

# LIST OF FIGURES

CHAPTER NO.	TITLE	PAGE NO.
2	REVIEW OF LITERATURE	
	Figure 2.1 Text Summarization Approaches	5
	Figure 2.2 Types of Extractive Summarization	6
	Figure 2.3 Recurrent Neural Network Architecture	9
	Figure 2.4 Types of Evaluations of Summaries	10
	Figure 2.5 Algorithms for Intrinsic Evaluation	12
	Figure 2.6 Types of ROUGE Measures	13
3	PROPOSED SOLUTION	
	Figure 3.1 Component Diagram	18
	Figure 3.2 Swimlane Diagram	22
4	IMPLEMENTATION	
	Figure 4.1 Extractive Text Summarization	23
5	RESULTS AND DISCUSSION	
	Figure 5.1 Average ROUGE-2 Recall Scores	30
	Figure 5.2 Average ROUGE-2 Precision Scores	31
	Figure 5.3 F-Measure Comparison	32
	Figure 5.4 Average F-Scores	32

## LIST OF TABLES

CHAPTER NO.	TITLE	PAGE NO.
5	RESULTS AND DISCUSSION	
	Table 5.1 Word Frequency Evaluation	28
	Table 5.2 TF-IDF Evaluation	29
	Table 5.3 Gensim's TextRank Evaluation	30

# LIST OF EQUATIONS

CHAPTER NO.	TITLE	PAGE NO.
2	REVIEW OF LITERATURE	
	Equation 2.1 Word Frequency	7
	Equation 2.2 Term Frequency	7
	Equation 2.3 Inverse Document Frequency	8
	Equation 2.4 TF-IDF	8
	Equation 2.5 Recall	11
	Equation 2.6 Precision	11
	Equation 2.7 Brevity Penalty	11
	Equation 2.8 BLEU Score	12
3	PROPOSED SOLUTION	
	Equation 3.1 F-Measure	17

## ABBREVIATIONS

ABBREVIATION	DESCRIPTION
TF-IDF	Term Frequency – Inverse Document Frequency
BLEU	Bilingual Evaluation Understudy
BP	Brevity Penalty
ROUGE	Recall-Oriented Understudy for Gisting Evaluation
LCS	Longest Common Subsequence
WOEID	Where On Earth IDentifier
NLP	Natural Language Processing
NLTK	Natural Language Toolkit (Python)
API	Application Program Interface
HTML	Hypertext Markup Language
XML	Extensible Markup Language

## ABSTRACT

With the boom of social media in the 21st century, it has grown to impact nearly every aspect of our life. This widespread nature of social media makes it a critical platform for the spread of information. This information more often than not is filled with flaws such as incomplete information, uncredited sources and may even go as far as being downright false. Platforms such as Facebook, twitter, Instagram and many more receive more than a billion active users every day with double of that active every month making them excellent platforms for the spread of information. There have always been suspicions and even some verified examples of these platforms being used to influence the lives of people by feeding them news from biased sources in order to influence the decisions they make in real life.[14]

Twitter being a key player in the propagation of information with 500 million tweets containing snippets of information being sent every day, proved to be the ideal platform on which we could implement a system to verify and then spread information. The decision was made to consider the trending topics on twitter to be the keywords on which more information must be found and propagated.

Bots proved to be the ideal form with which could implement said system as it reduces the probability of human error, removes work redundancy and is inherently unbiased. The bot would scrape data off articles from reliable sources, summarize them and then tweet them along with a link to the actual source.

Summarization of the selected articles however proved to be the greatest challenge. We decided to focus on extractive methods of summarization where sentences are lifted from the source verbatim instead of abstractive methods which generated sentences due to its high complexity and resource requirements. We compared a variety of methods such as word frequency, term frequency - inverse document frequency and a few others and have found that ideal extractive summarization is accomplished using a recurrent neural network.