American Journal of Engineering Research (AJER)
e-ISSN: 2320-0847 p-ISSN : 2320-0936
Volume-6, Issue-8, pp-253-260
www.ajer.org
Research Paper Open Access

Study of Abstractive Text Summarization Techniques

Sabina Yeasmin[1], Priyanka Basak Tumpa[2], Adiba Mahjabin Nitu[3], Md. Palash Uddin[4], Emran Ali[5] and Masud Ibn Afjal[6]

[3,4,5,6](Department of Computer Science and Engineering, Hajee Mohammad Danesh Science and Technology University, Dinajpur-5200, Bangladesh)

[1,2](B.Sc. in Computer Science and Engineering, Hajee Mohammad Danesh Science and Technology University, Dinajpur-5200, Bangladesh)

ABSTRACT:*Nowadays, people use the internet to find information through information retrieval tools such as Google, Yahoo, Bing and so on. Because of the increasing rate of data, people need to get meaningful information. So, it is not possible for users to read each document in order to find the useful one. Among all the modern technologies, text summarization has become an important and timely tool for the users to quickly understand the large volume of information. Automatic text summarization system, one of the special data mining applications that helps this task by providing a quick summary of the information contained in the documents. Text summarization approach is broadly classified into two categories: extractive and abstractive. Many techniques on abstractive text summarization have been developed for the languages like English, Arabic, Hindi etc. But there is no remarkable abstractive method for Bengali text because individual word of every sentence accesses domain ontology & wordnet and it must require the complete knowledge about each Bengali word, which is lengthy process for summarization. It has thus motivated the authors to observe, analyze and compare the existing techniques so that abstractive summarization technique for Bengali texts can be proposed. To do so, the authors have conducted a survey on abstractive text summarization techniques on various languages in this paper. Finally, a comparative scenario on the discussed single or multi-document summarization techniques has been presented.*

Keywords: *abstractive summarization, big data, data mining, text mining, text summarization*

Date of Submission: 10 -06-2017 Date of acceptance: 24-08-2017

I. INTRODUCTION

At present, rapid growth of bulky amount of data is required to process, store, and manage efficiently. Sometimes, it is so tough to find the correct information from huge amount of stored data or big data repositories. Nowadays, in the period of big data, rapid growing of textual data is available in many different languages. Big data has the potentiality to be mined for information and thus data mining is essential to find out the proper information what are needed. Search engines such as Google, AltaVista, Yahoo, etc., have been developed to retrieve specific information from this huge amount of data. But most of the times, the outcome of search engine is unable to provide expected result as the quantity of information is increasing enormously day by day and also the findings are abundant. As a field of data mining, text summarization is one of the most popular research areas which helps the reader to get a quick overview of an entire document. This process provides an abstract or a summary by selecting significant portion of the information from one or more texts and consequently reduces reader's time for finding the key information in the document(s) [1].

1.1 Data Mining

Today huge amount of information is available online. But reading all of the documents is very annoying and time elapsing process. Finding document from online resources, sometimes reader may skip the important information. To read all of the retrieved documents leads to wastage of time. When data is being accessed from such a huge repository of e-documents, hundreds and thousands documents are retrieved through data mining. Thus, data mining is the process of discovering potentially useful, interesting, and previously unknown patterns from a large collection of data. It is used to extract meaningful information and to develop significant relationships among variables stored in large data set or data warehouse [2]. Analysis of data from various perspectives and summarizing it into useful information is the main purpose of data mining [2].

1.2 Data Mining Algorithms and Applications

Data mining uses different techniques such as statistical, mathematical, artificial intelligence and machine learning as the computing techniques [4]. The algorithms for data mining are Naive Bayes decision theory, support vector machine (SVM), decision tree etc. for classification or logistic regression; multiple regression, SVM etc. for regression; minimum description length for attribute importance; one-class SVM for anomaly detection; orthogonal partitioning clustering, expectation maximization(EM) algorithm, K-means algorithm, enhanced K-means etc. for clustering; Apriority for Association; singular vector decomposition (SVD), principal components analysis (PCA), non-negative matrix factorization etc. for feature selection and extraction and so on [8]. As the importance of data analysis continues to grow, the companies are finding more and more applications for data mining and business intelligence. There are a number of commercial data mining systems available today and yet there are many challenges in this field. The applications include financial data analysis, retail industry, telecommunication industry, biological data analysis and other scientific applications such as data warehouses and data preprocessing, graph-based mining, visualization and domain specific knowledge etc. [16].

1.3Comparative Statement of Data Mining

The following table presents the comparative statement of various data mining trends from past to the future [3].

Table 1: Data mining trends comparative statement

	Algorithms/ Techniques employed	**Data formats**	**Computing Resources**	**Prime areas of applications**
Past	machine learning and statistical methods	structured and numerical data stored in conventional databases	evolution of 4thGeneration programming language and various related techniques	business
Current	machine learning, statistical, artificial intelligence, pattern reorganization techniques	heterogeneous data formats contains structured, semi structured and unstructured data	high speed networks, high end storage devices and parallel, distributed computing etc.	business, web, medical diagnosis etc.
Future	soft computing systems like fuzzy logic, neural networks and genetic programming	complex data objects i.e., high dimensional, high speed data streams, sequence, noise in the time series graph, multi instance objects, multi represented objects, temporal data	multi-agent technologies and cloud computing	business, web, medical diagnosis, scientific and research analysis fields (e.g., remote sensing), social networking etc.

1.4Text Summarization

Text Summarization plays an important role in the area of text mining and natural language processing. Text summarization aims to compress the source text into a shorter and concise form with preserving its information content and overall meaning [6]. Essentially, text summarization techniques are classified as extractive and abstractive. Extractive techniques perform text summarization by selecting sentences of documents according to some criteria. Abstractive techniques attempt to improve the coherence among sentences by eliminating redundancies and clarifying the contest of sentences. Both techniques are used for summarizing text either for single document or multi-document. Sentence scoring is the most used technique for extractive text summarization. So, extractive summarization involves assigning saliency measure to some units (e.g. sentences, paragraphs) of the documents and extracting those with highest scores to include in the summary. Abstractive summarization usually needs information fusion, sentence compression and reformulation. Abstractive summarization is complex because it requires deeper analysis of source document(s) and concept-to-text generation [5].

1.4.1 Extractive Summarization Techniques

Extractive summarizer [15] finds out the most relevant sentences in the document. It also avoids the redundant data. It is easier than abstractive summarizer to bring out the summary. Extractive summarization uses the following methods to summarize document(s).Term Frequency-Inverse Document Frequency (TF- IDF) method, Cluster based method, graph theoretic approach, machine learning approach etc. are the example of extractive summarization techniques.

1.4.2Abstractive Summarization Techniques

Abstractive summarization is classified into two categories: structured base and semantic based methods [4].**Structure based Abstractive Summarization Methods**: It includes several methods such as rule based method, tree based method, Ontology method, lead and body phrase method, graph based method etc.

Semantic based Abstractive Summarization Methods: It includes the several methods such as multimodal semantic model, information item based method, semantic graph based method, semantic text representation model etc. The basic form subject-verb-object of the sentences is considered for abstractive summarization method. Some steps of this method are given below:

Preprocessing: It is considered for each sentence to create a semantic graph. The actions performed in this step are sentence position, word position, stop word removal and stemming.

Sentence position: It is concerned with counting the words in the given sentences. In Bengali, sentence is segmented by identifying the boundary of sentence which ends with a দাড়ি (।).

Word position: It splits the sentence into words by identifying হাইফেন, কমা, সেমিকোলন etc. between the words.

The stop words: These are function words like এবং, অথবা, কিন্তু, অন্যথায়,কিংবা,মাত্র,ও,আর, অধিকন্তু etc.

Stemming – A word can be found in different forms in the same document. These words need to be converted to their original form for simplicity. The stemming algorithm is used to transform words to their canonical forms, like বাংলাদেশ, বাংলাদেশের, বাংলাদেশেও, etc. should be converted to their original form বাংলাদেশ.

A preprocessed sentence is generated after preprocessing. Domain ontology & wordnet will be worked as a database for each Bengali sentence which analyzes the selected sentence. It maintains the structure of the sentence, synonym & antonym of each word, available word, priority of each sentence, weight of the sentence, relationship of the sentence etc. After generating semantic graph from the original document, a reduced semantic graph is generated accessing Domain ontology &Wordnet. Domain ontology contains the information needed in the same domain of semantic graph to generate the final texts and Wordnet is accessed to generate multiple texts according to the synonym of the word.

II. REVIEWEDARTICLES

The previous works on single document or multi-document summarization have used different directions to show the best result. Various generic multi-document or single document abstractive based summarization techniques are already present. Most of them are on English or other languages but we have not found any work for Bengali language. In this section, we discuss some works on abstractive text summarization.

2.1 Paper I

A. Khan and N. Salim [8] presented a review on abstractive summarization methods to identify key sentences for summary which contains several methods proposed by different researchers. The following table illustrates some of the methods for abstractive text summarization:

Table 2: A Comparative study on abstractive summarization methods

Author/Year	Techniques/Methods	Text Representation	Content Selection	Summary Generation
Barzilay and McKeown, 1999	tree based	dependency based representation: dsynt tree	theme intersection algorithm	fuf/surge language generator
Harabagiu and Lacatusu ,2002	template based	template/frame having slots and fillers	linguistic patterns or extraction rules	ie based md summarization algorithm
Lee and Jian, 2005	ontology based	fuzzy ontology	classifier	news agent
Barzilay and McKeown, 2005	tree based	dependency tree	algorithm uses local alignment across pair of parsed sentences	algorithm for reusing and altering phrases from input sentences
Tanaka and Kinoshita, 2009	lead and body phrase	lead, body and supplement structure	revision candidates (maximum phrases of same head in lead and body sentences)	insertion and substitution operations on phrases

2.2 Paper II

J. Mohan, Sunitha, A. Ganesha, Jaya [9] presented a study on ontology based abstractive summarization. Ontology is a formal and explicit specification of a shared conceptualization. Generally, ontologies are defined for particular domain. Several mostly used ontology based methods for abstractive text summarization are illustrated in the following table:

Table 3: Ontology based abstractive summarization

Method	Goal	Description
OntoMetric	helps to choose the appropriate ontology for a new project	compares the importance of the project objectives and study the characteristics of the ontologies. gets for every candidate ontology a quantitative measure of its suitability.
Natural Language Application metrics	helps evaluate the content of ontologies with respect to various metrics: 1. precision and recall metrics 2. cost-based evaluation metrics 3. tennis measure 4. lexical comparison level measure	measures for each ontology (a) how many identified items are correct and (b) how many items that would have been identified are effectively identified. characterizes the performance in terms of the cost of errors or the value of correct things. gives a measure of the "fit" between an ontology and a corpus (domain knowledge) by using a vector space model of instances (terms).compares the contents of two ontologies without considering their conceptual structure
OntoClean	helps evaluate a formal ontology	cleans the taxonomical structure of ontologies compares the ontology vs. a predefined ideal taxonomical structure to detect inconsistencies
EvaLexon	helps evaluate ontologies created by ontology. the method stays at the linguistic level.	compares the vocabulary of the triples mined with the input text as such and with a set of words considered to be relevant for that text.

2.3 Paper III

I. F. Moawad, and M. Aref [10] presented a paper of semantic graph reduction approach for abstractive text summarization. This paper proposed an approach to summarize an input single document by creating a semantic graph called Rich Semantic Graph (RSG).The approach consists of three phases:

The Rich Semantic Graph Creation Phase

This phase starts with deep syntactic analysis of the input text, then generates typed dependency relations (grammatical relations), and syntactic and morphological tags for each word. After that, for each sentence, the model accesses the domain ontology to instantiate, interconnect and validate the sentence concepts. RSGs represent various semantic representations of the entire document, where the maximum ranked RSG is considered.

Preprocessing module: It consists of four main processes named entity recognition, morphological and syntactic analysis, cross-reference resolution, and pronominal resolution processes.

Rich Semantic Sub-graphs Generation module: Each preprocessed sentence is composed of a sequence of words: $S_i = [W_{i1}, W_{i2}, \ldots W_{in}]$, Each word is represented as a triple sequence $W_{ij} = [S_t, T, D]$, where S_t represents the word stem, T represents the set of tags and D represents the set of typed dependency relations.

Rich Semantic Graph Generation module: Finally, it generates the final rich semantic graphs from the highest-ranked rich semantic sub-graphs of the document sentences. The semantic sub-graphs of the input document will be merged to form the final rich semantic graph.

The Rich Semantic Graph Reduction Phase

In this phase, a set of heuristic rules are applied on the generated rich semantic graph to reduce it by merging, deleting, or consolidating the graph nodes. The heuristic rule of each sentence is composed of three nodes: Subject Noun (SN) node, Main Verb (MV) node, and Object Noun (ON) node.

The Summarized Text Generation Phase

Finally, the Summarized Text Generation Phase aims to generate the abstractive summary from the reduced rich semantic graph. This phase accepts a semantic representation in the form of RSG and generates the summarized text. The input graph contains the information needed to generate the final text. To achieve its task, the phase accesses the domain ontology, which contains the information needed in the same domain of RSG.

2.4 Paper IV

N. Moratanch [11] presented a survey on abstractive text summarization. This paper collectively summarizes the major methodologies improved, issues found, research and future directions in text summarization. Abstractive text summarization is classified into two types: structured based approach &semantic based approach.

Structure Based Approach

Structured primarily based approach translates most vital data from the document(s) through psychological feature schemas.

Tree based method: Using theme selection, the central theme of multiple documents is identified. Using clustering algorithm, sentences are ordered.

Template based method: Template is required for representation of the topic which extracts information from multiple documents containing slots and fillers.
Ontology based method: Ontology represents online documents that are domain connected which has own information structure. Fuzzy Ontology describes the domain knowledge for solving the uncertainty reasoning problems.
Lead and body phrase method: This method is based on the operations of phrases (insertion and substitution) that have same syntactic head chunk in the lead and body sentences in order to rewrite the lead sentence.
Rule based method: This method summarizes the documents in terms of categories and a list of aspects. Information Extraction Rules are used to answer one or more aspects of a category.
Graph based method: This method compresses and merges information based on word graph method. Vertices represent the words in the document and edge represents the adjacent relationship between two words.
Semantic Based Approach
This technique specializes in identifying noun phrases and verb phrases by processing linguistic data.
Multimodal semantic model: In this method, a semantic model, which captures concepts and relationship among concepts, is built to represent the contents (text and images) of multimodal documents.
Information item based method: The information about the summary are generated from abstract representation of supply documents, instead of sentences from supply documents.
Semantic Graph Model: This method summarizes a document by creating RSG for the initial document by reducing the linguistics graph and then generating the final abstractive outline from the reduced linguistics graph.
Semantic Text Representation Model: This technique aims to analyze input text using semantics of words rather than syntax/Structure of text.

2.5 Paper V

P. Genest, and G. Lapalme [12] proposed fully abstractive approach to guided summarization. This paper shows that in the context of guided summarization full abstraction can be accomplished and describes a work in progress that relies on information extraction, statistical content selection and natural language generation.
Guided summarization
Guided summarization task is an oriented multi-document summarization task in which a category is attributed to a cluster of 10 source documents to be summarized in 100 words or less. There are several categories: Accidents and Natural Disasters, Attacks, Health and Safety, Endangered Resources and Investigations/Trials. Each category is associated with a list of aspects to address in the summary. **The aspects for the Attacks category is given below: WHAT**: what happened? **WHEN:** date, time, other temporal placement markers. **WHERE:** physical location. **PERPETRATORS:** individuals or groups responsible for the attack. **WHY:** reasons for the attack. **WHO AFFECTED:** casualties (death, injury), or individuals otherwise negatively affected. **DAMAGES:** damages caused by the attack. **COUNTER MEASURES**: countermeasures, rescue efforts, prevention efforts, and other reactions.
Fully Abstractive Approach: Information Extraction: A large number of candidates are found by the IE rules for each aspect. The three resources have helped in designing extraction rules a thesaurus to and semantically related nouns and verbs. **Content Selection:** The content selection module selects the best ones and sends them to the generation module. The basic heuristic is to select the candidate and similarly for the choice of a preposition or a verb for generation**. Generation:** It takes a sentence structure and words in the root form as input and gives a sentence with resolved agreements and sentence markers as output.

2.6 Paper VI

I. Fathy, D. Fadl and M.Aref [13] proposed rich semantic representation based approach for text generation. This paper proposed a new model to generate an English text from RSG. This model accesses the WordNet ontology to generate multiple texts according to the word synonyms. There are five phases: Text planning, Sentence planning, Surface Realization, Writing Styles Selected Essay Generation and Text Evaluation.
The Text Planning Phase: This phase includes Content Determination which decides what information should be included in the generated text.
The Sentence Planning Phase: This phase receives noun and verb objects and generates enhanced semi-paragraphs. **The sentence planning consists of four main processes: Lexicalization Process:** To select the most appropriate synonyms for each noun/verb, a weight(W) is assigned for every synonym. $W = (E + \left(1 - \frac{NR}{RT}\right) + (NGS/TG))/3) * 10$ .Where E is the existence probability of the synonym in the input rich semantic graph, NR represents the synonym WordNet rank, RT represents the total value of all synonym ranks, and NGS represents the WordNet group by similarity for synonym, TG represents the total number of groups by similarity for all synonyms. **Discourse Structuring Process:** For each noun object, a pseudo-sentence is composed for each attribute, and a pseudo sentence is composed for each related verb to that object. **Aggregation Process:** This process decides how pseudo-sentences should be combined into semi-paragraphs. **Referring Expression Process:** This Process identifies and replaces the intended referent by its appropriate pronoun.
Surface Realization Phase: This phase aims to transform the enhanced semi paragraphs into paragraphs by correcting them grammatically and adding the required punctuation.

Writing Styles Selected Essay Generation Phase: This phase gives the output of the final paragraphs according to the writing style, which is selected by the end user.
Evaluation Phase: This phase evaluates and ranks the paragraphs according to coherence between paragraph sentences, and the most frequently used paragraph word synonyms. Coherence measure generates very close results, so the most frequently used paragraph word synonyms is used as an additional evaluation factor.

2.8 Paper VII

D.Bartakke, S. D.Sawarkar, and A. Gulati [14] proposed a semantic based approach for abstractive multi-document text summarization. This paper uses the multiple documents in order to create abstractive summarization. In this paper, semantic graph is generated for every sentence. Then generated graph is reduced & heuristic rules have been used to generate abstractive summary. The approach consists of three phases: i) Rich Semantic Graph Creation Phase. ii) Rich Semantic Sub-graphs Generation Phase iii) Summarized Text Generation Phase
Rich Semantic Graph Creation Phase: The main objective of the Rich Semantic Graph Creation Phase is to represent the input documents semantically using Rich Semantic Graph (RSG).It is composed of three modules: i) Preprocessing ii) Rich Semantic Sub-graphs Generation and iii) Rich Semantic Graph Generation modules.
The Preprocessing Module: It consists of four main processes: **Named entity recognition** defines categories such as person names, organizations. **Morphological analysis** divided each word into morphemes &**syntactic analysis** parses the whole sentence and build the parse tree, and typed dependencies to express the relationships between words. **Co-reference resolution &Pronominal resolution reference** resolution processes identify co-reference named entities and resolve pronominal references in the whole input text.
Rich Semantic Sub-graphs Generation: The main objective of the Rich Semantic Sub-graphs Generation module is to generate multiple rich semantic sub-graphs for each input preprocessed sentence. This module includes three processes: **Word Senses Instantiation** this process instantiates word concepts for both noun and verb senses based on the domain ontology. **Concepts Validation** the sentence concepts instantiated are interconnected and validated to generate multiple rich semantic sub-graphs. **Semantic Sentences Ranking processes** to generate single rich semantic graph, the process considers the first ranked rich semantic sub-graph.
The Rich Semantic Graph Generation module the semantic sub-graphs of the input document will be merged to form the final rich semantic graph.
The Rich Semantic Graph Reduction Phase: In this phase, a set of heuristic rules are applied on the generated rich semantic graph to reduce it by merging, deleting, or consolidating the graph nodes.
The Text Generation Phase: This phase aims to generate the abstractive summary from the reduced Rich Semantic Graph (RSG).**There are four modules namely**: **Text Planning**: It decides what information should be included in the generated text. **Sentence Planning:** The sentence planning consists of four main processes: **Lexicalization Process:** Synonyms are selected for each verb/noun object by accessing the Word Net ontology. **Discourse Structuring Process:** Containing the selected object synonyms in the form of pseudo-sentences. **Aggregation Process:** Deciding how pseudo-sentences should be combined into semi paragraphs. **Referring Expression Process:** This process identifies appropriate pronoun.
The Surface Realization module: Transforming the enhanced semi-paragraphs into paragraphs by correcting grammatically and adding the required punctuation.
The Evaluation module: Evaluating and ranking the paragraphs according to two factors: coherence between paragraph sentences and the synonyms of most frequently used paragraph word.

2.9 Paper VIII

M.Subramaniam, and V.Dalal [15]presented a paper of test model for Rich Semantic Graph representation for Hindi text using abstractive method. This approach is used to generate an abstractive summary automatically for the Hindi input text document using a new semantic graph called Rich Semantic Graph (RSG).This approach consists of three phases: The Rich Semantic Graph Creation Phase. The Rich Semantic Subgraph Reduction Phase and Summarized Text Generation Phase. **Rich Semantic Graph Creation Phase**: It is classified as three modules: **Preprocessing Module:** The preprocessing involves preparing text document for the analysis the text document based on **Sentence Position Model:** Identifying the topic of the entire text document ,sentence position information is very important .In Hindi, sentence is segmented by identifying the boundary of sentence which ends with a puramviram (।).**Word Position Model:** It divided the sentence into words by identifying spaces, commas, special symbols between the words. **Stop Word Removals:** Stop words are words which are filtered out before or after processing of natural language data .Some of the stop words such as “a”, “an”, “any”, “all”, “about”, etc. **Stemming:** Stemming is the process of removing prefixes and suffixes from words .Such as the word “computation” might be stemmed to “comput”.
Rich Semantic Sub Graph Generation: This phase aims to reduce the generated rich semantic graph of the original document to more reduced graph. In this phase, a set of heuristic rules are applied on the generated rich semantic graph to reduce the graph nodes.
The Summarized Text Generation Phase: This phase accepts a semantic representation in the form of RSG and generates the summarized text.

III. RESULT AND COMPARATIVE DISCUSSION

The comparative result of the reviewed papers is illustrated in the following table:

Table 4: Comparative study of the discussed articles

Paper no.	Language	Document type	Summarizing methods
Paper I [8]	English	Single or Multiple	Term Frequency, Cue Method, Title/Headline word, Sentence length, Similarity, Proper noun
Paper II [9]	English	Single or Multiple	Onto Metric, Natural Language Application metrics ,Onto Clean, Eva exon
Paper III [10]	English	Single	The Rich Semantic Graph Creation Phase, The Rich Semantic Graph Reduction Phase, The Summarized Text Generation Phase.
Paper IV [11]	English	Single or Multiple	Structured based (tree, template, ontology, lead & body, rule, graph).Semantic based (multimodal, information item, semantic graph, semantic text representation).
Paper V [12]	English	Single or Multiple	Categories and Aspects, Extraction rules and Generation patterns
Paper VI [13]	English	Single or Multiple	Text Planning, Sentence Planning(Lexicalization, Discourse Structuring, Aggregation, Referring Expression), Surface Realization, Writing Style Selected Essay Generation
Paper VII [14]	English	Multiple	The Rich Semantic Graph Creation Phase(Preprocessing(Named entity recognition, morphological and syntactic analysis, cross-reference resolution and pronominal resolution processes), Rich Semantic Sub-graphs Generation(Word Senses Instantiation, Concepts Validation and Semantic Sentences Ranking processes) and Rich Semantic Graph Generation modules.), The Rich Semantic Graph Reduction Phase, The Summarized Text Generation Phase(Text planning, Sentence Planning, The Surface Realization and The Evaluation modules).
Paper VIII [15]	Hindi	Single	Rich Semantic Graph Creation Phase((Preprocessing(Sentence Position, Word Position, Stop Word Removals,), Rich Semantic Sub-graphs Generation and Rich Semantic Graph Generation modules), The Rich Semantic Graph Reduction Phase, The Summarized Text Generation Phase.

IV. RECOMMENDATION FOR BENGALI ABSTRACTIVE SUMMARIZATION

Analyzing the discussed articles for abstractive text summarization, it has been observed that semantic graph based method may be an emerging work for Bengali text summarization over rule based and ontology based method. Prior to develop such kind of summarizer for Bengali text, domain ontology and wordnet must be constructed for the language. The work flow of such kind of summarizer may be illustrated in the following figure.

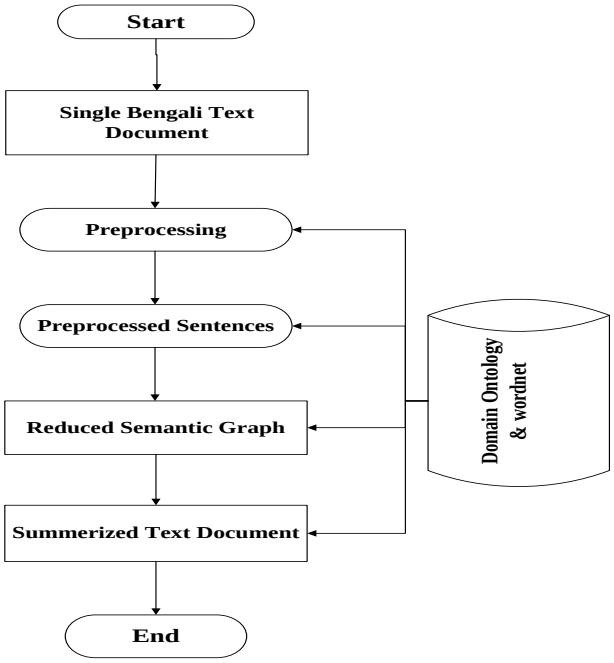

Figure 1:A proposal to Bengali abstractive text summarizer

V. CONCLUSION

Text summarization has become a burning issue in natural language processing and data mining applications. Extractive summarization techniques are now available for various natural languages including Bengali. However, there is no significant work for Bengali abstractive text summarization. In this paper, several works to summarize text documents using different abstractive data mining approaches for various languages are observed and analyzed. Then, a comparative statement of the discussed papers is made with a view to proposing a new abstractive summarization technique for Bengali language subsequently based on the recommendations discussed in section IV.

REFERENCES

[1]. K. Sarkar, Bengali text summarization by sentence extraction**,** Proceedings of International Conference on Business and Information Management (ICBIM), 233-245, India, 2012.

[2]. R. Pradheepa, K. Pavitha, A Survey on Overview of Data Mining, International Journal of Advanced Research in Computer and Communication Engineering, 5(8), India.

[3]. L.C. Reddy and Venkatadri. M, A Review on Data mining from Past to the Future, International Journal of Computer Applications, 15(7), 2012, India.

[4]. Unnamed, Data mining techniques and algorithm. [Online]. Available http://www.oracle.com/technetwork/database/enterprise-edition/odm-techniques-algorithms-097163.html

[5]. F. El-Ghannam, and T. El-Shishtawy, Multi-Topic Multi-Document Summarizer, International Journal of Computer Science & Information Technology (IJCSIT),5(6), 2013,India.

[6]. H. Dave, and S. Jaswal, Multiple Text Document Summarization System using Hybrid Summarization Technique, 1st International Conference on Next Generation Computing Technologies (NGCT-2015) ,804-808, 2015, India.

[7]. E. Turban and J. E. Aronson, Decision Support and Business Intelligent System, 6th edition, Prentice Hall, Upper Saddle River, NJ, 2001.

[8]. A. Khan, and N. Salim, A Review On Abstractive Summarization Methods,Journal of Theoretical and Applied Information Technology, 59(1), 64-71, Malaysia, 2014.

[9]. J. Mohan, M. Sunitha, A. Ganesha, Jaya , A Study on Ontology based Abstractive Summarization, Fourth International Conference Fourth International Conference on Recent Trends in Computer Science & Engineering,32-37,India,2016.

[10]. I. F. Moawad, M. Aref, Semantic Graph Reduction Approach for Abstractive Text Summarization, IEEE International Conference On Natural, 132-138, Egypt, 2012.

[11]. N. Moratanch, and. S. Chitrakala , A Survey Of Abstractive Text Summarization, IEEE International Conference on Circuit, Power and Computing Technologies, India, 2016.

[12]. P. Genest, and G. Lapalme, Fully Abstractive Approach to Guided Summarization,Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, 354–358, Canada, 2012.

[13]. I. Fathy, D. Fadl, and M. Aref , Rich Semantic Representation based Approach for Text Generation,The 8th International Conference on INFOrmatics and Systems (INFOS), 19-27, Egypt, 2012.

[14]. D. Bartakke, S. D. Sawarkar, and A. Gulati, A Semantic Based Approach for Abstractive multi-Document Text Summarization, International Journal of Innovative Research in Computer and Communication Engineering, 4(7), India, 2016.

[15]. M. Subramaniam, and V. Dalal, Test Model for Rich Semantic Graph Representation for Hindi Text using Abstractive Method, International Research Journal of Engineering and Technology (IRJET) , 2(2), India, 2015.

[16]. Unnamed, Data Mining applications & Trends, [Online]. Available: https://www.tutorialspoint.com/data_mining/dm_applications_trends.htm

Sabina Yeasmin " Study of Abstractive Text Summarization Techniques." American Journal of Engineering Research (AJER) 6.8 (2017): 253-260.