

# Summarized News Application using TF-IDF

Badreesh Shetty<sup>1</sup>, Vinayak Shetty<sup>2</sup>, Rohit Shinde<sup>3</sup>, Prof. Torana Kamble<sup>4</sup>

<sup>1,2,3</sup> Student, Final Year Computer Engineering, Navi Mumbai

<sup>4</sup> Professor, Dept. of Computer Engineering, Bharati Vidyapeeth College of Engineering, Maharashtra, India

\*\*\*

**Abstract** - Modern hand held devices such as smartphones and PDAs have become increasingly powerful in recent years. Dramatic breakthroughs in processing power along with the number of extra features included in these devices have opened the doors to a wide range of commercial possibilities. In Particular, most cell phones regularly include cameras, processors comparable to PCs from only a few years ago, and internet access.

As mobile devices become more like PCs they will come to replace objects we tend to carry around such as cameras, mp3 players, credit cards, etc. In short, we will be using them to accomplish our daily tasks. One application that falls into this category is the Summarized News App.

The prime objective of "Summarized News Application" is to create an Android Application that will give a summarized version of a long article, thereby reducing reading time of readers and providing only the crux of the article. Reader can search already loaded articles in the app which have been summarized. Reader can bookmark articles for further reference in the app. Articles are categorized for the Reader's convenience as the user can access news faster through Categories. Images are loaded according to relevance of the article.

The Project is developed in Android Studio using Java and Sublime for PHP with Cmdr as Command Panel.

**Key Words:** News, TF-IDF, News Summary, Bookmark, Push Notification.

## 1. INTRODUCTION

The era of mobile technologies opens the windows to the android app. The websites are vanishing and the mobile phones are emerging. We are introducing an android application software which would let us read news and also provide news in more summarized manner. It is helpful for those people who doesn't like to read the whole news. It acts as an overview of the whole news.

## 2. PROBLEM STATEMENT

To create an android application for readers to minimize their reading time. Reader needs to be updated with the latest news. So the android application should be easy and efficient to use. Readers have to often read long articles with the most important details not pointed out. A

concise and succinct news which is summarized assures maximum Reader satisfaction.

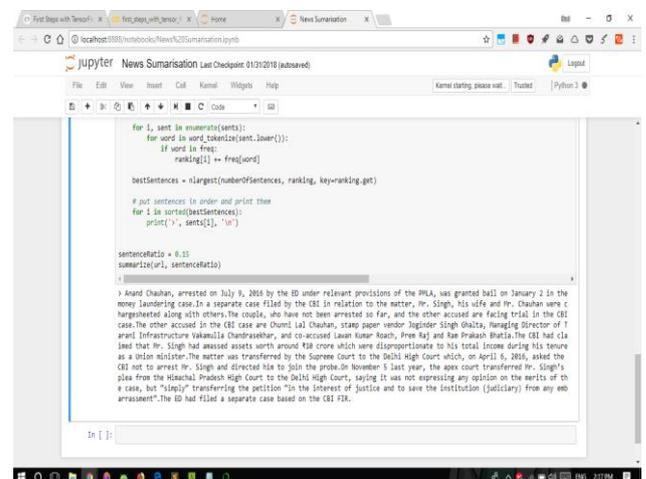
## 3. OBJECTIVES

- The primary objective of our New Application is to provide a quality summarized news to the users.
- To identify the top and trending news and provide it to the users
- Categorize the news into different sub categories and providing it to the users in effective and easy form.
- Summarize the news so that user's attention is not diverted and keeping the news easy and simple.
- It's a very useful and informative app. It keeps one updated and saves a lot of time by providing news in just 60 words that can be read in just 30 seconds.

## 4. PROPOSED MODULES

### A. Generation of Summarized News

Dedicated Admin Panel with Summarized News generated from TF-IDF Algorithm. Admin posts in the relevant categories. News is then visible to all the users.



```
for i, sent in enumerate(sentences):
    for word in word_tokenize(sent.lower()):
        if word in freq:
            ranking[i] += freq[word]

bestSentences = nlargest(numberOfSentences, ranking, key=ranking.get)

# put sentences in order and print them
for i in sorted(bestSentences):
    print("> ", sent[i], "\n")

sentenceRatio = 0.15
summarize(r1, sentenceRatio)
```

> Anand Chavhan, arrested on July 9, 2016 by the ED under relevant provisions of the PMLA, was granted bail on January 2 in the money laundering case. In a separate case filed by the CBI in relation to the matter, Mr. Singh, his wife and Mr. Chavhan were charged along with others. The couple, who have not been arrested so far, and the other accused are facing trial in the CBI case. The other accused in the CBI case are Churni Lal Chavhan, stamp paper vendor Jaginder Singh Ghaita, Managing Director of IT and Infrastructure Vahmaniia Chandrasekhar, and co-accused Laxmi Kumar Bhatnagar, Preet Raj and Ran Prakash Bhatnagar. The CBI had claimed that Mr. Singh had amassed assets worth around ₹18 crore which were disproportionate to his total income during his tenure as a Union minister. The matter was transferred by the Supreme Court to the Delhi High Court which, on April 6, 2016, asked the CBI not to arrest Mr. Singh and directed him to join the probe on November 9 last year, the apex court transferred Mr. Singh's plea from the Himachal Pradesh High Court to the Delhi High Court, saying it was not expressing any opinion on the merits of the case, but "simply" transferring the petition "in the interest of justice and to save the institution (judiciary) from any embarrassment". The ED had filed a separate case based on the CBI FIR.

Fig-1: Generation of Summary

**B. Categorized News**

Pre-categorized News is available to the User based on Relevance.

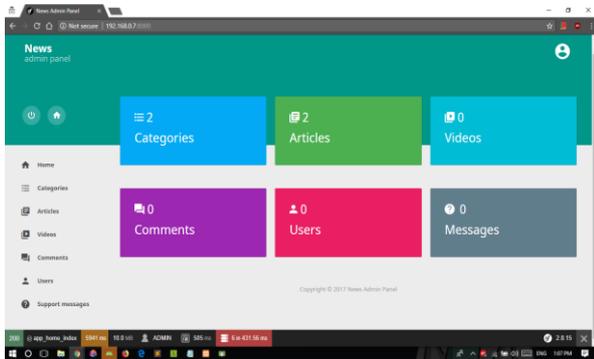


Fig-2: Admin Panel

**C. Home Feed**

News feed visible to user as per addition of articles to the Admin Panel. News is updated regularly.

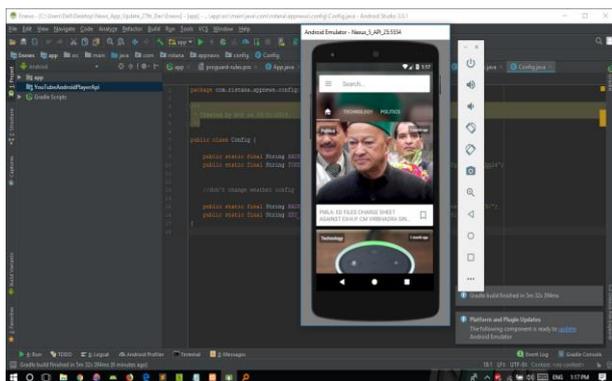


Fig-3: Home Feed of Device

**D. Push Notification**

Notification is sent to Readers where they can slide to avoid or tap on the news title to get further article details

**E. Bookmarked News**

Favorite News can be bookmarked for further reference of Reader. This function helps in reducing time for searching favorite news from the roster of articles

**F. Comment on Article**

Readers can give their view or opinion on by commenting on News article. All the Readers can view these comments on the articles opened.

**G. Sharing News**

Summarized News article can be shared to other Readers.

**5. WORKING OF TF-IDF ALGORITHM FOR SUMMARIZATION**

Before going through the Algorithm, we first need to understand how a TF-IDF works. A Term Frequency is a count of how many times a word occurs in a given document (synonymous with bag of words). The Inverse Document Frequency is the the number of times a word occurs in a corpus of documents. TF-IDF is used to weight words according to how important they are relevant to their context. Words that are used frequently in many documents will have a lower weighting while infrequent ones will have a higher weighting.

$$w_{i,j} = tf_{i,j} \times \log \left( \frac{N}{df_i} \right)$$

Where, w is weightage of word in context to article

tf is term frequency i.e. no of occurrences of word in article

df is inverse document frequency i.e. no. of occurrence of word in a corpus

N is the corpus document

**A. Import all Important Libraries**

Libraries like math for mathematical functions, re for Regex functions , requests for requesting URL Webpage, beautiful soup for scraping important details, stop words for flushing out stop words, corpus which is predefined document, sentence tokenize, word tokenize for tokenizing words as well as sentences.

**B. Request of URL and Scraping of Webpage**

Beautiful Soup is used for scraping web pages which have important details regarding. All text related to news is scraped leaving behind all images as well as ad related to the news article page.

**C. Extracting Page Data**

From requested web page article is extracted for performing the algorithmic functions. According to the title website the URL is parsed.

#### D. Summarization of Article

Tokens of Sentences and Words is generated and stopwords are not taken in to account as they weight wrong calculations. Frequency of tokens is found out with respect to the TF-IDF Algorithm. From sentence ratio the best sentences with have relevant information is generated. The generated Summary is then added to the Admin Panel in the System for posting it on the Android Application.

### 6. ADVANTAGES

A. When something is reportable anywhere around the world the news get updated regularly, where newspapers get typically printed once or most doubly every day. Whereas the online news is typically updated whenever there's one thing value news.

B. Online news are better as compared to regular newspapers as it saves a lot of time and money. This news doesn't need to be printed and there is no need for anyone to deliver them either. They are just published online and anyone from the world can view them with a few simple clicks.

C. News is available in summarized manner for people who are not interested to read the whole news or want just an overview of the news.

D. There is no limit to how many articles or news one can read. With newspapers, people can only read the articles or news contained in the newspaper.

### 7. REQUIREMENTS

#### 7.1 Hardware Requirements

Hardware	Specification
Processor	Any Snap dragon, Intel ,Exynos Chip
RAM	512 MB

#### 7.2 Software Requirements

Software	Specification
Operating System	Android
Platform	Android Studio, Sublime

### 8. FUTURE SCOPE

Features such as news recommender, news poll, news based on location, dictionary.

### 9. CONCLUSIONS

The application will help the readers in reading summarized news articles that saves a lot of reading time and the reader can send important articles to other readers. It is efficient and always available online.

Thus the application helps in keeping the reader updated on daily news. It is beneficial as it has no cost and always available online. Also, you will have a clear idea and understanding of what is happening in your country and the whole world.

### 10. ACKNOWLEDGEMENT

We take this opportunity to express our profound gratitude and deep regards to our guide Mrs. Torana Kamble for her exemplary guidance, monitoring and constant encouragement throughout the course of this project.

### 11. REFERENCES

- [1] <https://www.androidhive.info/2012/05/how-to-connect-android-with-php-mysql/>
- [2] <https://glowingpython.blogspot.in/2014/09/text-summarization-with-nltk.html>
- [3] <https://medium.com/@acrosson/summarize-documents-using-tf-idf-bdee8f60b71>
- [4] <https://www.codementor.io/flame3/send-push-notifications-to-android-with-firebase-du10860kb>