



AN IOT BASED AUTOMATIC TEXT SUMMARIZATION

ANJANA SEBASTIAN

Department of computer science
St. Joseph's college (Autonomous)
Irinjalakuda, Thrissur, Kerala

JASMINE JOSE

Department of computer science
St. Joseph's college (autonomous)
Irinjalakuda, Thrissur, Kerala

ABSTRACT - The rate of people reading the books, newspaper etc... Are reducing day by day. Lack of time and lack of patience are the two main reasons behind it. Our project finds a remedy for this situation. The aim our project is reduces any contents which is given and automatically finds its summary and tells the whole summary of that content.

The summary can be speak out using Text to Speech application and also it will be sent to cloud. Like this any number of contents can be summarized and can be further summarized if they are almost similar contents and find the resultant summary.

Keywords-Summary,Content,Text Speech

I. INTRODUCTION

Text Summarization deals with concise and compressed form of original documents[1][2][3]. There are several application areas of text summarization such as web mining, genre classification and search engines[4][5][6][7][8]. Techniques used for text summarization may be categorized as extractive and abstractive[1][3]. Extractive techniques generate summaries based on

information units from original text such as sentences or paragraphs[1][3][5].

An important concern in extractive summarization is to select important sentences or paragraph[1][5][9][3]. Text summarization also known as document summarization helps to overcome information overload problem by providing more information content is less space[1]. Techniques for text summarization are broadly classified as extractive and abstractive. We focus on extractive techniques as they are more commonly used for generating text summaries. An important issue in extractive techniques is how to select importance sentences. One class of sentence selection approaches are based on notation of similarity. A particular approach in this regards is to select sentences based on their similarities with the topic[10]. The topic may be explicitly given or may be created or generated from the given text such as topic based on most frequent words appearing in the text[11].

The rate of people reading the books, newspaper etc... are reducing day by day .Lack of time and lack of patience are the two main reasons behind it. Our project finds a remedy for this situation. The aim our project is reduce any

contents which is given and automatically finds its summary and tell the whole summary of that content. Initially we give some content as input and it gets split into sentences and paragraphs and then calculates the intersection between every two sentences. The score will be calculated by calculating the sum of all its intersection by splitting the sentences into words and format the sentences. Build the summary by splitting the contents into paragraphs and get the best summary and get the best sentence from the paragraph.

The summary can be speak out using Text to Speech application and also it will be sent to cloud. Like this any number of contents can be summarized and can be further summarized if they are almost similar contents and find the resultant summary.

II. EXISTING SYSTEM

The existing system focuses on the text summarization. Earlier, it was necessary to read the entire text in order to get its summary. The existing system had lots of disadvantage. One of the main disadvantage of the existing system is that it is really time consuming. Another disadvantage is that the user must know how to read. The rate of people reading the books, newspaper etc... are reducing day by day. Lack of time and lack of patience are the two main reasons behind it.

III. PROPOSED SYSTEM

The proposed system is that it is time managing system and the user does not want to waste their time and effort in reading the content. Also, the user who don't know how to read, can also understand the content as they can hear the voice playback of the final summary. Along with that they can read the summary and send via telegram. It requires less human effort. Techniques used for text summarization may be categorized as extractive and

abstractive. Extractive techniques generate summaries based on information units from original text such as sentences or paragraphs.

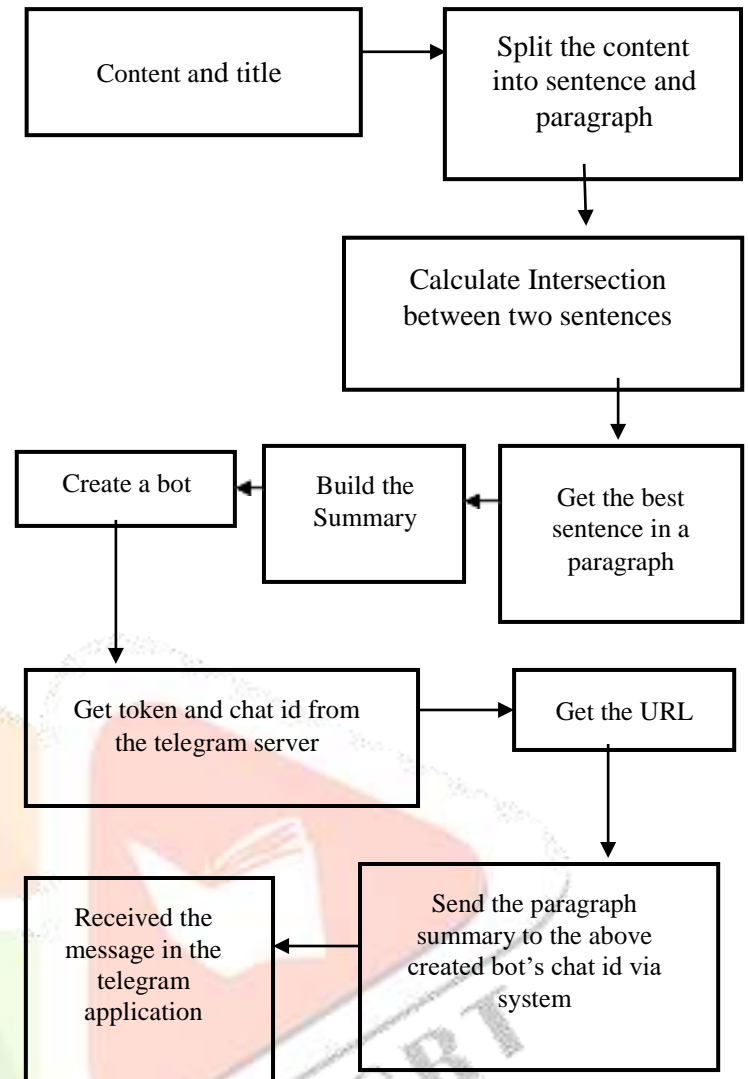


Figure 1. BLOCK DIAGRAM WORKING

IV. METHODOLOGY

The requirement is to reduce the reduce any contents which is given and automatically finds its summary and tell the whole summary of that content. Initially we give some content as input and it gets split into sentences and paragraphs and then calculates the intersection between every two sentences. The score will be calculated by

calculating the sum of all its intersection by splitting the sentences into words and format the sentences. Build the summary by splitting the contents into paragraphs and get the best summary and get the best sentence from the paragraph..The summary can be speak out using Text to Speech application and also it will be sent to cloud. Like this any number of contents can be summarized and can be further summarized if they are almost similar contents and find the resultant summary. The user does not want to waste their time and effort in reading the content. Also, the user who don't know how to read, can also understand the content as they can hear the voice playback of the final summary. Along with that they can read the summary and send via telegram.

V. CONCLUSION

Text summarization aims to generate brief and compact form of original text documents. Techniques used for text summarization may be categorized as extractive and abstractive. Extractive techniques generate summaries based on information units from original text such as sentences or paragraphs.

An important concern in extractive summarization is to select important sentences or paragraph. Text summarization also known as document summarization helps to overcome information overload problem by providing more information content is less space. Techniques for text summarization are broadly classified as extractive and abstractive. We focus on extractive techniques as they are more commonly used for generating text summaries. An important issue in extractive techniques is how to select importance sentences. One class of sentence selection approaches are based on notation of similarity.A particular approach in this regards is to select sentences based on their similarities with the topic. The topic may be explicitly given or may be created

or generated from the given text such as topic based on most frequent words appearing in the text.

REFERENCES

- [1] R. Barzilay and M. Elhadad, "Using lexical chains for text summarization," *Advances in automatic text summarization*, pp. 111–121, 1999.
- [2] Y. Gong and X. Liu, "Generic text summarization using relevance measure and latent semantic analysis," in *Proceedings of the 24th Annual International Conference on Research and Development in Information Retrieval*, pp. 19–25, 2001.
- [3] M. Pourvali and M. S. Abadeh, "Automated text summarization base on lexicales chain and graph using of wordnet and wikipedia knowledge base," *arXiv preprint arXiv:1203.3586*, 2012.
- [4] S. Afantenos, V. Karkaletsis, and P. Stamatoopoulos, "Summarization from medical documents: a survey," *Artificial intelligence in medicine*, vol. 33, no. 2, pp. 157–177, 2005.
- [5] V. Gupta and G. S. Lehal, "A survey of text mining techniques and applications," *Journal of emerging technologies in web intelligence*, vol. 1, no. 1, pp. 60–76, 2009.
- [6] M. Q. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth International Conference on Knowledge Discovery and Data Mining*, pp. 168–177, 2004.
- [7] P. Turney, "Coherent keyphrase extraction via web mining," 2003.
- [8] L. Yu, S. Y. Wang, and K. K. Lai, "A rough-set-refined text mining approach for crude oil market tendency forecasting," *International journal of knowledge and systems sciences*, vol. 2, no. 1, pp. 33–46, 2005.
- [9] Y. Ouyang, W. j. Li, R. X. Zhang, S. J. Li, and Q. Lu, "A progressive sentence selection strategy for document summarization," *Information processing management*, vol. 49, no. 1, pp. 213–221, 2013.
- [10] D. Shen, J. T. Sun, H. Li, Q. Yang, and Z. Chen, "Document summarization using conditional random fields," in *Proceeding of International Joint Conference on Artificial Intelligence*, vol. 7, pp. 2862–2867, 2007.
- [11] H. P. Luhn, "The automatic creation of literature abstracts," *IBM Journal of research and development*, vol. 2, no. 2, pp. 159–165, 1958.
- [12] D. Shen, J. T. Sun, H. Li, Q. Yang, and Z. Chen, "Document summarization using conditional random fields," in *Proceeding of International Joint Conference on Artificial Intelligence*, vol. 7, pp. 2862–2867, 2007.
- [13] X. Wan, J. Yang, and J. Xiao, "Manifold-ranking based topic-focused multi-document summarization," in *Proceeding of International Joint Conference on Artificial Intelligence*, vol. 7, pp. 2903–2908, 2007.