**IEEE TECHNICAL PROJECT EXHIBITION 2015, New Delhi**

**ABSTRACT**

**Author: Ayush Pareek**

**Project Name:** Automatic Text Organizer and Summarizer

**Abstract:**

PURPOSE::

We have built an extraction-based text summarizer along with text-organizing capabilities that can generate genre-specific, generic or user-configured summaries of a large amount of unorganized text using data mining. We are currently using a number of independent text-mining algorithms based on different statistical models to compute the summaries and combining them using some innovative configurable consensus techniques.   
  
Our Consensus method is dynamic in the sense that they could be based on assigning abstract weights to different algorithms and then take a weighted mean(in abstract sense) of the sentence ranking produced by each algorithm OR they could be manually-specified OR they could be based on the amount of relativity between the title of the text/certain keywords provided by the user and the summary produced by each algorithm. The overall procedure is fairly unique, sophisticated and better than any other being used for the same purpose.

We have also proposed several unprecedented and unorthodox applications of our project which have universal social and business value.

**Some Features of our Project:**

**(1)Unlimited number of algorithms which might have data-specific importance can be incorporated** under our unified project. For Eg; While making summaries of academic papers involving formulas and mathematical expressions, if an a few algorithms give better results than others, we can give them more weight-age during consensus.

**(2)Language Independent summaries**:- Except for the pre-processing stage, our algorithm works perfectly on any language that has the notion of words and sentences. Even for pre-processing, the stop words lists and stemming algorithms have already been developed and can be harnessed to make the summary platform multi-lingual.

**(3)Cloud Computing** :- Several Android and Desktop applications(see below) have been suggested like News-summary Maker, e-paper summarizer, Internet-plagiarism detector etc which can provide vast utility via data mining using a network of remote servers hosted on the Internet to store, manage, and process data.

**(4)Parallel Processing:-** The algorithms being used can work independently from each other, opening doors for processing them simultaneously using multiple processors while handling big-datasets.

**(5) A Generic summary:-** provides the main theme and topics of the text and keep redundancy to a minimum. On the other hand, Query-based text summaries are useful for answering such questions as whether a given document is important with respect to to the user’s query, and if yes, which part(s) of the document are important. Eg; If you want to revise some mathematics notes, then it would be able to extract a list of all the formulas in the text and so on...

**Few of the *numerous avant-garde applications which could be built manoeuvring our algorithm as the prime basis as follows:***

* **Sub-Heading and Index Creator:-** Our software can convert a big unorganized chunk of text in an organized format with sub-headings and sub-points which can be used to form INDEX for a data set (For Eg; a book)
* **Content Highlighter:-** Instead of a summary maker, our project can also work as a content highlighter. It can mark the most important parts in each paragraph of the entire text so that a person can decide whether to read it or not just by quickly-skimming through those parts
* **Browser Add-On:-** Whenever you come across a some text in a browser which you don’t want to read completely, we’ll give you a small button to convert it into a good summary
* **Subjective Exam sheet checker:-** This can also be used for evaluating subjective answers of students. Using our algorithm you can first make your own answer to every question and take the summary of the answer given by students to check if they are writing on the same basic points.
* **Making Abstract of Research papers and articles:-** It can also be used to create abstracts of a research paper by extracting the most important parts of the entire text.
* **Plagiarism Detector:-** Summaries of documents can be compared to check whether the most important parts of a document have been copied or not.
* **Hypertext context-link based summarizer:-** Now an advanced form of summary can also be generated by my software which makes every sentence a hypertext. Now if you don’t get the context of any of the sentences. Just click on that sentence and it'll temporarily take you to the place it is extracted from so that you can get a quick idea of context by reading a few sentences here and there.
* **Daily News feed summarizer / RSS :-** Another good application is that of a news summary maker. What it means is that there are millions of news articles published per day. Our software can take all those articles on a specific subject and provide you the news in as many lines as you want. The news generated will capture exhaustive information on the subject in question from a variety of sources.
* **Can be used in search engines/e-mail services to present compressed descriptions of the search results :-**.
* **In keyword directed subscription of news which are summarized and pushed to the user.**