International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue V, May 2019- Available at www.ijraset.com

Smart Question Generation System using TextRanking

Prof. M. D. Nirmal[1], Swapnil Sonawane[2], Prajakta Dighe[3] Anuradha Adhav[4]
[1, 2, 3, 4]*Department of Computer Engineering Pravara Rural Engineering College Loni, Savitribai Phule Pune university*

Abstract: Indirect regular study plays important role in student's academics and provide solution for improving performance of student in exams. Also, students want to study from the various sources , but as it is a large source of data it becomes difficult to gain relevant and necessary information. So, we can make them learn at least gist of various units over the period of a semester. The application should provide a short summary of topics in the syllabus. It will explain those topics by providing a brief view about the information. This will provide the students a way to quickly summerise the syllabus topics. We can achieve this with the help of machine learning. By mining key text fragments from text, extractive data isgenerated. It uses statistical analysis of individual or mixed surface level features to find the location from where the sentences are to be extracted. The process contains different techniques like tokenization, removal of stop words, lexical analysis, feature extraction, clustering and sentense extraction. The existing system can be used for the purpose of the text extraction and question generation. The new proposed system will provide additional feature of generating questions on the extracted data in order to sure deeper study of a specific topic.

Keywords: Data Mining, Machine learning, Text Summarization, NLP, Lemmatization, tokenization

I. INTRODUCTION

A. Motivation

We have observed that professor from Every college face problem as students are not interested to attend regular classes. So, student prefer to do last night study before the exam, result into reducing the final score. We can help those professors by developing an application which will increases students interest in regular study. Such that, student have habit to read topics quickly and regularly.

Students want to study from the Internet, but as it is a large source of data it becomes difficult to gain relevant and necessary information. So, we can make them learn at least list of various units over the period of a semester

B. Problem Statement

Text summarization has grown into a important and exact engine for supporting and provide text content in the latest speedy emergent information age. It's far very difficult for humans to physically summarize oversized documents of text. There is a wealth of textual content available on the internet. But, usually, the internet contributes more data than is desired.

Students want to study from the Internet, but as it is a large source of data it becomes difficult to gain relevant and necessary information. So, we can make them learn at least list of various units over the period of a semester

II. LITERATURE SURVEY

The work which has been done on summarizing the single document is mainly on technical documents. The most cited technical paper on summarization is Luhn, 1958 . This paper describes the research work done in 1950s at IBM. In 1958, Baxendale claimed that the sentence position provides an insight into its relative importance in the document .

In 1969, Edmundson illustrated a structure that constructs document extracts . In 1995, Kupiec et al. described a method which was derived from Edmundson , which was capable to be trained from data . Each sentence is worthy or not will be categorized by the classification function, by applying Naive-Bayes classifier. In 2018 Jeong-Woo Son, Wonjoo Park, Sang-Yun Lee and Sun-Joong Kim authored this topic and explained there idea as Titles of videos are the most important aspect to provide various services.. Since videos in such services are often automatically generated by segmenting a video, these contents cannot have their own titles. As a result, titles of the video fragments are annotated by human hands. To reduce the cost for manual annotation of video titles, this paper proposes a novel method to generate titles of videos by selecting informative sentences from closed captions. The proposed method utilizes explicit and implicit relations among words occurred in closed captions by constructing a stochastic matrix. And then, the proposed method picks important words up based on their weights estimated with TextRank.

Paper Name	Author	Related Work
Automatic Extractive Text Summarization using K-Means Clustering	Krithi Shetty , Jagadish S Kallimani (In 2017)	In this Paper ,Summarization of the text can informally be defined as the act of condensing the document from its original size without significantly compromising the semantics.
Multi-document text summarization	Amol Tandel , Priyasha Gupta (In 2017)	This application will allow the user to automatically summarize relevant information from various sources.
Video Scene Title Generation based on Explicit and Implicit Relations among Caption Words	Jeong-Woo Son, Wonjoo Park, Sang-Yun Lee, Sun-Joong Kim (Feb 2018)	In this paper proposes a novel method to generate titles of videos by selecting informative sentences from closed captions.

In 2017 Krithi Shetty and Jagadish S Kallimani authored this paper explained that The rise in the dimension of the World Wide Web has made an explosion of the amount of accessible information. As the textual data involves several instances of redundancy, omission of part of sentences or entire sentences is possible without altering the meaning of the document. Summarization of the text can informally be defined as the act of condensing the document from its original size without significantly compromising the semantics. For the purpose of generating an appropriate summary, the raw text is first pre-processed which involves removing non-ASCII characters and stop-words, tokenizing and stemming. Appropriate features are extracted from the data, tf-idf values for each word are computed and the entire pre-processed data is then transformed into a tfdf matrix. Every sentence of the document will be represented as a vector in the dimensional space of the document's vocabulary. To obtain a concise summary, sentences are appropriately clustered based on the degree of separation of vectors in the Euclideanplace. Association of sentences to a cluster using K-means method is totally based on cosine similarity. The count of the clusters is to be formed is predefined. As the number of clusters increase the accuracy of the summary increases. From each of the clusters the sentences which are informative are picked to form the final summary. Using recall and precision measures, the effectiveness of the summary is verified.

III.PROPOSED SYSTEM

A. Goal And Objectives

1) To make effective use of Google Search API.
2) To make effective use of raw data from web.
3) Remove Stop Words.
4) Apply Lexicon Grammar.
5) Apply Text Ranking.

B. Statement Of Scope

In this System we Use Text Ranking Algorithm instead of K-mean Algorithm
It generates two datasets one will be lexical analysis bases and another will ranking based dataset

1) *Search Web:* The System can be used to analyze the data i.e. searched from the web and download it using Google Search Engine API. Then the stop should also be removed from the raw data.
2) *Question Generator:* The System can be used to analyze and rank the data after lexicon analysis and rank the data according to the strength of each keyword that is to be defined first as a training dataset. It is achieved by using Text Ranking Algorithm.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue V, May 2019- Available at www.ijraset.com

IV. SYSTEM ARCHITECTURE

A. *System Architecture*

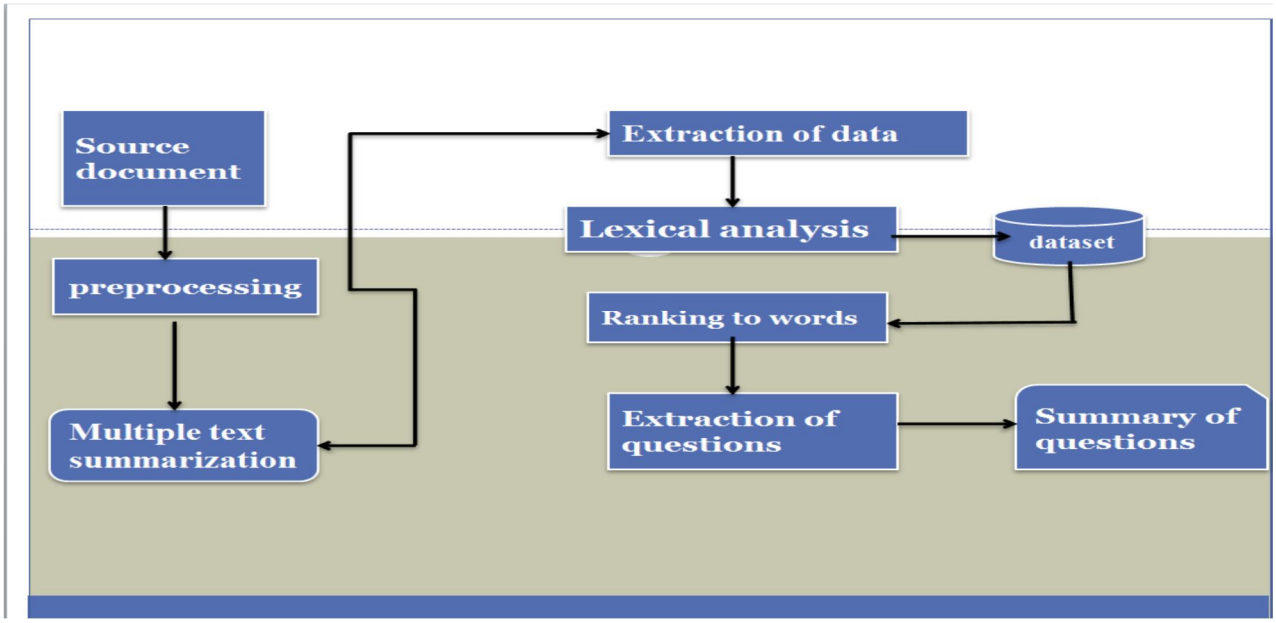

Fig: System Architecture

B. *Algorithum*

TextRank is a well-known method to extract keywords among words with their relations. Basically, TextRank has been adopted in text summarization. TextRank determines centralities of words with an iterative process. Then, a set of words are extracted with respect to their centralities To go from an string of text to a list of scored sentences based upon how much they represent the overall text, we need to go through below steps:

1) Tokenize the text into sentences
2) Tokenize every sentence into a collection of words
3) Convert the sentences into graphs
4) Score the sentences via pageRank.

V. APPLICATIONS

A. This application is help to get abstract information from the Internet and generate MCQ Based Questions
B. The proposed system will provide additional feature of generating questions with sort summery of answer on the extracted data in order to sure deeper study of a specific topic.

VI. CONCLUSION

In this paper, model proposes a Question generation method on given syllabus for students ,since question are efficient way to connect students with syllabus it is much importance factor for various services In the proposed method the questions are generated automatically with Rankwise in closed captions. The importance of question is determine with word weight given to it in the predefined dataset

REFERENCES

[1] A survey paper of Alguliyev, Rasim, Ramiz Aliguliyev, and Nijat Isazade. "A sentence selection model and HLO algorithm for extractive text summarization." Application of Information and Communication Technologies (AICT), 2016 IEEE 10th International Conference on. IEEE, 2016.
[2] Moratanch, N., and S. Chitrakala. "A survey on extractive text summarization." Computer, Communication and Signal Processing (ICCCSP), 2017 International Conference on. IEEE, 2017.
[3] Tandel, Amol, et al. "Multi-document text summarization-a survey." Data Mining and Advanced Computing (SAPIENCE), International Conference on. IEEE, 2016.
[4] GOTO, Takuya, et al. "An Automatic Generation of Multiple-choice Cloze Questions Based on Statistical Learning."