

A PROJECT
ON

Scribe - Natural Language Understanding and Generation Engine

BY

Vandita Shah B-827
Sumer Shende B-834
Vineet Trivedi B-848
Rishabh Vig B-855

Under the guidance of

Internal Guide
Prof. Dnyaneshwar Dhangar

MCT
MANJARA CHARITABLE TRUST
RAJIV GANDHI INSTITUTE OF TECHNOLOGY
Juhu-Versova Link Road Versova, Andheri(W), Mumbai-53

Department of Computer Engineering

University of Mumbai

April – 2016



Juhu-Versova Link Road Versova, Andheri(W), Mumbai-53

Department of Computer Engineering

CERTIFICATE

This is to certify that

1. Vandita Shah B-827
2. Sumer Shende B-834
3. Vineet Trivedi B-848
4. Rishabh Vig B-855

Have satisfactory completed this project entitled

“Scribe - Natural Language Understanding and Generation Engine”

Towards the partial fulfillment of the

**BACHELOR OF
ENGINEERING IN
(COMPUTER ENGINEERING)**

as laid by University of Mumbai.

Guide
Prof. Dnyaneshwar Dhangar

H.O.D.
Dr. S. B. Wankhade

Principal
Dr. Udhav Bhosle

Internal Examiner

External Examiner

Project Report Approval for B. E.

This project report entitled “Scribe - Natural Language Understanding and Generation Engine” by Vandita Shah, Sumer Shende, Vineet Trivedi and Rishabh Vig is approved for the degree of ***Bachelor of Computer Engineering***.

External Examiners

1.-----

2.-----

Internal Examiner

1.-----

2.-----

HOD

Date:

Place:

DECLARATION

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Date:

ACKNOWLEDGEMENT

We wish to express our sincere gratitude to Dr. U. V. Bhosle, Principal and Dr. S. B. Wankhade, H.O.D of Computer Department of RGIT for providing us an opportunity to do our project work on “Scribe - Natural Language Understanding and Generation Engine”.

This project bears on imprint of many peoples. We sincerely thank our project guide Prof. Dnyaneshwar Dhangar for his/her guidance and encouragement in carrying out this synopsis work.

Finally, we would like to thank our colleagues and friends who helped us in completing the Project work successfully.

ABSTRACT

“SCRIBE” is an AI driven system which automatically generates output in a user defined format based on raw data that is fed to it. It is a tedious and progressively difficult task for a human to process large volumes of data and deliver meaningful output in a scenario where things are changing by the minute. Thus “SCRIBE” reorganizes raw data into furnished output as required by the user whether it be in the form of an essay, article, post or letter. “SCRIBE” collects raw data and stores it in a file system. It applies adequate business rules for data cleaning, standardization and mining.

The polished data is then formatted as per user requirement using narrative science. In essence “SCRIBE” approaches the highest epoch of artificial intelligence to automate creative human tasks like writing.

TABLE OF CONTENTS

NAME	PAGE NO.
Title Page	i
Certificate	ii
Project Report Approval	iii
Declaration	iv
Acknowledgement	v
Abstract	vi
Table of Contents	vii
List of Figures	viii
Chapter 1) Introduction	1
Chapter 2) Problem Statement	3
Chapter 3) Literature Survey	7
Chapter 4) Requirement Analysis	11
Chapter 5) Project Design	12
Chapter 6) Implementation Details	17
Chapter 7) Technologies used	24
Chapter 8) Test Cases	26
Chapter 9) Project Timeline	29
Chapter 10) Task Distribution	30
Chapter 11) Conclusion and Future Work	32
Chapter 12) Reference	33
Chapter 13) Appendix	35

TABLE OF FIGURES

NAME OF FIGURES	PAGE NOS
Figure 2.1 Block Diagram	6
Figure 5.1 Overall Design	12
Figure 5.2 Breakdown of sentences	12
Figure 5.3 Architecture of system	14
Figure 6.1 Flowchart	19
Figure 6.2 Preprocessing output	20
Figure 6.3 Cleaning phase 1 output	20
Figure 6.4 Cleaning phase 2 output	21
Figure 6.5 Exporting to database	21
Figure 6.6 Checking the imported data	22
Figure 6.7 Checking for diversity of data	22
Figure 6.8 Querying lemmas in the database	23
Figure 6.9 Production of sentences	23
Figure 9.1 Timeline for implementation part 1	29
Figure 9.2 Timeline for implementation part 2	29