Parshvanath Charitable Trust's

A. P. SHAH INSTITUTE OF TECHNOLOGY

(Approved by AICTE New Delhi & Govt. of Maharashtra, Affiliated to University of Mumbai)
(Religious Jain Minority)

Department of Information Technology

Academic Year: 2017-18
Semester: VII
Class / Branch: BEIT
Subject: Project
Name of Instructor:

Name of Student: Pratiksha Patil
Student ID:16204011
Date of Performance: 24/08/18
Date of Submission:24/08/18

Project Presentation 1

GROUP NUMBER :

Name of Group Member:
1.Riddhi Prajapati(Leader)
2.Pratiksha Patil
3.Aafreen Shaikh

Guide:
Co-Guide:

INTRODUCTION

- Nowadays sifting through lots of documents can be difficult and time consuming.
- Without an abstract or summary, it can take minutes just to figure out what the heck someone is talking about in a paper or report.
- For this reason, advanced tools for categorizing, indexing and extracting information from multimedia documents will become indispensable in the near future .
- Automatic text summarizer is a dominant automated software tool for processing large amount of online information.
- It creates a summary with major points of the original document. It is a part of Data Mining.

CONTINUE...

- The main idea is to find a subset of data which contains the "Information" of entire set.
- Automatic summarization of text works by first calculating the word frequencies for the entire text document.
- Each sentence is then scored based on how many high frequency words it contains, with higher frequency words being worth more.
- Finally, the top X sentences are then taken , and sorted based on their position in the original text.

ABSTRACT

- An Automatic Text Summarizer should produce the summary quickly with no redundancy.
- There are two algorithm to Automatic Text Summarization NLP(natural language processing) and LEXICONS.
- ATS evaluation techniques fall into intrinsic and extrinsic. Here we are using intrinsic evaluation.
- The rate at which the information is growing is tremendous. Hence it is important to build a multilingual summarization system.
- The goal of this project is to explore methods for automatically summarizing spoken documents.

PROPOSED SYSTEM

Some features are used to score sentences:

- Word Frequency
- Length of sentence
- Position of sentence
- Title summary

SUMMARIZATION FLOW:

Source Document(Input)
↓
Preprocessing
↓
Feature Extraction
↓
Sentence scoring &Ranking
↓
Sentence extraction
↓
Main Summary(Output)

TECHNOLOGY STACK

- R Programming Language it is open source language.
- Machine Learning is used which is closely related to the Data Mining which is used for classification.

Project Planning

Sr.No	Duration	Activity
1	Research about project Technology and algorithm	August End
2	Learning	Mid of September
3	Start project Designing	End of October
4	Development	1st week of December
5	Testing	End of December
6	Modification	End of February

THANK YOU…