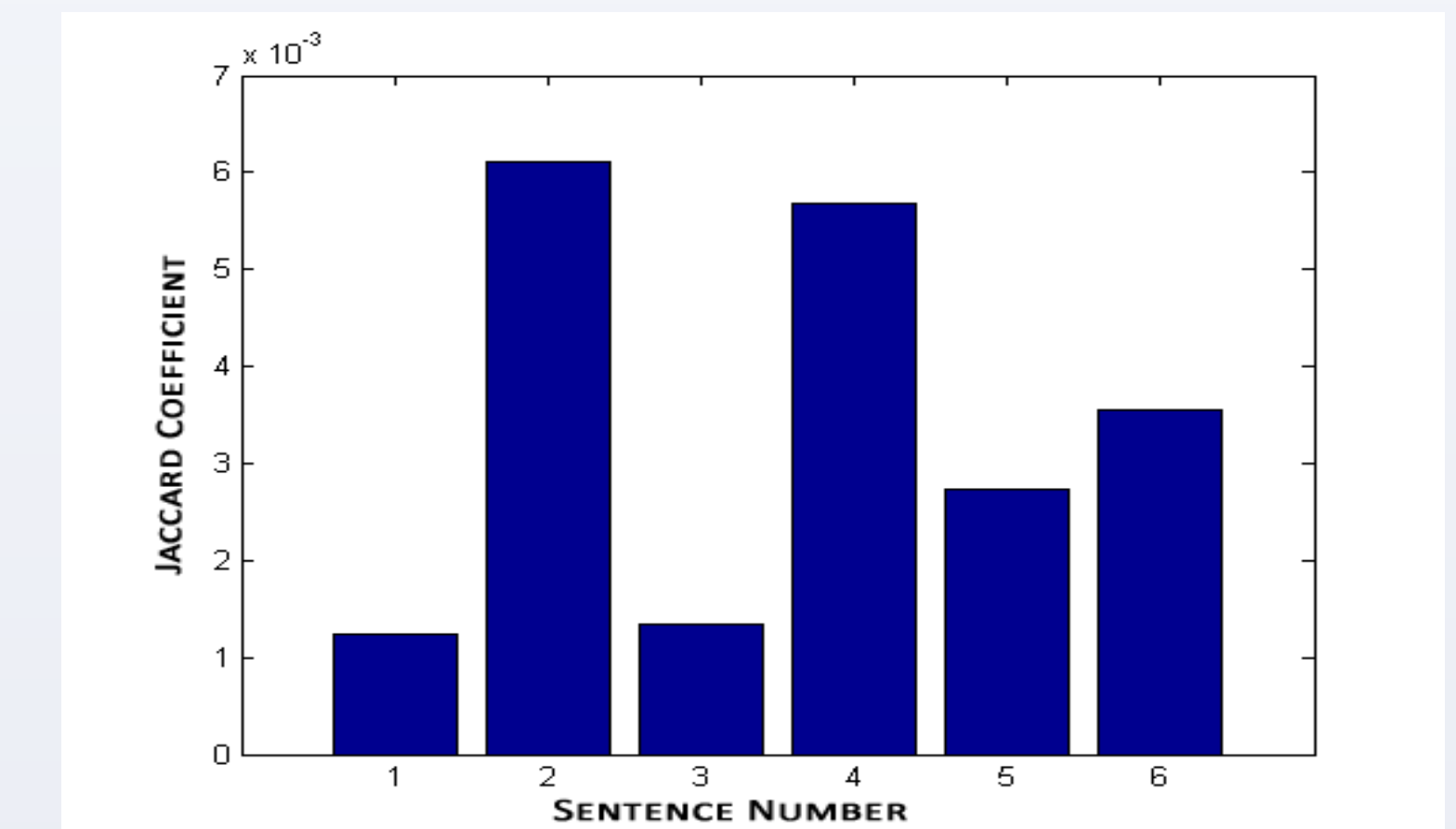
- Narrows down the pool of questions
- Allows to pick out the questions having the highest degree of abstraction

Addition of question phrases :

- "What" Questions
- Transform Do and Does questions to "why" questions
- Addition of "why" questions
- Opinion based questions

Modification of subjectivity according to question tag

RESULTS



SCORE	SENTENCES
0.00354	This role is analogous to the category of people that Gladwell identifies as “Connectors,” who play an inordinate role in helping new trends begin to “tip,” or spread rapidly.
0.00567	The three “rules of epidemics” that Gladwell identifies are: the Law of the Few, the Stickiness Factor, and the Power of Context.
0.00611	In most of these scenarios, whether the event in question is the spread of syphilis in Baltimore’s mean streets or the sudden spike in the popularity of Hush Puppies sales, there is a crucial juncture, which Gladwell terms the “tipping point,” that signals a key moment of crystallization that unifies isolated events into a significant trend.

SCORE	QUESTIONS GENERATED BY US
0.2411	Why do connectors play an inordinate role in helping new trends ?
0.125	What are the three rules of epidemics that Gladwell identifies : ? Discuss in detail.
0.125	What are : the Law of the Few , the Stickiness Factor, and the Power of Context ? Discuss in detail.
0.125	Are the three rules of epidemics that Gladwell identifies : the Law of the Few , the Stickiness Factor, and the Power of Context ?
0.1116	Gladwell asserts that most trends , styles , and phenomena are born and spread according to routes of transmission and conveyance that are strikingly similar , Do you agree or disagree with the author ?
0.1071	What is analogous to the category of people that Gladwell identifies as Connectors begin to tip , or spread rapidly ? Discuss in detail.
0.1041	How many rules of epidemics that Gladwell identifies are : the Law of the Few , the Stickiness Factor , and the Power of Context ?
0.0312	Who play an inordinate role in helping new trends ? Also explain Why?

FUTURE WORK

- To train the model for sentence extraction using a larger and more varied dataset.
- To conduct a user-study in which actual students and teachers test the system and mark those system-generated questions that proved useful