Content based Sentence Ordering using Spanning Tree Algorithm for Improved Multi Document Summarization

Ansamma John
Asst. Professor
Department of Computer Science & Engg
T.K.M College of Engg, Kollam

Dr M Wilscy
Head of the Department
Department of Computer Science
University of Kerala, Kariavattom,

ABSTRACT
Due to the availability of required information in the web, as multiple documents, the need for summarizing these multiple documents and ordering of the sentences in the summary in an efficient way become a relevant task in data mining. We present a novel sentence ordering method based on maximum cost spanning tree algorithm to improve the readability and cohesion of the summary obtained by extraction method from related multiple documents. It is based on extracting candidate sentences for the summary from multiple documents by ranking the sentences using cosine similarity measure and reducing the redundancy in the summary by Maximal Marginal Relevance (MMR) technique. Sentences in the summary are organized by constructing a graph where each sentence represents nodes of graph and edges are maintained between every pair of vertices which represents the similarity between the sentences. Most important task of our work is to find the first sentence to be placed in the ordered summary, by identifying the sentence which has minimum similarity with the sentences in the extracted summary. Ordering of remaining sentences in the summary is fixed one by one using Prim's Maximum Cost Spanning tree algorithm. The proposed algorithm is tested with DUC 2002 data set and found that summary generated after ordering has better readability and cohesion than that generated without ordering. It is noted that results are more impressive as the summary size increases.

General Terms
Data Mining

Keywords
Cosine similarity, Maximal Marginal Relevance, Spanning Tree, Prim's algorithm

1. INTRODUCTION
Availability of online information and use of mobile devices increases very rapidly day by day. Mechanisms to retrieve and interpret the relevant information in a limited time and to represent it in a precise form have become an inevitable task [1]. Automatic text summarization is a process of generating a compressed and meaningful form of source document or documents by taking salient information from source document or documents according to users need. Two categories of document summarization exist extractive and abstractive. Extractive summaries take relevant sentences from the source documents as it is. Abstractive summaries contain words and phrases which are not available in the source documents and it is similar to the human generated summaries. Summaries originate mainly from two categories of text sources, a single document or from a document set [2]. Building summaries for the single document is simple. Multi document summaries have some merits over single document summaries. Multi document summarization technique is capable of offering domain summary of a topic based on the document set or identical information in various documents or association between information in different documents or distinct information in separate documents based on the implementation. But the domain summary generated from multiple documents must satisfy the three properties. Summaries must be relevant so they should contain major concepts of all documents. Summaries should not include multiple sentences that convey the same meaning. Summaries should be bounded by size. Techniques used in single document summarization are also employed in multi document summarization. But more attention is required for multi document summarization. Degree of redundancy in the multi document summary is higher due to the existence of same information in more than one document. Order of same information in different documents may be different. So the multi document summary generated by extraction techniques may be unappealing to read due to the lack of information flow in the summary. This may lead to a very wide topic shift. Order of the sentence within the summary also depends upon the order in which the input documents are processed by the summarizer. Ordering the sentences within the summary is the most relevant and tedious task. Some researchers proposed algorithms for ordering the sentences in the summary by checking the linear ordering of sentences in the input document, by using the time stamp method and by finding shortest path method.

We present here an algorithm to order the sentences in the summary generated by extraction technique. First step of this task is to detect candidate sentences for the summary from multiple documents. Using cosine similarity measure all sentences in the input documents are ranked. Required numbers of high ranked sentences are selected to the summary, which depends upon the summary size specified by the user. Once the relevant sentences are selected, the summarizer decides the order in which these sentences are to be presented, such that whole summary maintains readability and cohesion. Main task of this work is to recognize the first

sentence in the ordered summary. This is accomplished by taking pair wise maximum similarity between the sentences in the summary. While taking this similarity we also considered the fact that each sentence in the summary has some relation with few sentences appearing before it in the paragraph of the input document where it appears. Influences of the previous sentences are to be considered while taking the pair wise similarity. Here three previous sentences are considered while calculating the pair wise similarity. From the computed maximum similarities, sentence with minimum similarity value is selected as the first sentence in the summary. Further, from first sentence a maximum cost spanning tree is constructed by selecting next sentences one by one using the similarity measure. The order in which sentences are included in the spanning tree indicates the order of sentences in the summary. Quality of ordering is very hard to evaluate because it is subjective. Sentences can be arranged in multiple ways to get the same meaning. However summary generated after ordering increases readability and provides better information flow.

Rest of the paper is organized as follows. Section 2 presents related work, section 3 describes the way how sentence are extracted and ordered. Further we discuss the need of selecting previous sentences of the sentences in the summary, while taking the pair wise similarity between the sentences in the summary. Section 4 covers result obtained and experimental part of the research.

2. RELATED WORKS

Majority ordering and chronological ordering are the algorithms developed for ordering the sentences in the summary[1]. Majority ordering relies on the original ordering of the sentences in the input documents. Input documents are decomposed into different themes and pair wise orders between the themes are recognized to obtain a linear order that maximizes the agreement between the ordering provided by the input text. For each pair of themes two counts are calculated. Linear order is identified by marinating a precedence graph, where nodes are themes and edges has a weight. This algorithm constructs the maximum cost spanning tree. This method is well suited when all input documents follows similar organization of themes. Chronological ordering algorithm takes as input a set of themes and orders the sentences from the themes based on the timestamp of the theme. This cannot provide acceptable ordering if the input documents are not event based.

A graph based ranking model is introduced by R. Mihalcea for summarization keyword extraction and other NLP applications [9]. Here the sentences for the summary are extracted by representing them as the vertex of a graph and importance of the vertex within the graph is obtained from the global information in a recursive manner. This method only extracts the sentences and no specific ordering is computed within the extracted sentences.

Jonas Sjoberg and Kenji Araki introduced extraction based method for automatic summarization by finding the shortest path algorithm. In this method N shortest paths of length two to length .This method will work if there is a path from first node to the last node and it is not considering the influence of preceding neighborhood sentences while constructing the path.

Ordering the selected sentence is crucial for readability and cohesion. A novel method for sentence ordering is proposed and it is explained in the following section.

3. PROPOSED SPANNING TREE METHOD FOR SENTENCE ORDERING

Required numbers of sentences are selected from the multiple input documents using the extraction method and ordering algorithm is applied to generate a well structured summary. Fig.1 portraits the proposed system architecture.

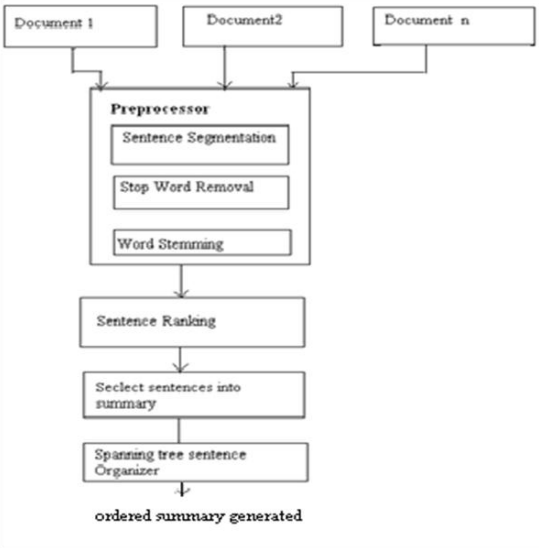

Fig. 1 system architecture of summarizer

A set of preprocessing steps are applied on the input documents before the sentences are selected for summarization.

3.1 Preprocessing

Preprocessing organizes the input documents in a structured way. Multiple documents are preprocessed using the techniques namely, sentence segmentation, removing stop words and sentence stemming. Sentence Segmentation separates each sentence from the input document using the delimiter full stop. Stop word removal removes stop words from the input documents since they do not convey any relevant meaning like “have”, “been”, “it”, etc from the documents. Stemming algorithm identifies stem or radix of each word and use the stem for further processing. This is achieved by removing the prefixes and suffixes of each word using porter stemmer algorithm [18]. This step is very much essential because the same stem word can occur in different sentences in multiple forms. Similarity between the sentences will be more accurate if the word stems are used for finding similarity.

3.2 Extraction of Relevant Sentences

Sentences in the preprocessed documents are ranked based on the relevance of the sentence with respect to the whole documents. Ranking is achieved by representing each preprocessed sentence as a vector and all input documents as another vector. Vector of a sentence or document object is generated by keeping frequencies of distinct stemmed words in the object, other than stop words. This representation is very efficient for selecting appropriate sentences for the summary by using cosine similarity measure. Cosine similarity measure is used to calculate the similarity between the vectors and it calculates the cosine of the angle between them. Similarity between the two sentence vectors X and Y is calculated using the formula as given below.

$$\text{Similarity} = \frac{\text{X.Y}}{|\text{X}|.|\text{Y}|}$$

$$X.Y = \sum_{i=1}^{n} x_i \, . \, y_i$$

$$|\text{X}|.|\text{Y}| = \sqrt{\sum_{\text{i}=1}^{\text{n}} \text{x}_\text{i}{}^2} + \sqrt{\sum_{\text{i}=1}^{\text{n}} \text{y}_\text{i}^2}$$

where x_i and y_i are the frequency of i^{th} word in X and Y vector

N
$-$ *number of non stopwords in the whole input documents*

Rank of a sentence directly indicates the relevance of it in the whole document. During the summarization process high ranked sentences are selected one by one till the required sized summary is generated. This may leads to the redundancy of information in the summary. As the relevance increases, the redundancy of information contained in the summary increases which may adversely affect the quality of summary. During the summarization process the relevance is desirable at the same time redundancy should be kept to a minimum. So for the sentence selection we followed a technique that helps to strike a balance between redundancy and relevance. Major task in the summary generation is to identify the most appropriate sentence which has maximum relevance with respect to sentences in the input documents and has minimum redundancy with the already generated sentences in the summary. Maximal Marginal Relevance (MMR) method is used for selecting most appropriate sentences for the summary [5]. For selecting most relevant and novel sentence we measure relevance and novelty independently and provide a linear combination as a metric. This linear combination is called marginal relevance. Maximal Marginal Relevance for each sentence 'S' is calculated as follows

$$\text{MMR(S)} = \lambda\, \text{Relevance(S)} - (1-\lambda)\text{Redundancy(S)}$$

where Relevance(S)

$=$ Similarity (S, Poolof sentences in theinput documents)

$$Redundancy\,(S) = Similarity\,(S, Partiallyconstructed\,Summary)$$

The λ value determines the level of redundancy and a good λ value depends on the input document set to the summarizer**.** The **λ** is a value in the interval [0, 1] the linear combination of both criteria is optimized. So we have to choose appropriate value for λ. Here value of λ is fixed as 0.5. Sentence selection using MMR is done in an iterative fashion**.** During successive iteration the MMR score of the sentences in the input document which is not included in summary is recalculated and the sentence with highest score is selected for the inclusion in the summary. This is removed from the input pool and included in the partially constructed summary.

3.3 Sentence Ordering

Proposed graph based sentence ordering system is developed based on the observation that every sentence in any type of document is related to few sentences that occur immediately before it in the paragraph where it lies. Readability and cohesion depends on the connectivity of each sentence with the previous sentences in the document. We are considering three previous sentences in the current paragraph for maintaining the ordering. Once we extracted the sentences for generating summary, we have to order them to increases readability.

Formally, let G = (V, E) be a directed graph with set of vertices V and set of edges E. Each vertex in V represents a sentence in the summary. A complete graph is constructed using the sentences in the summary. Edge weight represents the similarity between sentence in the head vertex and the three sentences, preceding the sentence available at the tail. Calculate edge weight for every ordered pair of vertices. Score of a vertex Vi is defined as the maximum edge cost of the edge <Vi,Vj>, for all Vj Є V and i ≠ j. Vertex with minimum score will have minimum similarity with other and it is selected as the first sentence. Starting from the vertex that represents first sentence, use Prim's algorithm to construct a maximum cost spanning tree of the graph that spans all the vertices. The order in which vertices are removed from the priority queue to generate the spanning tree indicates the order of the corresponding sentences in the summary.

4. RESULTS AND DISCUSSIONS

DUC 2002 dataset of multiple documents written about the same topics are used for the evaluation of our system. This system is tested with all categories of data in the dataset. Different categories of data available are documents about a single natural disaster event and created within at most seven day window, documents about a single event in any domain created with at most a seven day window, document about multiple distinct event of a single type (no limit on the time

window) and documents about biological information mainly about a single individual.

Using MMR technique summary is generated by extracting sentences from different dataset. By extracting sentences from different documents may not convey meaningful information if the sentences are not organized appropriately. Results of our spanning tree based ordering system shows that ordering considerably improves the quality of the summaries in terms of readability and cohesion. There is no well defined method to evaluate the ordering of sentences in the summary since it is subjective. Ten human judges were asked to evaluate our system by comparing the unordered summary with ordered summary. Thirty two datasets of DUC 2002 that include all four categories of data are used for evaluating the system. The study was done by generating 50 word, 100 word, 200 word and 500 word summaries. All judges has the opinion that ordered summary generated by our system is better than unordered summary in terms of readability and cohesion.

Tables given below shows improvement achieved in the summary after ordering by considering different types of document set and by varying the summary size.

4.1 Summary Generated From Five Documents about a Single Natural Disaster.

Table 1.1 Summary generated with 50 words

Before ordering :

The National Hurricane Center in Miami reported its position at 2 a.m. Hurricane Gilbert, packing 110 mph winds and torrential rain, moved over this capital city today after skirting Puerto Rico, Haiti and the Dominican Republic. Hurricane Gilbert, one of the strongest storms ever, slammed into the Yucatan Peninsula Wednesday and leveled thatched homes, tore off roofs, uprooted trees and cut off the Caribbean resorts of Cancun and Cozumel.

After ordering :

Hurricane Gilbert, one of the strongest storms ever, slammed into the Yucatan Peninsula Wednesday and leveled thatched homes, tore off roofs, uprooted trees and cut off the Caribbean resorts of Cancun and Cozumel. The National Hurricane Center in Miami reported its position at 2 a.m. Hurricane Gilbert, packing 110 mph winds and torrential rain, moved over this capital city today after skirting Puerto Rico, Haiti and the Dominican Republic.

Table 1.2 Summary generated with 100 words.

Before ordering :

The National Hurricane Center in Miami reported its position at 2 a.m. Hurricane Gilbert, packing 110 mph winds and torrential rain, moved over this capital city today after skirting Puerto Rico, Haiti and the Dominican Republic. Havana Radio, meanwhile, reported Monday that 25,000 people were evacuated from coastal areas in Guantanamo Province on the nation's southeastern coast as Gilbert's winds and rain began to brush the island. Hurricane Gilbert, one of the strongest storms ever, slammed into the Yucatan Peninsula Wednesday and leveled thatched homes, tore off roofs, uprooted trees and cut off the Caribbean resorts of Cancun and Cozumel.

After ordering:

Hurricane Gilbert, one of the strongest storms ever, slammed into the Yucatan Peninsula Wednesday and leveled thatched homes, tore off roofs, uprooted trees and cut off the Caribbean resorts of Cancun and Cozumel. The National Hurricane Center in Miami reported its position at 2 a.m. Hurricane Gilbert, packing 110 mph winds and torrential rain, moved over this capital city today after skirting Puerto Rico, Haiti and the Dominican Republic. Havana Radio, meanwhile, reported Monday that 25,000 people were evacuated from coastal areas in Guantanamo Province on the nation's southeastern coast as Gilbert's winds and rain began to brush the island.

Table 1.3 Summary generated with 200 words

Before ordering :

The storm was approaching from the southeast with sustained winds of 75 mph gusting to 92 mph. The National Hurricane Center in Miami reported its position at 2 a.m. Hurricane Gilbert, packing 110 mph winds and torrential rain, moved over this capital city today after skirting Puerto Rico, Haiti and the Dominican Republic. It looks like the eye is going to move lengthwise across that island, and they're going to bear the full brunt of this powerful hurricane, Sheets said. Havana Radio, meanwhile, reported Monday that 25,000 people were evacuated from coastal areas in Guantanamo Province on the nation's southeastern coast as Gilbert's winds and rain began to brush the island. The National Weather Service warned that the Caymans could expect high waters and large waves "which may undermine buildings along the beaches. Hurricane Gilbert, one of the strongest storms ever, slammed into the Yucatan Peninsula Wednesday and leveled thatched homes, tore off roofs, uprooted trees and cut off the Caribbean resorts of Cancun and Cozumel. Despite the intensity of the onslaught and the ensuing heavy flooding, officials reported only two minor injuries. EDT, the center of the hurricane was about 100 miles south of the Dominican Republic and 425 miles east of Kingston, Jamaica.

After ordering:

Hurricane Gilbert, one of the strongest storms ever, slammed into the Yucatan Peninsula Wednesday and leveled thatched homes, tore off roofs, uprooted trees and cut off the Caribbean resorts of Cancun and Cozumel. The storm was approaching

from the southeast with sustained winds of 75 mph gusting to 92 mph. The National Hurricane Center in Miami reported its position at 2 a.m. Hurricane Gilbert, packing 110 mph winds and torrential rain, moved over this capital city today after skirting Puerto Rico, Haiti and the Dominican Republic. It looks like the eye is going to move lengthwise across that island, and they're going to bear the full brunt of this powerful hurricane, Sheets said. Havana Radio, meanwhile, reported Monday that 25,000 people were evacuated from coastal areas in Guantanamo Province on the nation's southeastern coast as Gilbert's winds and rain began to brush the island. The National Weather Service warned that the Caymans could expect high waters and large waves "which may undermine buildings along the beaches. Despite the intensity of the onslaught and the ensuing heavy flooding, officials reported only two minor injuries. EDT, the center of the hurricane was about 100 miles south of the Dominican Republic and 425 miles east of Kingston, Jamaica.

Table 1.4 Summary generated with 500 words

Before ordering:

The storm was approaching from the southeast with sustained winds of 75 mph gusting to 92 mph. The National Hurricane Center in Miami reported its position at 2 a.m. The National Weather Service in San Juan, Puerto Rico, said Gilbert was moving westward at 15 mph with a broad area of cloudiness and heavy weather rotating around the center of the storm. Strong winds associated with the Gilbert brought coastal flooding, strong southeast winds and up to 12 feet to Puerto Rico's south coast. Hurricane Gilbert, packing 110 mph winds and torrential rain, moved over this capital city today after skirting Puerto Rico, Haiti and the Dominican Republic. It looks like the eye is going to move lengthwise across that island, and they're going to bear the full brunt of this powerful hurricane, Sheets said. Forecasters say Gilbert was expected to lash Jamaica throughout the day and was on track to later strike the Cayman Islands, a small British dependency northwest of Jamaica. Prime Minister Edward Seaga of Jamaica alerted all government agencies, saying Sunday night: Hurricane Gilbert appears to be a real threat and everyone should follow the instructions and hurricane precautions issued by the Office of Disaster Preparedness in order to minimize the danger. Flights were canceled Sunday in the Dominican Republic, where civil defense director Eugenio Cabral reported some flooding in parts of the capital of Santo Domingo and power outages there and in other southern areas. "People were running around in the main lobby of our hotel (on Grand Cayman Island) like chickens with their heads cut off, " said one man. Havana Radio, meanwhile, reported Monday that 25,000 people were evacuated from coastal areas in Guantanamo Province on the nation's southeastern coast as Gilbert's winds and rain began to brush the island. " The storm also skirted Puerto Rico without causing any damage to military facilities, Ross said The National Weather Service warned that the Caymans could expect high waters and large waves "which may undermine buildings along the beaches. Hurricane Gilbert, one of the strongest storms ever, slammed into the Yucatan Peninsula Wednesday and leveled thatched homes, tore off roofs, uprooted trees and cut off the Caribbean resorts of Cancun and Cozumel. Despite the intensity of the onslaught and the ensuing heavy flooding, officials reported only two minor injuries. Bands of 25-30 youths roamed the streets of Cancun Wednesday, looting stores with shattered windows, said Alfredo Moro Sanchez, investigative coordinator of the Quintana Roo state judicial police. The National Hurricane Center said a hurricane watch was in effect on the Texas coast from Brownsville to Port Arthur and along the coast of northeast Mexico from Tampico north. Forecasters said the hurricane was gaining strength as it passed over the ocean and would dump heavy rain on the Dominican Republic and Haiti as it moved south of Hispaniola, the Caribbean island they share, and headed west. EDT, the center of the hurricane was about 100 miles south of the Dominican Republic and 425 miles east of Kingston, Jamaica. Tropical Storm Gilbert formed in the eastern Caribbean and strengthened into a hurricane Saturday night.

After ordering :

Tropical Storm Gilbert formed in the eastern Caribbean and strengthened into a hurricane Saturday night. The storm was approaching from the southeast with sustained winds of 75 mph gusting to 92 mph. The National Hurricane Center in Miami reported its position at 2 a.m. The National Weather Service in San Juan, Puerto Rico, said Gilbert was moving westward at 15 mph with a broad area of cloudiness and heavy weather rotating around the center of the storm. Strong winds associated with the Gilbert brought coastal flooding, strong southeast winds and up to 12 feet feet to Puerto Rico's south coast. Hurricane Gilbert, packing 110 mph winds and torrential rain, moved over this capital city today after skirting Puerto Rico, Haiti and the Dominican Republic. It looks like the eye is going to move lengthwise across that island, and they're going to bear the full brunt of this powerful hurricane, Sheets said. Forecasters say Gilbert was expected to lash Jamaica throughout the day and was on track to later strike the Cayman Islands, a small British dependency northwest of Jamaica. Prime Minister Edward Seaga of Jamaica alerted all government agencies, saying Sunday night: Hurricane Gilbert appears to be a real threat and everyone should follow the instructions and hurricane precautions issued by the Office of Disaster Preparedness in order to minimize the danger. Flights were canceled Sunday in the Dominican Republic, where civil defense director Eugenio Cabral reported some flooding in parts of the capital of Santoomingo and power outages there and in other southern areas. "People were running around in the main lobby of our hotel (on Grand Cayman Island) like chickens with their heads cut off, " said one man. " The storm also skirted Puerto Rico without causing any damage to military facilities, Ross said. Hurricane Gilbert, one of the strongest storms ever, slammed into the Yucatan Peninsula Wednesday and leveled thatched

homes, tore off roofs, uprooted trees and cut off the Caribbean resorts of Cancun and Cozumel. EDT, the center of the hurricane was about 100 miles south of the Dominican Republic and 425 miles east of Kingston, Jamaica. Havana Radio, meanwhile, reported Monday that 25,000 people were evacuated from coastal areas in Guantanamo Province on the nation's southeastern coast as Gilbert's winds and rain began to brush the island. The National Hurricane Center said a hurricane watch was in effect on the Texas coast from Brownsville to Port Arthur and along the coast of northeast Mexico from Tampico north. Bands of 25-30 youths roamed the streets of Cancun Wednesday, looting stores with shattered windows, said Alfredo Moro Sanchez, investigative coordinator of the Quintana Roo state judicial police. The National Weather Service warned that the Caymans could expect high waters and large waves "which may undermine buildings along the beaches. Despite the intensity of the onslaught and the ensuing heavy flooding, officials reported only two minor injuries. Forecasters said the hurricane was gaining strength as it passed over the ocean and would dump heavy rain on the Dominican Republic and Haiti as it moved south of Hispaniola, the Caribbean island they share, and headed west.

4.2 Summary of Three Documents about a Single Individual

Table 2.1 Summary generated with 50 words

Before ordering:

John Warner of Virginia, who served with Quayle on a Senate subcommittee, told an audience that whenever he heard the Indiana senator speak, ``I'd say, `there stands the next Thomas Jefferson. What Quayle thinks of those references isn't known. Republican vice presidential candidate Dan Quayle's official resume says he was the Indiana Consumer Protection Division's chief investigator a year before the unit was created, state records show.

After Ordering.

Republican vice presidential candidate Dan Quayle's official resume says he was the Indiana Consumer Protection Division's chief investigator a year before the unit was created, state records show. John Warner of Virginia, who served with Quayle on a Senate subcommittee, told an audience that whenever he heard the Indiana senator speak, ``I'd say, `there stands the next Thomas Jefferson. What Quayle thinks of those references isn't known.

Table 2.2 Summary generated with 100 words

Before ordering:

John Warner of Virginia, who served with Quayle on a Senate subcommittee, told an audience that whenever he heard the Indiana senator speak, ``I'd say, `there stands the next Thomas Jefferson. What Quayle thinks of those references isn't known. Wendell C. Phillippi _ who worked for Quayle's grandfather, Eugene C. Pulliam, as managing editor of The Indianapolis News _ said he contacted the National Guard on behalf of Quayle when the young man applied in 1969 and ``recommended him very highly," the News reported. Republican vice presidential candidate Dan Quayle's official resume says he was the Indiana Consumer Protection Division's chief investigator a year before the unit was created, state records show. Her statement was released by the Quayle campaign

After ordering :

Republican vice presidential candidate Dan Quayle's official resume says he was the Indiana Consumer Protection Division's chief investigator a year before the unit was created, state records show. John Warner of Virginia, who served with Quayle on a Senate subcommittee, told an audience that whenever he heard the Indiana senator speak, ``I'd say, `there stands the next Thomas Jefferson. What Quayle thinks of those references isn't known. Wendell C. Phillippi _ who worked for Quayle's grandfather, Eugene C. Pulliam, as managing editor of The Indianapolis News _ said he contacted the National Guard on behalf of Quayle when the young man applied in 1969 and ``recommended him very highly," the News reported. Her statement was released by the Quayle campaign.

Table 2.3 Summary generated with 200 words

Before ordering:

John Warner of Virginia, who served with Quayle on a Senate subcommittee, told an audience that whenever he heard the Indiana senator speak, ``I'd say, `there stands the next Thomas Jefferson. What Quayle thinks of those references isn't known. ``I think Dan Quayle, although he's 41 years of age, he's going to send a message to the American people that the Republican Party respects young people who achieve," said Sen. Bush, who followed his running mate to the podium, said he was ``proud to have Dan Quayle at my side. Wendell C. Phillippi _ who worked for Quayle's grandfather, Eugene C. Pulliam, as managing editor of The Indianapolis News _ said he contacted the National Guard on behalf of Quayle when the young man applied in 1969 and ``recommended him very highly," the News reported. Republican vice presidential candidate Dan Quayle's official resume says he was the Indiana Consumer Protection Division's chief investigator a year before the unit was created, state records show. Quayle was hired as an entry-level research assistant by the state Attorney General's office in 1970 after he returned from six months of basic and advanced training for the Indiana National Guard. Her statement was released by the Quayle

campaign

After ordering :

Quayle was hired as an entry-level research assistant by the state Attorney General's office in 1970 after he returned from six months of basic and advanced training for the Indiana National Guard. John Warner of Virginia, who served with Quayle on a Senate subcommittee, told an audience that whenever he heard the Indiana senator speak, ``I'd say, `there stands the next Thomas Jefferson. What Quayle thinks of those references isn't known. ``I think Dan Quayle, although he's 41 years of age, he's going to send a message to the American people that the Republican Party respects young people who achieve," said Sen. Bush, who followed his running mate to the podium, said he was ``proud to have Dan Quayle at my side. Wendell C. Phillippi _ who worked for Quayle's grandfather, Eugene C. Pulliam, as managing editor of The Indianapolis News _ said he contacted the National Guard on behalf of Quayle when the young man applied in 1969 and ``recommended him very highly," the News reported. Republican vice presidential candidate Dan Quayle's official resume says he was the Indiana Consumer Protection Division's chief investigator a year before the unit was created, state records show. Her statement was released by the Quayle campaign.

4.3 Summary of Six Documents about a Single Event in a Domain Created with at Most a Seven Day Window

Table 3.1 Summary generated with 50 words

Before ordering:

The Big Mac meal, consisting of a hamburger, soft drink and french fries costs the equivalent of $2.57, or about as much the similar meal would cost in numerous Pljeskavica joints around town. The 500-seat McDonald's restaurant in a three-story building is operated by McDonald's Restaurant Shenzhen Ltd., a wholly owned subsidiary of McDonald's Hong Kong. After that, we'd like to open five every year

After ordering :

The 500-seat McDonald's restaurant in a three-story building is operated by McDonald's Restaurant Shenzhen Ltd., a wholly owned subsidiary of McDonald's Hong Kong. The Big Mac meal, consisting of a hamburger, soft drink and french fries costs the equivalent of $2.57, or about as much the similar meal would cost in numerous Pljeskavica joints around town. After that, we'd like to open five every year

Table 3.2 Summary generated with 100 words

Before ordering:

The Big Mac meal, consisting of a hamburger, soft drink and french fries costs the equivalent of $2.57, or about as much the similar meal would cost in numerous Pljeskavica joints around town. Under the sign of the golden arches, accented by the Soviet hammer-and-sickle flag, hundreds lined up for the long-awaited grand opening at 10 a.m. The 500-seat McDonald's restaurant in a three-story building is operated by McDonald's Restaurant Shenzhen Ltd., a wholly owned subsidiary of McDonald's Hong Kong. Opening the second McDonald's restaurant in Moscow, along with a 12-storey office block, Mr George Cohon, head of McDonald's Canada, could well be described as the Russian authorities' idea of a model investor. After that, we'd like to open five every year.

After ordering:

The 500-seat McDonald's restaurant in a three-story building is operated by McDonald's Restaurant Shenzhen Ltd., a wholly owned subsidiary of McDonald's Hong Kong. The Big Mac meal, consisting of a hamburger, soft drink and french fries costs the equivalent of $2.57, or about as much the similar meal would cost in numerous Pljeskavica joints around town. Under the sign of the golden arches, accented by the Soviet hammer-and-sickle flag, hundreds lined up for the long-awaited grand opening at 10 a.m. Opening the second McDonald's restaurant in Moscow, along with a 12-storey office block, Mr George Cohon, head of McDonald's Canada, could well be described as the Russian authorities' idea of a model investor. After that, we'd like to open five every year.

Table 3.3 Summary generated with 200 words

Before ordering:

`I just wanted to taste genuine American hamburgers," said Milica Nikolic, a high school student who waited for three hours to taste her first Big Mac. The Belgrade media have suggested that the success of McDonald's in Yugoslavia depends on its acceptance by citizens long accustomed to a hamburger-like fast-food dish called the Pljeskavica: ground pork and onions on a bun. The Big Mac meal, consisting of a hamburger, soft drink and french fries costs the equivalent of $2.57, or about as much the similar meal would cost in numerous Pljeskavica joints around town. Under the sign of the golden arches, accented by the Soviet hammer-and-sickle flag, hundreds lined up for the long-awaited grand opening at 10 a.m McDonald's hopes to open a restaurant in Beijing later. The 500-seat McDonald's restaurant in a three-story building is operated by McDonald's Restaurant Shenzhen Ltd., a wholly owned subsidiary of McDonald's Hong Kong. Opening the second McDonald's restaurant in Moscow, along with a 12-storey office block, Mr George Cohon, head of

McDonald's Canada, could well be described as the Russian authorities' idea of a model investor. However, the competition is heating up, and there is "a good possibility that South Korea already has reached the fast-food saturation point for its current average income bracket," an official for the South Korea Chamber of Commerce and Industry says. After that, we'd like to open five every year.

After ordering :

However, the competition is heating up, and there is "a good possibility that South Korea already has reached the fast-food saturation point for its current average income bracket," an official for the South Korea Chamber of Commerce and Industry says. ``I just wanted to taste genuine American hamburgers," said Milica Nikolic, a high school student who waited for three hours to taste her first Big Mac. The Belgrade media have suggested that the success of McDonald's in Yugoslavia depends on its acceptance by citizens long accustomed to a hamburger-like fast-food dish called the Pljeskavica: ground pork and onions on a bun. The Big Mac meal, consisting of a hamburger, soft drink and french fries costs the equivalent of $2.57, or about as much the similar meal would cost in numerous Pljeskavica joints around town. Under the sign of the golden arches, accented by the Soviet hammer-and-sickle flag, hundreds lined up for the long-awaited grand opening at 10 a.m. McDonald's hopes to open a restaurant in Beijing later. The 500-seat McDonald's restaurant in a three-story building is operated by McDonald's Restaurant Shenzhen Ltd., a wholly owned subsidiary of McDonald's Hong Kong. Opening the second McDonald's restaurant in Moscow, along with a 12-storey office block, Mr George Cohon, head of McDonald's Canada, could well be described as the Russian authorities' idea of a model investor. After that, we'd like to open five every year.

5. CONCLUSION

Automatic multi-document summarization system is developed and spanning tree based sentence ordering algorithm is used to improve the readability and cohesion of the summary generated. MMR technique is used to extract the relevant sentences from input documents and to reduce the redundancy of information in the summary. DUC 2002 dataset is used to evaluate the ordered summary generated. The experimental results showed that the incorporation of ordering algorithm improved the quality of the summary by increasing the readability. Readability of a document depends upon the connectivity of each sentence with previous sentences. So to order the sentences in the summary we considered few previous sentences in the paragraph where it occurs in the original document. Quality of the ordering is observed by varying the size of the summary. It is observed that as the size of the summary increases readability and cohesion also increases. This summarization and ordering is not suited for document with multiple events. Work can be extended to handle these situations. More over study on the influence of previous sentences in the ordering can be extended to fix the number of previous sentences to be considered while ordering depending upon the nature of the document and also to tune this value using neural network.

6. REFERENCES

[1] Barzilay. R. Elhadad. N and Kathleen R. McKeown . 2002 Inferring Strategies for Sentence Ordering in Multidocument News summarization . In Journal of Artificial Intelligence Research , vol.17, pp.35-55, 2002 .

[2] Liang Zhou, Miruna Ticrea and Eduard Hovy, 2004 Multidocument Biography Summarization, In Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2004), Barcelona, Spain, July 25-26, 2004.

[3] Shiyan Ou, Christopher S.G , Khoo and Dion H. Goh, 2008 Design and development of a concept-based Multidocument system for research abstract, In Journal of Information Science, vol. 34, no. 3, pp. 308-326, June 2008

[4] You Oungag , Wenji Li, Qin Lu, 2009, An integrated multi-document summarization approach based on word Hierarchical Representation, In proceedings of the ACL-IJCNLP, Singapore, pp 109-112, 2009

[5] Jade Goldstein, Vibhu Mittal, Jaime G. and Carbonell, ,2000, Multi document Summarization by Sentence Extraction, In NAACL-ANLP 2000 Workshop on Automatic Summarization (Seattle, Washington). Vol. 4, pp 40-48, 2000.),

[6] R. Michalcea., 2004, Graph-based ranking algorithm for sentence extraction , applied to text summarization, In Proceedings of the 42nd Annual Meeting of Association for Computational Linguistics (ACL) Barcelona, Spain.

[7] Lovins, J.B, 1968, Development of a stemming algorithm In Mechanical Translation and Computational Linguistic, Vol 11, pp 22-31, 1968.

[8] R. Mihalcea, P. Tarau, and E.Figa, 2004, PageRank on semantic networks, with application to word sense disambiguation. In Proceedings of 20st International Conference on Computational Linguistics (COLING), Geneva, Switzerland.

[9] R. Mihalcea, P. Tarau , 2004, TextRank: Bringing order into texts In Proceedings of EMNLP , Vol 4, Issue 4, pp 404-411 Barcelona: ACL (2004)

[10] Jiaming Zhanb, Han Tong Lohb and Ying Liua., 2009, On macro and micro-level information in multiple documents and its influence on summarization. In International Journal of Information Management, Vol. 29, No. 1, pp. 57-66, February 2009.

[11] Lin Zhao, Lide Wua and Xuanjing Huang, Using query expansion in graph-based approach for query-focused multidocument summarization, In Information Processing and Management . Vol. 45, No.1, pp. 35-41,January 2009

[12] Lovins .J.B . , Development of stemming algorithm In Mechanical Translation and Computational Linguistics vol 11 pp 22-31 1968

Notes