

# INTERNSHIP REPORT

**ZNZ Systems Pvt. Ltd**

**Submitted By**

**Omkar Shivaji Motale**

**T184061.**

**0120160208.**

##### DEPARTMENT OF COMPUTER ENGINEERING

MIT ACADEMY OF ENGINEERING

ALANDI (D), PUNE



##### DEPARTMENT OF COMPUTER ENGINEERING

*CERTIFICATE*

This is to certify that the “**Internship report”** submitted by **Mast. Omkar Shivaji Motale.** **PRN. No.: 0120160208 and Examination Seat No.: SCET184061** is work done by him/her and submittedduring 2018 – 2019 academic year.

**Mentor for Internship Department Internship Coordinator**

**(Mr. Rudragouda Patil) (Mrs. R. R. Badre)**

**Institute Internship Coordinator Head of the Department**

**(Mrs. R. R. Badre) (Mrs. R. R. Badre)**

### 

### Internship certificate provided by the internship institution

###### ACKNOWLEDGEMENT

I, Mr. Omkar Shivaji Motale the student of Computer Engineering Department at MIT Academy of Engineering Pune (An Autonomous Institute), has been completed Summer Internship in “ZNZ Systems Pvt. Ltd.”.

I take this opportunity to record my profound gratitude to , Mr. Niranjan Patil Mr. Mehul Patil ,Miss Snehal Gawade cointerns ,ZnZ Systems Pvt. Ltd, for their,valuable advices, constant encouragement throughout my internship.

I express my deep sense of gratitude to Mr. Chaitanya Patil sir Director, ZNZ Systems Pvt. Ltd, for his continuous inspiration and encouragement.

Finally, I would like to acknowledge and express my special thanks to Mr. Rudragauda patil Assistant Professor, Computer Engineering Department, MIT Academy of Engineering Pune (An Autonomous Institute), my family and other employees for their patience, encouragement, support they have made during the period of this internship.

**(Omkar Shivaji Motale)**

## Table of Contents

|  |  |  |  |
| --- | --- | --- | --- |
| **Topic** | | | **Page No.** |
| 1. | **Introduction** | | 06. |
|  | 1.1 | About the organization. | 06. |
|  | 1.2 | Scope of the work. | 09. |
| 2. | **Internship Discussion** | | 13. |
|  | 2.1 | Learning Experience. | 13. |
|  | 2.2 | Most challenging task performed. | 18. |
|  |  |  |  |
| 3. | **Conclusion** | | 21. |
| 4. | **Bibliography** | | 23. |

**1) Introduction.**

1.1 About the Organization.

### Specialties:

Product Engineering Services, Web Development Services, Machine learning projects .

**Company Type:** Privately Held.

**Size:** 11-50 employees.

**Website:** <https://www.missionworldpeace.com/>

[www.ChaitanyaPatil.in/vm](http://www.ChaitanyaPatil.in/vm)

Znz System is Startup Company in 2017 with a team of highly professional skilled people team having huge experience in IT industry with excellent knowledge in Product Development, providing Professional Services across various domain and industries.

The company primarily provides professional services in the niche skills areas of development of world peace and interesting projects in the field of machine learning for automation.

**Mission:**

* We Come Together (Digitally). We Think. We Discuss. We Share & Improve Ideas. We Work Smartly.
* **Our Theme is:** Mission World Peace.

**Services:**

ZnZ Systems has an array of services onboard from which you are free to take your pick.

1. Mobile App Development: Mobile Apps are loved for two things - Striking Layout and Responsive functionality. They provide the best of both with adaptability to any screen size.
2. Web Development: They believe in creating the simple, creative and flexible Design websites right from scratch. Services of company are adept enough to meet the toughest challenges to achieve optimal client satisfaction.

**Working:**

ZnZ System teaches effective plan about working process. The plan is that “To bring your idea alive online, it is required to pass through a sequence of intricate processes. These processes ensure that your idea reaches their deserved platform in top form”.

Plan has different stages:

1. Planning: A successful project is built upon the minutest details. Knowing them and their requirements provides the basis for a clear plan of action.
2. Structure: Structure is the foundation of any product or process that’s online. Be it mobile apps, website development or digital marketing. You name it, we achieve it!
3. Design: A good design possesses an attractive look that appeals to your target audience. While a visual design is something we excel at, we are just as good in designing strategies.
4. Development: We are the bridge that transitions your idea into a successful project. To ensure your idea achieves perfection in its digital form, the latest technologies and designs are brought onboard to create your perfect project.
5. Digital: Once everything needed is in place, it’s time to get your business online. That’s when our digital marketing wizards get at work with the assurance of securing the best results on search engines.
6. Deployment: Your website is ready just like you envisioned it to be. Now, we upload it onto servers to get it live for the world to use.

**The period of Internship: Two Months.**

**Starting Date:** 04 June 2019.

**Ending Date:** 28 July 2019.

1.2 Scope of the work.

**Purpose of Project:**

We all interact with an application which uses text summarization. Many of those applications are for the platform which publishes articles on daily news, entertainment, sports. With our busy schedule, we prefer to read the summary of that article before we decide to jump in for reading entire article. Reading a summary help us to identify the interest area, gives a brief context of the story. Summarization can be defined as a task of producing a concise and fluent summary while preserving key information and overall meaning.

**Objective of Project:**

Summarizing the contents of a text file.

**Functional Requirements:**

1. Accept the file as the input for summarizing.

2. Pass it through the code and obtain a summary of the contents of text file.

**2) Internship Discussion.**

2.1 Learning Experience.

1. Knowledge acquired:

**How text summarization works?**

In general there are two types of summarization, abstractive and extractive summarization.

1. Abstractive Summarization:

Abstractive methods select words based on semantic understanding; even those words did not appear in the source documents. It aims at producing important material in a new way. They interpret and examine the text using advanced natural language techniques in order to generate a new shorter text that conveys the most critical information from the original text.

It can be correlated to the way human reads a text article or blog post and then summarizes in their own word.

**Input document → understand context → semantics → create own summary.**

1. **Extractive Summarization:**

Extractive methods attempt to summarize articles by selecting a subset of words that retain the most important points. This approach weights the important part of sentences and uses the same to form the summary. Different algorithm and techniques are used to define weights for the sentences and further rank them based on importance and similarity among each other.

**Input document → sentences similarity → weight sentences → select sentences with higher rank.**

**EXTRACTIVE SUMMARIZATION METHODS:**

A. Term Frequency-Inverse Document Frequency (TF-IDF) method

B. Cluster based method

C. Graph theoretic approach

D. Machine Learning approach

E. Text summarization with neural networks

F. Automatic text summarization based on fuzzy logic.

Code flow:

Input article → split into sentences → remove stop words → build a similarity matrix → generate rank based on matrix → pick top N sentences for summary.

Text Rank Algorithm:

Before getting started with the Text Rank algorithm, there’s another algorithm which we should become familiar with – the PageRank algorithm. In fact, this actually inspired Text Rank! PageRank is used primarily for ranking web pages in online search results. Let’s quickly understand the basics of this algorithm with the help of an example.

PageRank Algorithm



Suppose we have 4 web pages — w1, w2, w3, and w4. These pages contain links pointing to one another. Some pages might have no link – these are called dangling pages.



Web page w1 has links directing to w2 and w4

 w2 has links for w3 and w1

 w4 has links only for the web page w1

 w3 has no links and hence it will be called a dangling page

In order to rank these pages, we would have to compute a score called the **PageRank score**. This score is the probability of a user visiting that page.

To capture the probabilities of users navigating from one page to another, we will create a square **matrix M**, having n rows and n columns, where **n** is the number of web pages.



Each element of this matrix denotes the probability of a user transitioning from one web page to another. For example, the highlighted cell below contains the probability of transition from w1 to w2.



The initialization of the probabilities is explained in the steps below:

1. Probability of going from page i to j, i.e., M[ i ][ j ], is initialized with **1/(number of unique links in web page wi)**

2. If there is no link between the page i and j, then the probability will be initialized with **0**

3. If a user has landed on a dangling page, then it is assumed that he is equally likely to transition to any page. Hence, M[ i ][ j ] will be initialized with **1/(number of web pages)**

Hence, in our case, the matrix M will be initialized as follows:



Finally, the values in this matrix will be updated in an iterative fashion to arrive at the web page rankings.

Let’s understand the TextRank algorithm, now that we have a grasp on PageRank. I have listed the similarities between these two algorithms below:

 In place of web pages, we use sentences

 Similarity between any two sentences is used as an equivalent to the web page transition probability

 The similarity scores are stored in a square matrix, similar to the matrix M used for PageRank

**TextRank is an extractive and unsupervised text summarization technique.** Let’s take a look at the flow of the TextRank algorithm that we will be following:



The first step would be to concatenate all the text contained in the articles

 Then split the text into individual sentences

 In the next step, we will find vector representation (word embeddings) for each and every sentence

 Similarities between sentence vectors are then calculated and stored in a matrix

 The similarity matrix is then converted into a graph, with sentences as vertices and similarity scores as edges, for sentence rank calculation

 Finally, a certain number of top-ranked sentences form the final summary.

* 1. Most Challenging task performed

The limited study is available for abstractive summarization as it requires a deeper understanding of the text as compared to the extractive approach.

Purely extractive summaries often times give better results compared to automatic abstractive summaries. This is because of the fact that abstractive summarization methods cope with problems such as semantic representation, inference and natural language generation which is relatively harder than data-driven approaches such as sentence extraction.

For Abstractive Text summarization we will be building a document or text summarization graphical user interface app with tkinter and spacy, sumy, gensim and nltk.

# Summaryzer\_GUI : simply text/ document summarization gui with python tkinter

* Using Spacy,Sumy,Gensim and NLTK
* Fetch text from url and summarize
* Summarize files as well as links
* Compare various summarizers

Spacy summarization:

* Text Preprocessing (remove stopwords,punctuation).
* Frequency table of words/Word Frequency Distribution – how many times each word appears in the document
* Score each sentence depending on the words it contains and the frequency table
* Build summary by joining every sentence above a certain score limit

Gensim summarization:

This module implements TextRank, an unsupervised algorithm based on weighted-graphs from a [paper by Mihalcea et al](https://web.eecs.umich.edu/~mihalcea/papers/mihalcea.emnlp04.pdf). It is built on top of the popular [PageRank](https://en.wikipedia.org/wiki/PageRank) algorithm that Google used for ranking webpages. TextRank works as follows:

* Pre-process the text: remove stop words and stem the remaining words.
* Create a graph where vertices are sentences.
* Connect every sentence to every other sentence by an edge. The weight of the edge is how similar the two sentences are.
* Run the PageRank algorithm on the graph.
* Pick the vertices(sentences) with the highest PageRank score

# Prerequisites:

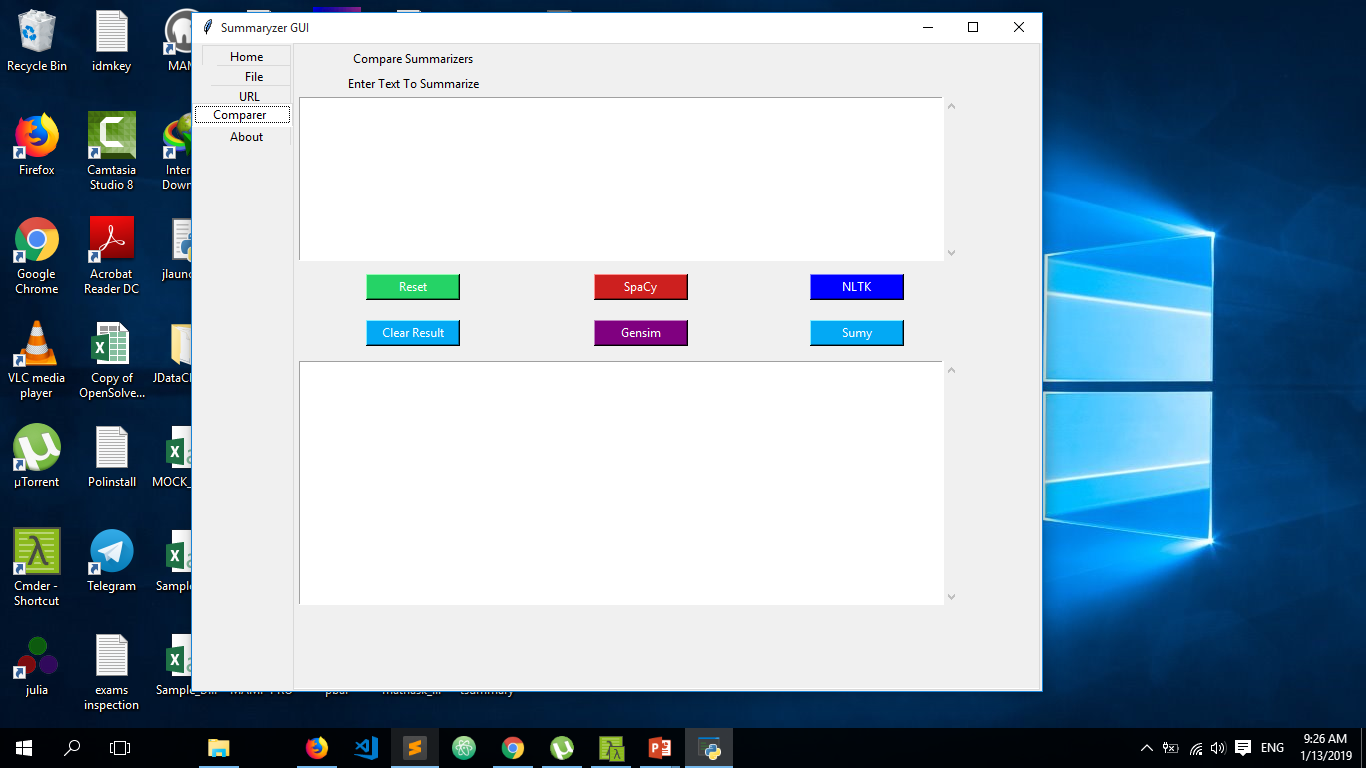
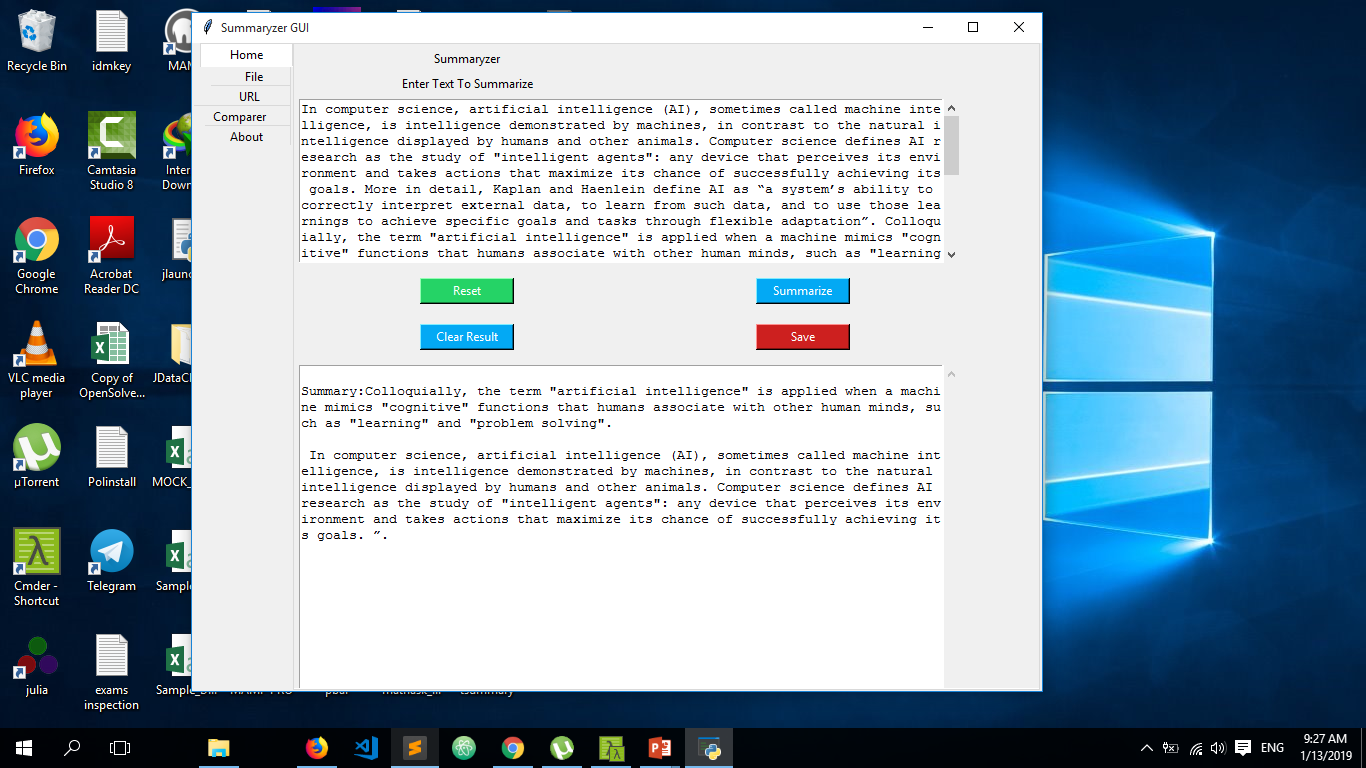
1. Python 3
2. Spacy
3. NLTK
4. Gensim
5. Sumy

# Python GUI – tkinter

Python offers multiple options for developing GUI (Graphical User Interface). Out of all the GUI methods, tkinter is most commonly used method. It is a standard Python interface to the Tk GUI toolkit shipped with Python. Python with tkinter outputs the fastest and easiest way to create the GUI applications. Creating a GUI using tkinter is an easy task.

**To create a tkinter:**

1. Importing the module – tkinter
2. Create the main window (container)
3. Add any number of widgets to the main window
4. Apply the event Trigger on the widgets.

GUI:

**3) Conclusion:**

We have developed an application which can summarize text from the given text file also from the inserted text and also from the given url.This automated text summarization will efficiently summarize the paragraph. This application is to be used in answer sheet evaluation where summary of answers are to be crosschecked,. Hence it can efficiently provide summary.

### Applications of Text Summarization

* + 1. News: There are multiple applications of this technique in the field of News. It includes creating an introduction, Generating headlines, Embedding captions on pictures.  
         
       2. Scientific Research: Algorithms are used to dig out important information from Scientific research papers. AI is outranking human beings in doing so.  
         
       3. Social Media Posting: Content on Social media is preferred to be concise. Companies use this technique to convert long blog articles into shorter ones suited for the audience.   
         
       4. Creating Study Notes: Many applications use this process to create student notes from vast syllabus and content.  
         
       5. Conversation Summary: Long conversations and meeting recording could be first converted into text and then important information could be fetched out of them.  
         
       6. Movie Plots and Reviews: The whole movie plot could be converted into bullet points through this process.  
         
       7. Deliverable Feeds: They are the short piece of information derived from the complete informative articles. These are generally delivered to people through emails or feed delivery services.   
         
       8. Content Writing: Not from the scratch though but on providing a topic and points an outlined summary could be generated.

**Future Scope:**

The rate at which the information is growing is tremendous. Hence it is very important to build a multilingual summarization system and this research could be a stepping stone towards achieving that goal provided there is availability of online lexical databases in other languages. The work presented by the thesis can also be applicable to multi document summarization by using minimal extensions.