# Extractive Manuscript Summarization of Motivational Blogs in Natural Language Processing

Deepali Vaijinath Sawane[✉], Sanjay Y. Azade, Shabeena Naaz Khan, and Sarfaraz Pathan

MGM University, Aurangabad, MS, India
{dsawane,sazade}@mgmu.ac.in

**Abstract.** The growing use of search engines business surveys, market updates, news reports, patients medical records etc., the use of automatic manuscript summarization is becoming more and more popular. In this process, the original data is compressed without changing its meaning. It is a growing field of Natural Language Processing; which is a branch of linguistics, computer science, and artificial intelligence. It studies how computers and human language interact, with a focus on how to design computers to process and analyze massive volumes of natural language data. The NLP also provides users to get valuable knowledge with one click. It is a very beneficial process, particularly when dealing with large article and source manuscripts. It provides the meaningful summary with-in a fraction of seconds, and it is a systematic way to manage large amount of data available on search engines, otherwise, it becomes very tedious for managing and reading. Summarize Bot, Resoomer, SMMRY, Text Summarization and Text Compactor, are the top five online tools for text summarization, among these all tools, we chose the text compactor online tool for summarization with different percentages and parameter.

**Keywords:** MS (Manuscript Summarization) · TC (Text Compactor) · NLP (Natural Language Processing)

## 1 Introduction

The main motive of manuscript report is to compress the original source into smallest version, which in turn or creates a small document on multiple documents of huge manuscripts and articles [1]. The manuscript techniques increasingly get evolves in styles using which one can obtain imperative data, however the increasing need of compression is varied because of availability of information. The automatic script reporting is particularly useful for converting large manuscript into smaller and smaller version with economical and accurate contained as same as original source [2]. It is a turning into highly regarded space that belongs to tongue processing [3]. The main features of script report are that, it consists of term frequency, location, cure method, heading and proximity [4]. In old days the text summarization was able to seen in variety of
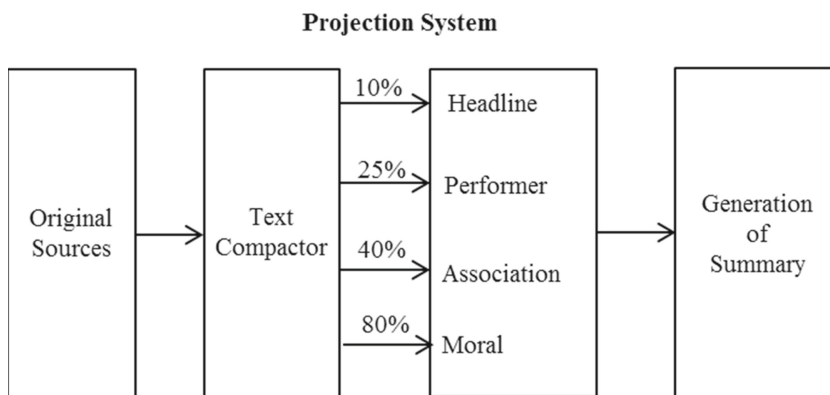
Projection System



**Fig. 1.** Projection System Architecture

contacts like news report summary, email summary, small news on mobile, business meeting summary, government sectors, research fellows and readers through online search browser obtain the summary of appropriate articles. In addition to these medical records of patients with prescriptions and advance cure treatments as per physicians [5]. In all application, summarization acts as an essential behavior. In this research we are going to summarize motivation blogs supported by a little quantity of unstructured moreover as we continue to use some system victimization text compactor [6].

## 2 Proposed Work

After going through rigorous study of existing compactors, we propose new projection system architecture as shown in Fig. 1.

As shown in Fig. 1, in the initial stage of the working process of automatic text summarization, we have collected some motivational blogs as input source manuscript. After that the rules are designed, which are applied to the summary generation procedure, which is framed for enrichment of the projected task of generating high quality summary with significant contents [7]. The source manuscript is considered as a motivational blog and passed to the text compactor for extracting the necessary data and projected stages as shown in figure, which represents the architecture and flow of proposed work.

### 2.1 Input/Source Data

We have considered five motivational blogs, which are available in variety of sizes and submitted it to text compactor for transitional outcomes. While doing this we have considered motivational blogs of maximum length 60 and minimum length of 35 sentences (Fig. 2).

### 2.2 Text Compactor

As the text compactor is online text compression tool it measured the frequency of words and sentences of original source, and sorted out the rate of reoccurrence of words and
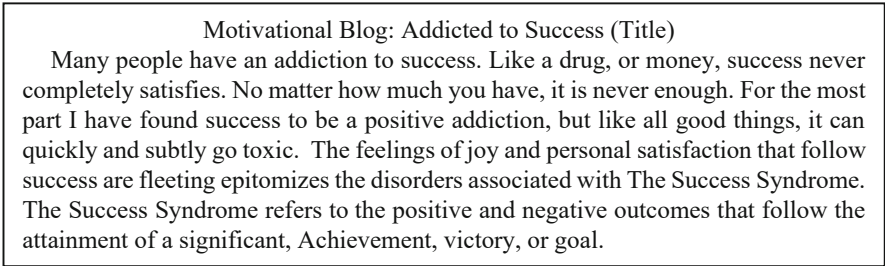
---

Motivational Blog: Addicted to Success (Title)

Many people have an addiction to success. Like a drug, or money, success never completely satisfies. No matter how much you have, it is never enough. For the most part I have found success to be a positive addiction, but like all good things, it can quickly and subtly go toxic. The feelings of joy and personal satisfaction that follow success are fleeting epitomizes the disorders associated with The Success Syndrome. The Success Syndrome refers to the positive and negative outcomes that follow the attainment of a significant, Achievement, victory, or goal.

---

**Fig. 2.** Motivational Blog as original Source

highly ranked sentences from the original source [8]. The extracted words and sentences are then calculated in percentage form each sentence is calculated on the basis of words regularity and association of the words with sentence it consists of [9]. The most essential and appropriate sentence is considered highly in terms of rate of reoccurrence [10].

## 2.3 Production Rules

We have framed the rule for the attribute of a source manuscript as like the title, performer, relation of the performer, activity going on at a particular location explained in sources, activity of the performer, and most important, motivation of the blogs. For summarization we have finalize some parameters which plays a significant role in generating quality summaries, and creates the abstraction the motivational blogs [11]. The parameters in rule based are framed as follows:

Rule 1: The first sentence of the outcome manuscript should be considered a heading or title.

Rule 2: All the performers in the blogs should be notified as "actors".

Rule 3: The place of the performer should be considered the "location".

Rule 4: The activity of the performer should be considered an "event".

Rule 5: The last line of the outcome of the summary is considered the moral or motivational of the blogs.

1. Performer ----- Relation ---- other performer.
2. Performer ----- Relation ---- Particular Event.
3. Performer ----- Relation ---- Specific Location.
4. Performer ----- Relation ---- Particular through/self, to be created as an association in the final manuscript.

## 2.4 Outcome

The source manuscript is passed through an online tool such as a text compactor at the initial stage of all tasks at 10%, 25%, 40%, and 80%. We have summarized motivational blogs in different ratios of percentage as like 10%, 25%, 40%, and 80% with the use of text compactor [12]. Overall details about the working of text compactor are how it can compress the text at which level and up to the ration of percentage in the result. We have defined it in a specific tables and figures as shown in Experimental result session3.

> Motivation Blog: Addicted to Success (Title)
> Many people have an addiction to success.

**Fig. 3.** Text compactor 10% summary

**Table 1.** 10% Summary Generation

| Motivational Blog | Words | Generated Summary |
|---|---|---|
| Addicted To Success | 258 | Heading $1^{st}$ Line |
| Deep Existence | 230 | Heading $1^{st}$ Line |
| Tiny Buddha | 268 | $2^{nd}$ Sentences of $2^{nd}$ passage, $1^{st}$ Sentences of $3^{rd}$, $4^{th}$, & $5^{th}$ passage |
| Fearless Motivation | 568 | $1^{st}$ line of $5^{th}$ passage, $1^{st}$, & $2^{nd}$ sentence of $5^{th}$ passage, $1^{st}$ sentence of $8^{th}$ and $9^{th}$ passage. |
| Successes Consciousness | 224 | Heading $1^{st}$ Line |

> Motivational Blog : Addicted to Success(Title)
>
> Many people have an addiction to success. The feelings of joy and personal satisfaction that follow success are fleeting epitomizes the disorders associated with the success syndrome.

**Fig. 4.** Text compactor 25% summary

## 3 Experimental Result

### 3.1 Result of 10% Summary Generation

Table 1 shows 10% summary generation: For the 10% compaction of the blogs, the heading is included, and the rest of the passages' first and second sentences have probably been included. Therefore, 10% of the data generated for the summary is very few and extremely significant (Fig. 3).

### 3.2 Result of 25% Summary Generation

Table 2 shows 25% Summary Generation: In the 25% compaction of the blogs, a summary is generated with the content, blog title, first to third sentences of all the passages, and some redundant words (Fig. 4).

### 3.3 Result of 40% Summary Generation

Table 3 shows 40% summary generation. The 40% summary generation procedure involved the title of blogs and probably the first to fifth sentences of motivational blogs.

**Table 2.** 25% Summary Generation

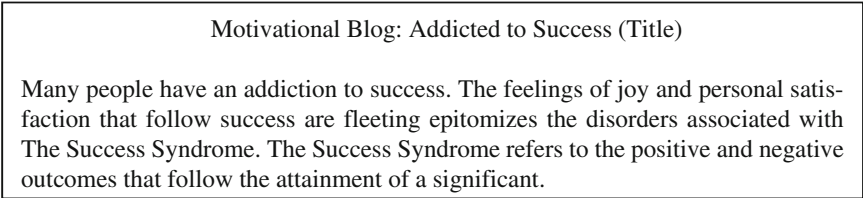| Motivational Blog | Words | Generated Summary |
|---|---|---|
| Addicted To Success | 258 | Title $1^{st}$ line of 1, 3 and $4^{th}$ passage. |
| Deep Existence | 230 | Title, 1st Line of 1, 2nd passage |
| Tiny Buddha | 268 | 2nd sentence of 2nd paragraph, 1st sentence of 3, 4, 5th Para graph. |
| Fearless Motivation | 568 | 2nd sentence in 4rd passage, 1st sentence of 4th, 5th, 6th, 7th, 8th, 9th passage. |
| Successes Consciousness | 224 | Heading, 1st sentence of 1st, 3rd passage, 2nd sentence of 2nd passage. |

---

Motivational Blog: Addicted to Success (Title)

Many people have an addiction to success. The feelings of joy and personal satisfaction that follow success are fleeting epitomizes the disorders associated with The Success Syndrome. The Success Syndrome refers to the positive and negative outcomes that follow the attainment of a significant.

---

**Fig. 5.** Text compactor 40% summary

**Table 3.** 40% Summary Generation

| Motivational Blog | Words | Generated Summary |
|---|---|---|
| Addicted To Success | 258 | Title $1^{st}$ and $2^{nd}$ sentences of $1^{st}$ paragraph, $1^{st}$ sentence of $3^{rd}$, $4^{th}$ paragraph. |
| Deep Existence | 230 | Title, $1^{st}$ Line of $1^{st}$ and $2^{nd}$ paragraph |
| Tiny Buddha | 268 | $2^{nd}$ sentence of $2^{nd}$ paragraph, $1^{st}$ sentence of $3^{rd}$, $4^{th}$ and $5^{th}$ paragraph. |
| Fearless Motivation | 568 | $2^{nd}$ and $3^{rd}$ sentence of $3^{rd}$ paragraph, $1^{st}$ and $2^{nd}$ sentence of $4^{th}$ paragraph, $1^{st}$ and $3^{rd}$ sentence of $5^{th}$ para graph, $1^{st}$ and $2^{nd}$ sentence of $7^{th}$ paragraph. $1^{st}$ sentence of $8^{th}$ and $9^{th}$ paragraph sentence of $6^{th}$ and $2^{nd}$. |
| Successes Consciousness | 224 | Title, $1^{st}$ sentence of $1^{st}$ and $2^{nd}$ sentences of $2^{nd}$ paragraph, $1^{st}$ sentence of $3^{rd}$ & $6^{th}$ paragraph. |

In the 40% of text that is compressed, it contains some relevant data and sentences, which define significant words as information (Fig. 5).

**Table 4.** Summary Match at 10%

| Motivational Blog | Similarity in Percentage |
|---|---|
| Addicted To Success | 14.07% |
| Deep Existence | 24.26% |
| Tiny Buddha | 52.62% |
| Fearless Motivation | 17.64% |
| Successes Consciousness | 14.92% |

**Table 5.** Summary Match at 25%

| Motivational Blog | Similarity in Percentage |
|---|---|
| Addicted To Success | 42.22% |
| Deep Existence | 52.40% |
| Tiny Buddha | 56.48% |
| Fearless Motivation | 28.36% |
| Successes Consciousness | 34.16% |

## 4  Result Analysis

After analyzing all 10%, 25%, 40%, and 80% results the observed information has been combined for the summary. It is observed that 80% compression is as good as the original one, so we have chosen on end line. This is recognized by our proposed system. Through we have conducted research simultaneously, then also we divided it into 50% stages, but it is observed that the outcomes match 40% and that's why the compression is not considered at the end of the analysis. We have used following Eq. 1 for calculating the matching compression ratio at different compression stages.

$$\text{Matching Percentage } = \text{ (Matched Words/Length of the sentences)} * 100 \quad (1)$$

$$\text{Error } = \text{ } 100 - \text{Average Matching Percentage} \quad (2)$$

The summary is matched at 10% compare with original sources as shown in Table 4.

Error at 10% Compression $= 100 - 24.26 = 75.74\%$

The summary is matched at 25% as compare with original sources as shown in Table 5.

Error at 25% Compression $= 100 - 40.78 = 59.22\%$

The summary is matched at 40% as compare with original sources in following Table 6.

Error at 40% Compression $= 100 - 45.50 = 54.50\%$

**Table 6.**  Summary Match at 40%

| Motivational Blog | Similarity in Percentage |
| --- | --- |
| Addicted To Success | 41.29% |
| Deep Existence | 50.46% |
| Tiny Buddha | 54.68% |
| Fearless Motivation | 40.08% |
| Successes Consciousness | 44.20% |

The major lacunas in the proposed system are that the length of blog must be 468 words or 48 sentence, after compression the most important perceptions are seen the outcome where the scope for improvement is still there.

## 5   Future Work

After looking at every 10%, 25%, 40%, and 80%, the information has been combined for the summary. It is observed that 80% compression is as good as the original one. As per our experimental analysis, we observed that even when the motivation blogs are proceeding for compaction, some huge blog sentences are not involved in the final summary. Therefore, the outcomes of the summary are incomplete due to the absence of some essential sentences. This is the challenge of text compression. So we will plan for the task of generating the summary from large sources by looking for other compaction tools and techniques.

## 6   Conclusion

Manuscript summarization is an emergent field that belongs to NLP as the stipulate for compressive consequential data. We completed the task on motivational blogs in this research paper. Following an analysis of various motivational blogs at 10%, 20%, 25%, 40%, and 80% compaction stages using a text compactor, four production rules for constructing the consequential summary of a motivational blog by collecting the entity of blogs from various compaction steps were defined. The overall result was beneficial and encouraging, as illustrated in the outline. There are still, some gaps have been identified, and because this is only appropriate for short blogs, we must investigate the potential outcomes for large blog summarization using different methodologies.

## References

1. Nenkova, Ani, and Kathleen McKeown. "A survey of text summarization techniques." In Mining text data, pp. 43–76. Springer, Boston, MA, 2012.
2. Tas, Oguzhan, and Farzad Kiyani. "A survey automatic text summarization." PressAcademia Procedia 5, no. 1 (2007): 205-213.

3. Allahyari, Mehdi, Seyedamin Pouriyeh, Mehdi Assefi, Saeid Safaei, Elizabeth D. Trippe, Juan B. Gutierrez, and Krys Kochut. "Text summarization techniques: a brief survey." arXiv preprint arXiv:1707.02268 (2017).
4. Modi, Shivangi, and Rachana Oza. "Review on Abstractive Text Summarization Techniques (ATST) for single and multi documents." In 2018 International Conference on Computing, Power and Communication Technologies (GUCON), pp. 1173–1176. IEEE, 2018.
5. Indu, M., and K. V. Kavitha. "Review on text summarization evaluation methods." In 2016 International Conference on Research Advances in Integrated Navigation Systems (RAINS), pp. 1–4. IEEE, 2016.
6. Masum, Abu Kaisar Mohammad, Sheikh Abujar, Md Ashraful Islam Talukder, AKM Shahariar Azad Rabby, and Syed Akhter Hossain. "Abstractive method of text summarization with sequence to sequence RNNs." In 2019 10th international conference on computing, communication and networking technologies (ICCCNT), pp. 1–5. IEEE, 2019.
7. Moratanch, N., and S. Chitrakala. "A survey on extractive text summarization." In 2017 international conference on computer, communication and signal processing (ICCCSP), pp. 1–6. IEEE, 2017.
8. El-Kassas, Wafaa S., Cherif R. Salama, Ahmed A. Rafea, and Hoda K. Mohamed. "Automatic text summarization: A comprehensive survey." Expert Systems with Applications 165 (2021): 113679.
9. Widyassari, Adhika Pramita, Supriadi Rustad, Guruh Fajar Shidik, Edi Noersasongko, Abdul Syukur, and Affandy Affandy. "Review of automatic text summarization techniques & methods." Journal of King Saud University-Computer and Information Sciences (2020).
10. Bhatia, Neelima, and Arunima Jaiswal. "Automatic text summarization and it's methods- a review." In 2016 6th International Conference-Cloud System and Big Data Engineering (Confluence), pp. 65–72. IEEE, 2016.
11. Liao, Chechen, Pui-Lai To, and Chuang-Chun Liu. "A motivational model of blog usage." Online Information Review (2013).