Automatic Text Summarization of News Articles

Prof.Asha Rose Thomas[1], Prof.Teena George[2], Prof. Sreeresmi T S

[1,2,3]Assistant Professor, Dept. of Computer Science and Engineering, Adi Shankara Institute of Engg and Technology Kalady, India

---***---

Abstract *Text Summarization has forever been a locality of active interest within the domain. In recent times, even supposing many techniques have been developed for automatic text summarization, potency continues to be a priority. Given the rise in size and range of documents on the market online, associate degree economical automatic news summarizer is that the want of the hour during this paper, we have a tendency to propose a way of text summarizer that focuses on the matter of distinctive the foremost necessary parts of the text and manufacturing coherent summaries. In our methodology, we have a tendency to don't need full linguistics interpretation of the text, instead we have a tendency to produce a outline employing a model of topic progression within the text derived from lexical chains. We have a tendency to gift associate degree optimized and economical algorithmic program to come up with text outline exploitation lexical chains and exploitation the WordNet synonym finder. Further, we have a tendency to conjointly overcome the constraints of the lexical chain approach to come up with a decent outline by implementing function word resolution and by suggesting new rating techniques to leverage the structure of reports articles.*

Key Words: Extractive Text summarization, Lexical Chains, News account, language process, Anaphora Resolution.

1. INTRODUCTION

With the supply of World Wide internet in each corner of the globe lately, the quantity of knowledge on the web is growing at associate degree exponential rate. However, given the feverish schedule of individuals and also the large quantity of knowledge on the market, there's increase in want for data abstraction or account. Text account presents the user a shorter version of text with solely very important data and so helps him to know the text in shorter quantity of your time. The goal of automatic text account is to condense the documents or reports into a shorter version and preserve necessary contents.

1.1 Summarization Definition

Natural Language process community has been work the domain of account for nearly the second half century Radev et al, 2002 [3] defines outline as "text that's created from one or a lot of texts, that conveys necessary data within the original text(s), which is not than 1/2 the initial text(s) and typically considerably but that." 3 main aspects of analysis on automatic summarization are painted by this definition:

- Summaries could also be created from one document or multiple documents,

- Summaries ought to preserve necessary data,

- Summaries ought to be short

1.2 Need for Automatic Summarization

The main advantage of account lies within the incontrovertible fact that it reduces user's time in looking out the necessary details within the document. Once humans summarize a piece, they 1st browse and perceive the article or document and so capture the small print. They then use these small prints to come up with their own sentences to speak the gist of the article. Even supposing the standard of outline generated may well be wonderful, manual account could be a time overwhelming method. Hence, the requirement for automatic summarizers is kind of apparent. The foremost necessary task in extractive text account is selecting the necessary sentences that may seem within the outline. Distinctive such sentences could be a actually difficult task. Currently, automatic text account has applications in many areas like news articles, emails, analysis papers and on-line search engines to receive outline of results found.

2. Lexical Chains

Morris, Jane and Hirst 1st introduced the thought of lexical chains. In any given article, the linkage among connected words will be utilised to generate lexical chains. A lexical chain could be a logical cluster of semantically connected words that depict a thought within the document. The relation between the words will be in terms of synonyms, identities and hypernyms/hyponyms. As an example, we will place words along when:

• 2 noun instances area unit identical, and area unit employed in an equivalent sense. (The cat within the room is giant. The cat likes milk.)

• 2 noun instances needn't be identical however area unit employed in an equivalent sense (i.e., area unit synonyms). (The bike is red. My motorbike is blue.) • The senses of 2 noun instances have a hypernym/hyponym relation between them. A word could be a word with a broad which means constituting a class into that words with additional specific meanings fall. (Daniel gave Maine a flower. it's a Rose.)

• The senses of 2 noun instances area unit siblings within the hypernym/hyponym tree. (I just like the fragrance of Rose. but helianthus is far higher.)

These relations will be accustomed cluster noun instances in a very lexical chain given the condition that every noun is appointed to just one chain. The difficult task here is decisive the chain to that a selected noun are appointed since it's going to have multiple senses or contexts. Also, although there's one context for the noun usage, it would be still ambiguous to see the lexical chain. The rationale being, as an example, one lexical chain would possibly correspond to word relation of the noun whereas the opposite would possibly correspond to its equivalent word relation. Hence, to be ready to resolve such ambiguities, the nouns should be sorted in such how that it creates longest or strongest lexical chains. If a series contains many nouns concerning same which means, then we have a tendency to decision that chain as longest chain. Similarly, the lexical chain with highest score are termed as strongest chain.

Generally, a procedure for constructing lexical chains follows 3 steps:

1)   Choose a collection of eligible words like nouns, adjectives, adverbs, etc. In our case, we decide solely nouns;

2)   For every eligible word, look for AN corresponding chain reckoning on a connexion criterion among members of the chains;

3)   If it's found, insert the word within the chain and update it consequently.

An identical path was followed by Hirst and St- Onge (H&S) in their approach of report. In opening, all words within the document labelled as nouns in WordNet area unit picked up. Within the next step, their connexion is measured supported the distance between their occurrences and their association within the WordNet wordbook. 3 styles of relation area unit outlined extra- strong (between a word and its repetition), strong (between 2 words connected by a Wordnet relation) and medium-strong once the link between the synsets of the words is longer than one (only methods satisfying sure restrictions area unit accepted as valid connections).

2.1 Barzilay and Elhadad Approach

Barzilay and Elhadad[7] planned lexical chains as an intermediate step within the text report method. They planned to develop a chaining model consistent with all potential alternatives of word senses and so select the simplest one in all them.

Their approach will be illustrated victimization the subsequent example - mister. Kenny is that the person who fictional an Anesthetic machine that uses micro-computers to regulate the speed at that an anesthetic is pumped up into the blood. Such machines area unit nothing new. However his device uses 2 micro-computers to realize a lot of nearer observation of the pump leading the anesthetic into the patient. First, a node for the word "Mr" is formed [lex "Mr.", sense adult male, Mr. future candidate word is "person" it's 2 senses: "human being" (person-1) and "grammatical class of pronouns and verb forms" (person – 2) the selection of sense for "person" splits the chain world to 2 completely different interpretations as shown in Figure one.
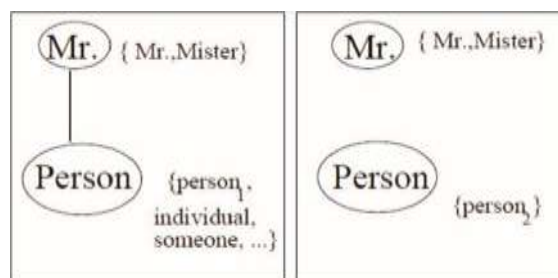
(a) Figure 1

They outline a part as a listing of interpretations that area unit exclusive of every different. Part words influence one another within the choice of their individual senses. Future candidate word "anesthetic" isn't associated with any word in the primary part, in order that they produce a replacement part for it with one interpretation.

The word "machine" has five senses - machine(1) to machine(5). In its 1st sense, "an economical person", it's associated with the senses "person" and "Mr". It thus influences the choice of their senses, so "machine" should be within the 1st part. Once its insertion the image of the primary part becomes the one shown in Figure two.
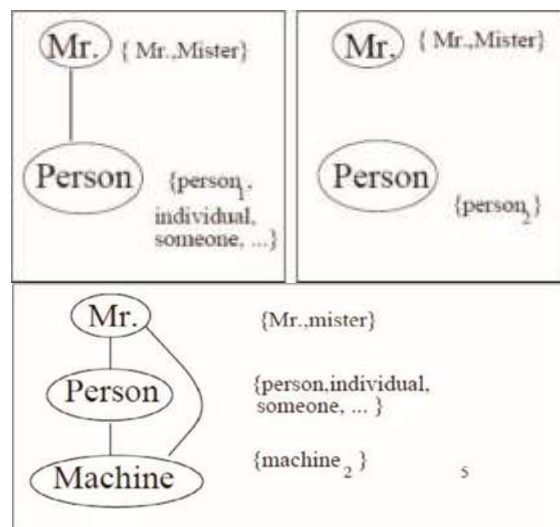

Figure 2

Under the belief that the text is cohesive, they outline the simplest interpretation because the one with the foremost connections (edges within the graph). They outline the score of an interpretation because the add of its chain scores. A series score is decided by the amount and weight of the relations between chain members. Through an experiment, they fastened the load of reduplication and equivalent word to ten, of opposite to seven, and of word and whole name to four. Their rule computes all potential interpretations, maintaining each while not self-contradiction. Once the amount of potential interpretations is larger than a definite threshold, they prune the weak interpretations i.e. interpretations having low scores consistent with this criterion, this is often to stop exponential growth of memory usage. In the end, they choose from every part the strongest interpretation.

2.2 OUR APPROACH

The lexical chain generation rule projected by Barzilay and Elhadad delineate in previous section has exponential run time that was improved by Silber and McCoy algorithm and has linear run time complexness. Hence we tend to adopted Silber and McCoy algorithm to construct the fundamental lexical chain model. Further, we've got conjointly tried to resolve the problems in each algorithms by implementing closed-class word resolution and increased sentence marking to leverage the structure of reports articles.

The following steps describe our rule for text summarisation.

1) When receiving the input, we tend to initial perform closed-class word resolution on the text.

2) In closed-class word resolution, we tend to try and notice the most effective representative noun for a closed-class word. When finding them, we tend to replace them in our sentences.

3) To exchange the pronouns, we tend to initial tokenize the passage into individual sentences.

4) Every of those sentences is more tokenized into words, and when getting the thesaurus, we tend to replace the closed-class word with the representative noun, and re-construct the sentence.

5) We tend to then notice the Part-of-Speech tag for each word to separate the nouns.

6) These nouns are then used for lexical chain construction.

7) Each lexical chain consists of closely connected nouns supported their semantic relation.

8) We tend to then score the lexical chains supported our marking criteria, and decide the sturdy chains whose score is bigger than the set threshold.

9) Exploitation the sturdy lexical chains, we will then score the individual sentences, and select those sentences to be in our outline whose score is bigger than the set threshold.

10) We tend to conjointly score the right nouns within the passage primarily based on their frequency within the passage.

11) We tend to choose a set of those correct nouns whose score is bigger than the set threshold. Later, we tend to decide the sentences that contain the primary prevalence of those correct nouns and add them to our outline.

12) Finally, the sentences are ordered consistent with their prevalence within the passage, and also the obtained sets of sentences represent the outline of the newspaper article.

2.2.1 Sentence Tokenization

We tend to take the article as input in our system and tokenize it into sentences. We tend to perform this tokenization on the idea of punctuation marks valid for locating sentence termination points as known by the foundations of English descriptive linguistics. To tokenize it into valid sentences, we tend to use the NLTK library, that is predicated on the required language of the text, that in our case is English.

2.2.2 Part of Speech Tagging for Tokenized Words

When tokenizing the article into sentences, we tend to drive our specialize in each sentence within the article to extract vital options associated with the article. We tend to more tokenize a sentence more into words. For every word, we tend to establish that POS (Part of Speech) tag it relates to a part of Speech tag helps to spot the relation of the word to at least one of the broad categories of words outlined within the English language, like Nouns, Pronouns, Verbs, etc. Its significance in our situation are going to be even later within the report. We tend to initial tokenize the given sentence to an inventory of words within the sentence. For half of Speech tagging, we tend to once more visit the NLTK library. The NLTK library maintains an outsized corpus of English words, that identifies the word and conjointly stores the a part of Speech tag it relates to. Thus, we tend to generate a brand-new list of things, wherever every item may be a tuple consisting of the word in our sentence, alongside it's a part of Speech tag.
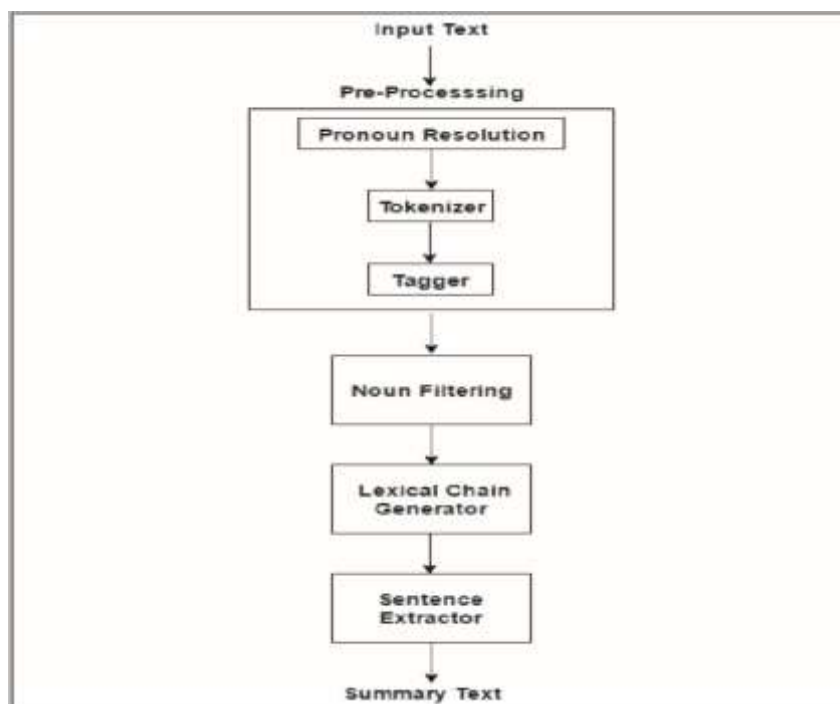

Figure 3

2.2.3 Closed-Class Word Resolution

English passages use heaps of pronouns to unceasingly refer some nouns in a piece of writing, to exchange their over usage. Thus, if we wish to spot vital nouns in a very passage, we must always resolve each closed-class word in it to relate to their various noun prevalence. Drawback of closed- class word resolution has been known as a really onerous problem as a result of it needs a grammar further as linguistics understanding of the passage. There exist varied algorithms for an equivalent, some exclusively on the grammar options, etc all. that use machine learning techniques to coach their system to spot and perceive the linguistics relations within the text. For our drawback, we tend to visit associate degree existing answer enforced by the Stanford human language technology cluster, Stanford CoreNLP, which is a suite providing varied language analytics tools, together with closed-class word resolution. We've got enforced their native library as a neighborhoods server on a machine, and perform API calls thereto. Thus, we tend to perform associate degree API decision to its native server from our program, passing our passage alongside the necessary choices to perform closed-class word resolution and that we receive associate degree output describing the relations of varied pronouns and also the noun it'd be concerning. With this info, we tend to replace the pronouns within the passage with the documented noun.

2.2.4 Lexical Chain formation

We tend to had known each word's a part of speech tag and conjointly resolved the closed-class word occurrences with the various nouns. Our next step towards summarization is to spot the most conception the passage focuses on. We tend to try and notice the most conception on the idea of the nouns within the passage. The intuition behind this can be since we tend to are concerning news articles, they contain heaps of nouns and customarily direct their specialize in a specific set of nouns, whether or not the newspaper article belongs to the class of World News, Political News, Sports News, Technology News, etc. Thus, if we tend to are ready to spot a group of nouns that kind the core of the newspaper article, extracting sentences a lot of targeted on them generates an elliptical and relevant outline. To spot the vital nouns within the passage, we tend to implement the technique of lexical chain formation, conferred by Morris, Jane and Hirst, and enforced for text summarization by Barzilay and Elhadad. Exploitation lexical chains, we tend to try and cluster along similar nouns into chains so establish sturdy chains on the idea of a marking criteria. When characteristic the sturdy chains, varied extraction techniques are often wont to extract a set of sentences from the newspaper article. In our implementation of lexical chain formation, we tend to initial try and notice all the attainable meanings or senses a noun is employed. This was achieved by exploitation WordNet. WordNet may be a massive electronic information service of English. WordNet superficially resembles a wordbook, in this it teams' words along supported their meanings. Nouns, verbs, adjectives and adverbs are sorted into sets of psychological feature synonyms (synsets), every expressing a definite conception. Synsets are interlinked by suggests that of conceptual-semantic and lexical relations. To come up with a lexical chain in our case, we tend to use the information structure wordbook, wherever every found that means can represent an inventory of these nouns within the article having this in concert of their meanings. During this means, we tend to are ready to capture each noun and their attainable senses in our lexical chain structure. Exploitation this structure, we discover the vital noun sets which can be used for sentence extraction supported our marking and sentence extraction techniques.

2.2.5 Scoring Mechanisms

The steps mentioned until currently are used for extracting important info from the newspaper article, which can facilitate America to come up with an outline from the article. During this step, we tend to discuss the varied aspects of the text which can be scored, and can be used for outline extraction.

1)   Lexical Chain Scoring: we've got fashioned the lexical chains exploitation all the nouns within the article, aside from correct nouns. The matter with correct nouns is that they often don't mean something. Thus, they cannot be accessorial to any lexical chain. One will try and assign genders to correct nouns so notice a series consequently. One methodology to seek out the gender of correct nouns might be to coach a system through world examples in order that it returns the gender that it finds the foremost probable. However, this methodology conjointly wouldn't guarantee high success rate, as multiple lexical chains may have an equivalent gender parts. We tend to had to currently puzzle out a heuristic perform which might facilitate America score the lexical chains that were fashioned higher than. For marking these chains, there are often multiple heuristics attainable. The heuristic we tend to enforced makes use of the vital criteria known by Barzilay and Elhadad, i.e. chain length, distribution within the text and also the text span lined by the text. The subsequent parameters are sensible predictors of the strength of a chain:

Length: the quantity of occurrences of the members of the chain.

Homogeneity Index:

Length –

Number of distinct occurences/length

2) Sentence Scoring:

Using the scores computed for lexical chains, now we wish to find the scores for the sentences. After our sentences are scored, we can extract a particular subset of the sentences which would have a score above a decided threshold, and would form a part of the summary.

To identify the strong chains, we use the following criteria to rank the chains.

$$Score(Chain) > Average(Scores) + 2*Standard\ Deviation(Scores)$$

3) Proper noun Scoring:

News articles contain an outsized variety of correct nouns. These correct nouns can't be accessorial to our lexical chain structure, as there is not any acceptable approach to spot the usage sense of those nouns. But correct nouns are associate degree integral a part of news articles. Thus, their prevalence can't be fully unheeded. Our basis for marking correct nouns in a piece of writing is its frequency within the article. We tend to may not notice the other considerable characteristic for them. The key reason we tend to may argue for exploitation solely the frequency was thanks to non-existence of options they might have concerning the language. Correct nouns are freelance of the language of the article and so,

We evaluated the score of a series as:

Language specific criteria doesn't exist. Our marking formula for correct nouns is

$$Score\ (Proper\ Noun) = Frequency\ (Proper\ Noun)$$

3.Outline Extraction

When marking varied aspects of our newspaper article, we tend to currently gift varied ways that we tend to enforced along to come up with a relevant outline.

3.1Extraction supported our Article Category:

The articles that our summarizer focuses on are news articles. We tend to tried to spot some relevant feature specific to a newspaper article that we tend to accessorial to our summarizer for outline generation. News articles are well- structured and arranged. The writers tend to keep up a correct flow of data in them. The primary number of sentences sometimes contain most of the relevant info on what the article would later elaborate and discuss on. Thus, these sentences ought to have a better importance over the later sentences within the article.

We tend to then parsed varied news articles from websites like BBC, Times of Bharat, the Hindu, etc to do to spot what number sentences ought to run the most relevancy. When evaluating some articles from every of those websites, we tend to came to the conclusion that solely the primary sentence ought to run a better priority over different sentences. The primary sentence inthese articles is usually long and covers the most gist of the newspaper article.

Thus, for our outline generation, the inclusion of this sentence is critical. We tend to accessorial this feature in our summarizer to essentially extract this sentence and add it to our outline.

3.2Exploitation Sentence Scoring:

Within the earlier section, we tend to delineate our marking technique wherever we tend to created a lexical chain organisation for similar nouns and used it to attain our sentences. Our threshold for choosing a sentence is that if its score is bigger than the common of the several all the sentences. These sentences have a better concentration of vital nouns compared to different sentences, and so ought to be a part of the outline. Thus, exploitation sentence marking, solely those sentences would be a part of the outline wherever

$$Score\ (Sentence) > Average\ (Sentence\ Scores)$$

1) **Exploitation strong Lexical Chains**:

We tend to delineate our methodology of sentence extraction wherever we tend to used the sentence scores to extract relevant sentences. The lexical chains compete a significant role here as they were wont to score the sentences.

To extend the importance of the sturdy lexical chains, we tend to applied another heuristic here delineate by Barzilay and Elhadad in his text summarisation techniques.

For every chain within the outline illustration, select the primary sentence that contains the primary look of a representative chain member within the text. We tend to enforce this system for text extraction on our sturdy chains. The most reason behind this can be that if sentence extraction is predicated on each chain, it'd end in massive outline size further as increase the likelihood of adding impertinent sentences to the outline due to the low scored lexical chains. Thus, for sentence extraction, we discover the primary sentence that contains one in all the chain members for each sturdy chain.

2) **Exploitation proper name Scoring**:

Since we tend to are summarizing news articles, some extraction on the idea of correct nouns is critical. We tend to earlier delineate our marking technique for correct nouns. Exploitation these scores, we tend to tried to explore a correct heuristic which might facilitate generate relevant and elliptical outline. When testing for multiple heuristics, the ultimate accepted heuristic is to extract the primary sentence for all those correct nouns whose score is bigger than common fraction of the quantity of sentences within the article.

Score(Proper Noun) > 1/3*Count(Sentences)

The main plan behind keeping the heuristic to match the quantity of sentences to the right noun score, i.e. the frequency of the right noun within the text was that, if a correct noun happens sizable amount of times in a very newspaper article, it's to be relevant to the topic of the article. If we tend to had compared the score of {a correct} noun with the common of the several proper nouns, it'd have chosen correct nouns relative to different correct nouns within the newspaper article. But our main aim here is to seek out some proper name that dominates the newspaper article in itself. Thus, we tend to use the count of the sentences. When testing for multiple values, common fraction was found to come up with the foremost acceptable outline. Thus, when selecting the vital correct nouns, we tend to extract the primary sentence that happens within the article. All of those techniques would extract a set of the sentences from the article. Our final outline would comprise of union of all the distinctive sentences extracted by every of the higher than delineate techniques.

The outline generated by our approach consists of the vital sentences known by lexical chains further because the sentences containing very important info concerning the subjects (proper noun occurrences) within the article. We've got mentioned the results of our experiments on one such newspaper article within the next section.

4. EXPERIMENTAL RESULTS

We tested our algorithm at all stages on various inputs of various lengths to understand its strength and weaknesses. Following is one such article we tested our algorithm on along with the generated summary

4.1 ARTICLE

Islamic State (ISIS) hackers have printed a" hit list" of over seventy United States of America military personnel WHO are concerned in drone strikes against terror targets in Syria and asked their followers to "kill them where they are". According to 'The Sunday Times', the hackers have links with United Kingdom of Great Britain and Northern Ireland and decision themselves 'Islamic State Hacking Division' and circulated on-line the names, home addresses and pictures of over seventy United States of America employees, together with ladies and urged supporters: "Kill them where they're, play their doors and kill them, stab them, shoot them within the face or bomb them." The cluster conjointly claimed that it'd have a mole within the UK's ministry of defense and vulnerable to publish "secret intelligence" within the future that would identify Britain's Royal Air Force (RAF) drone operators. The new list options the ISIS flag higher than the heading: 'Target – u.

s. Military' and therefore the document, circulated via Twitter and denote on the Just Paste web site, states: "You crusaders that may solely attack the troopers of the Muslim State with joysticks and consoles, die in your rage! "Your military has no spirit, neither has your president as he still refuses to send troops. Thus instead you press buttons thousands of miles away in your feeble plan to fight United States of America. "A nation of cowards that holds no bravery as you resort to causing your pilotless unmanned Reaper and Predator drones to attack United States of America from the skies. Thus this is often for you, America. "These seventy five crusaders are denote as targets for our brothers and sisters in America and worldwide to search out and

kill." The cluster conjointly warned: "In our next leak we tend to could even disclose secret intelligence the Muslim State has simply received from a supply the brothers within the United Kingdom of Great Britain and Northern Ireland have spent it slow exploit from the ministry of defense in London as we tend to slowly and on the Q.T. infiltrate European country and therefore the USA on-line and off. "At very cheap of the ISIS document is a picture of the sculpture of Liberty with its head bring to an end. The ISIS hacking division was antecedent crystal rectifier by Junaid Hussain, a former British Muslim pc hacker from Birmingham WHO was killed in a very United States of America drone strike in Syria last August. His wife, Sally Jones, a Muslim convert from Kent within the United Kingdom of Great Britain and Northern Ireland, continues to be believed to be concerned within the organization, that within the past has urged "lone wolf" attacks against RAF bases within the United Kingdom of Great Britain and Northern Ireland. Inquiries created by 'The Sunday Times' found that the names on the Yankee list are real. However, the data printed by ISIS doesn't seem to be the results of a leak or real hack. Instead, the cluster appears to own fastidiously gleaned the names of Reaper and Predator drone operators from news articles and military newsletters, before matching them to addresses, photos and different personal details from in public accessible sources on the net. A number of the data seems to own been taken from social media sites, together with Facebook and LinkedIn.

5. SUMMARY GENERATED

Islamic State (ISIS) hackers have revealed a "hit list" of over seventy US military personnel WHO are concerned in drone strikes against terror targets in Syria and asked their followers to "kill them where they are". The cluster additionally claimed that it'd have a mole within the UK's ministry of defense and vulnerable to publish "secret intelligence" within the future that might determine Britain's Royal Air Force (RAF) drone operators. "A nation of cowards that holds no bravery as you resort to causation your unmanned remote- controlled Reaper and Predator drones to attack US from the skies. The cluster additionally warned: "In our next leak we tend to could even disclose secret intelligence the Islamic State has simply received from a supply the brothers within the United Kingdom have spent it slow effort from the ministry of defense in London as we tend to slowly and in secret infiltrate European country and therefore the USA on-line and off." The ISIS hacking division was antecedent light-emitting diode by Junaid Hussain, a former British Muslim laptop hacker from Birmingham WHO was killed in a very US drone strike in Syria last August. Inquiries created by "The Sunday Times" found that the names on the yank list area unit real. Instead, the cluster looks to possess fastidiously gleaned the names of Reaper and Predator drone operators from news articles and military newsletters, before matching them to addresses, photos and different personal details from publically obtainable sources on the web.
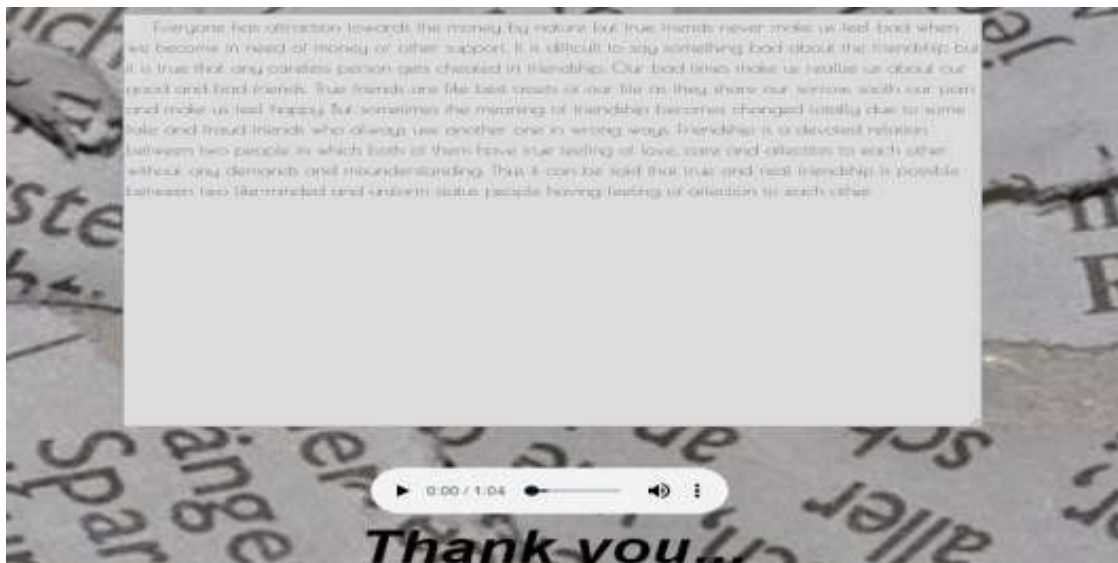
6. SNAPSHOTS

Figure 4.1 Input Text
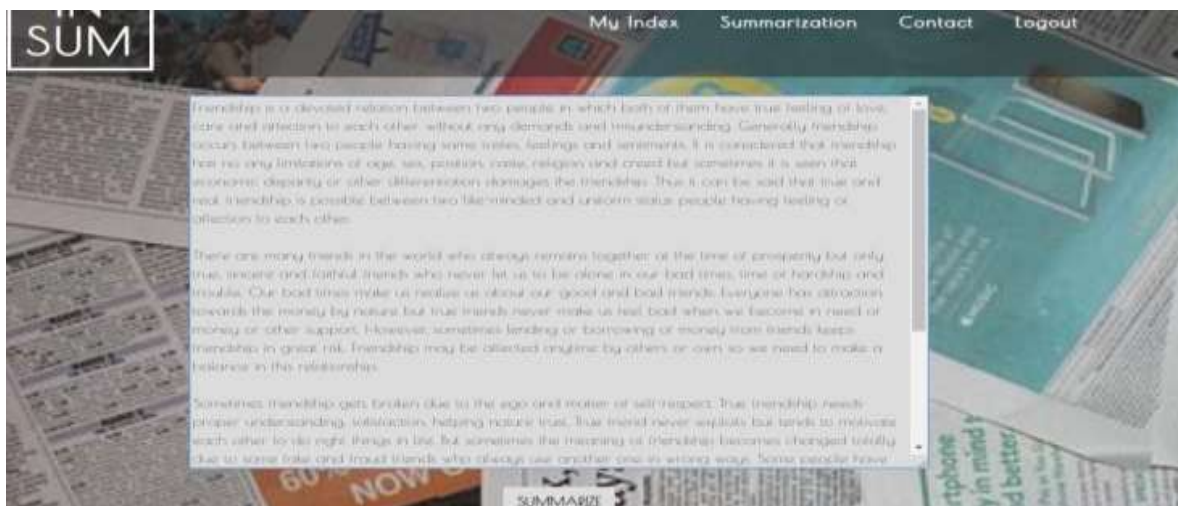
Figure 4.2 Output

7. CONCLUSION

We were able to aut o-summarize news articles and compare summaries generated by them to research what rating parameters would result in higher results. Within the method, we tend to tweaked strategies we tend to had researched on to leverage the very fact that we tend to were addressing news articles solely. We tend to found that journalists follow a set pattern to write down an article. They begin with what happened associated once it happened within the initial paragraph and continue with an elaboration of what happened and why it happened within the following paragraphs. We tend to needed to use this data whereas rating the sentences by giving the nouns showing within the initial sentence a better score. However once reviewing the preliminary results of our rating methodology as represented in Barzilay and Elhadad, we tend to complete that the primary sentence continually got a high score since it had nouns that were continual many times within the article. This is often intuitively consistent since the primary sentence of the article continually has nouns that the article talks regarding i.e. the subject of the article. In, lexical chains were obtaining created in exponential time. We tend to enforced a linear time formula as represented in Silber and McCoy to explore graph primarily based algorithms for sentence grading and extraction.

8. REFERENCES

[1] H. Saggion and T. Poibeau," Automatic text summarization: Past, present and future", Multi- source, Multilingual Information Extraction and Summarization, ed: Springer, pp. 3- 21., 2013

[2] M. Haque, et al.," Literature Review of Automatic Multiple Documents Text Summarization", International Journal of Innovation and Applied Studies, vol. 3, pp. 121- 129, 2013.

[3] D. R. Radev, et al.," Introduction to the special issue on summariza- tion", Computational Linguistics, vol. 28, pp. 399-408, 2002.

[4] C. Fellbaum," WordNet: An Electronic Lexical Database",Cambridge, MA: MIT Press., 1998

[5] G. A. Miller," WordNet: A Lexical Database for English", Communications of the ACM, Vol. 38, No. 11: 39-41., 1995

[6] Morris, Jane and Graeme Hirst. " Lexical cohesion computed by thesaural relations as an indicator of the structure of text", Computational linguistics 17.1, 21-48, 1991

[7] R. Barzilay, & M. Elhadad," Using lexical chains for text summarization", Advances in automatic text summarization, 111-121, 1999

[8] H. G. Silber, & K. F. McCoy, "Efficiently computed lexical chains as an intermediate representation for automatic text summarization", Computational Linguistics, 28(4), 487-496. , 2002

[9] M. Galley, K. McKeon, E. Fosler-Lussier, and

H. Jing, Discourse segmentation of multi-party conversation, in Annual Meeting- Association for Computational Linguistics, vol. 1. Association for Computational Linguistics, 2003

[10] S. Brin and L. Page, The anatomy of a large- scale hyper textual Web search engine, Computer Networks and ISDN Systems, vol. 30, no. 17, pp. 107117, 1998