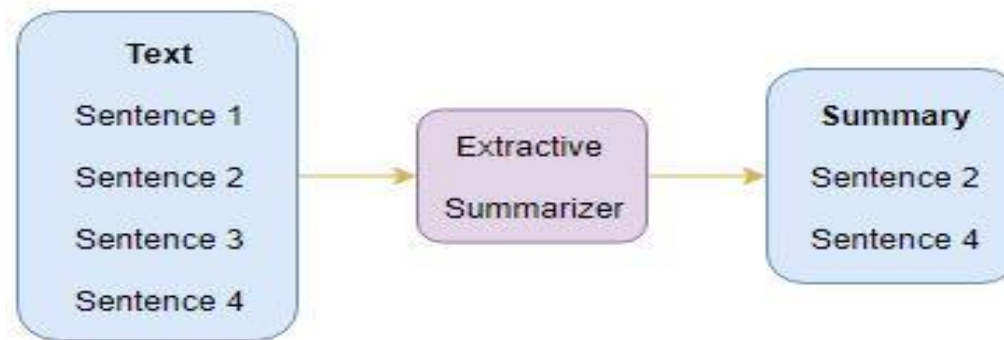


Text Summarization

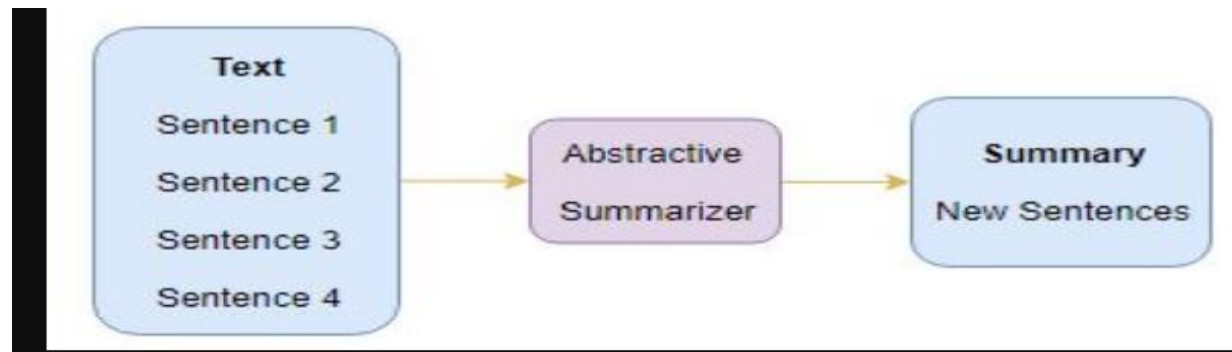
- Summarization means to reduce the size of the document without changing its meaning.
- Automatic text summarization is used to create a short, accurate and fluent summary from a longer document.
- Summarization methods are greatly needed to consume the ever-growing amount of text data available online.
- Applying text summarization reduces reading time, accelerates the process of researching for information and increases the amount of information that can fit in an area.
- Types of Summarization:
 - 1. Extraction Based Summarization
 - 2. Abstraction Based Summarization

- Extraction Based Summarization
- **It identify the important sentences or phrases from the original text and extract only those from the text.** Those extracted sentences would be our summary. The below diagram illustrates extractive summarization



- Example:
- Source Text: Joseph and Mary came to attend the annual event in Delhi.
- In the city, Mary gave birth to a child named Jessus.
- Extractive Summary: Joseph and Mary attend event Delhi. Mary birth Jessus.

- Abstraction- Based Summarization:
- Here, we generate new sentences from the original text. This is in contrast to the extractive approach we saw earlier where we used only the sentences that were present. The sentences generated through abstractive summarization might not be present in the original text



- Abstractive Summary: Joseph and Mary came to Delhi where Jessus was born.

- Abstractive text summarization for Marathi Documents:
- The idea is to summarize an input Marathi document by creating semantic graph called rich semantic graph for the original document, reducing the generated semantic graph, and further generating the final abstract summary.
- This approach consists of the following phases:
 - A. Marathi text document as input.
 - B. Rich Semantic Graph (RSG) creation phase:
 - In this phase, analysis of the input text document is done, finds the sentence and produces tokens for the complete document.
 - For every word it creates POS tags and detects the words into predefined categories such as person's name, location and organization.
 - After this it generates the graph for every sentence and concatenates rich semantic sub-graphs.
 - At last the sub graphs are mixed together to show the complete document correctly.

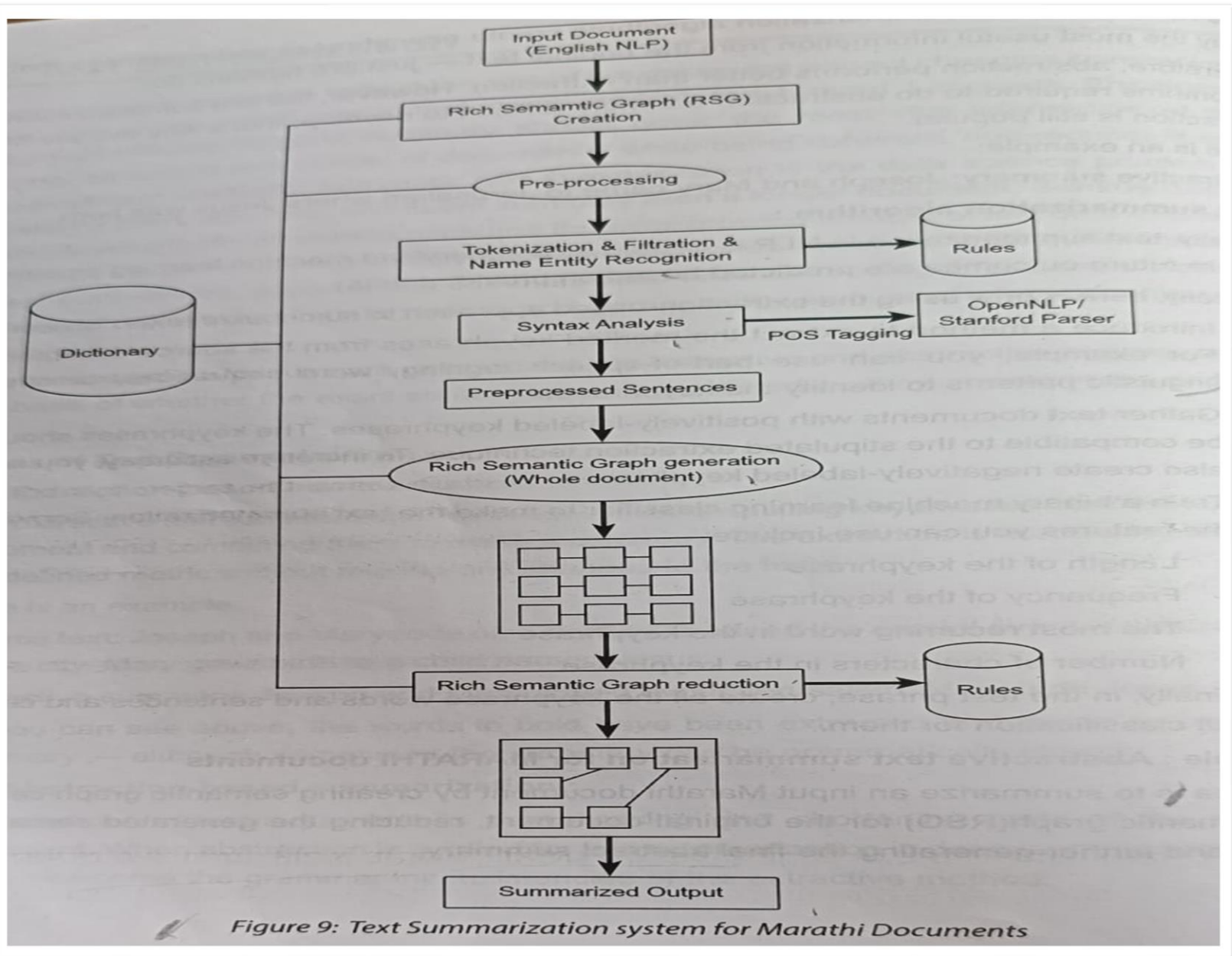


Figure 9: Text Summarization system for Marathi Documents

- C) Rich Semantic Graph Reduction Phase:
 - This phase targets to reduce the obtained rich semantic graph of the source document to more reduced graph.
 - Here set of rules are applied on the obtained rich semantic graph to reduce it by merging, consolidating or deleting the graph nodes.
- D) Summary Generation from reduced RSG:
 - This phase targets to obtain the abstractive text summary from reduced rich semantic graph.
 - To reach the target, this phase accesses the domain ontology: it has the data required in the same domain of RSG to obtain the final output.
 - The Word Net ontology is used to obtain multiple texts according to the word synonyms.
 - The obtained multiple texts are accessed and ranked, the most ranked text is considered.

- Example:
- Input: Single text documents in Marathi Language.
- मुरलीधर देवीदास आमटे उर्फ बाबा आमटे हे एक थोर मराठी समाजसेवक होते. बाबा आमटेचा जन्म डिसेंबर २३ १९१४ रोजी महाराष्ट्रतील वर्धा जिल्ह्यात झाला. बाबा आमटेना भारत सरकार कडून १९७१ मध्ये पद्मश्री पुरस्कार प्राप्त झाला. तसेच बाबा आमटेना भारत सरकार कडून १९८६ मध्ये पद्मविभूषण पुरस्कार प्राप्त झाला.
- Output: Reduced Meaningful Summary:
- मुरलीधर देवीदास आमटे उर्फ बाबा आमटे हे एक थोर मराठी समाजसेवक होते. त्यांचा जन्म डिसेंबर २३ १९१४ रोजी महाराष्ट्रतील वर्धा जिल्ह्यात झाला. त्यांना भारत सरकार कडून १९७१ मध्ये पद्मश्री पुरस्कार आणि १९८६ मध्ये पद्मविभूषण पुरस्कार प्राप्त झाला.