

*Major Project Report on*

# **POPULARITY ANALYSIS ON SOCIAL NETWORK**

By

**Alphonsius Sungoh (BT/IT/1002)**

**Brandon Victor Syiem (BT/IT/1013)**

**Moidul Islam (BT/IT/1035)**

**Pranjal Deka (BT/IT/1041)**

**Sailen Kalita (BT/IT/1048)**

*In partial fulfillment of requirements for the award of degree in*  
**Bachelor of Technology in Information Technology**  
**(2010-14)**

*Under the Supervision of*  
**Dr. Hemanta Kalita (Associate Prof., Deptt. Of IT)**  
**Mr. Sufal Das (Assistant Prof., Deptt. Of IT)**



**Department of Information Technology**  
**NORTH EASTERN HILL UNIVERSITY**  
**UMSHING, SHILLONG, MEGHALAYA - 793022**



सूचना प्रौद्योगिकी विभाग  
पूर्वोत्तर पर्वतीय विश्वविद्यालय  
उमशिगं, शिलांग-७९३०२२.

**Department of Information Technology**  
NORTH-EASTERN HILL UNIVERSITY

UMSHING, SHILLONG - 793 022.

---

**Date:**

### **To whom it may concern**

This is to certify that Alphonsius Sungoh (BT/IT-1002), Brandon Victor Syiem (BT/IT-1013), Moidul Islam (BT/IT-1035), Pranjal Deka (BT/IT-1041) and Sailen Kalita (BT/IT-1048) worked in the project “POPULARITY ANALYSIS ON SOCIAL NETWORK” from 18<sup>th</sup> February 2014 to 30<sup>th</sup> July 2014 and has successfully completed the major project, in order to partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Information Technology under our supervision and guidance.

Mr. Sufal Das (*Assistant Prof.*)

Dr. Hemata kr. Kalita (Associate prof.)

Department of Information Technology

North Eastern Hill University

Umshing Shillong-793022



सूचना प्रौद्योगिकी विभाग  
पूर्वोत्तर पर्वतीय विश्वविद्यालय  
उमशिगं, शिलांग-७९३०२२.

**Department of Information Technology**  
**NORTH-EASTERN HILL UNIVERSITY**

UMSHING, SHILLONG – 793 022.

Phone No.: +91 364-2723611

Mobile No: +91 9436337792

Email: [hodit@nehu.ac.in](mailto:hodit@nehu.ac.in)

---

No:

Date:

### **To whom it may concern**

This is to certify that Alphonsius Sungoh (BT/IT-1002), Brandon Victor Syiem (BT/IT-1013), Moidul Islam (BT/IT-1035), Pranjal Deka (BT/IT-1041) and Sailen Kalita (BT/IT-1048) worked in the project “POPULARITY ANALYSIS ON SOCIAL NETWORK” from 18<sup>th</sup> February 2014 to 30<sup>th</sup> July 2014 and has successfully completed the major project, in order to partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Information Technology.

External Examiner

Head

Department of Information Technology

North Eastern Hill University

Umshing, Shillong-793022

## **ACKNOWLEDGEMENT**

If words are considered as a symbol of approval and tokens of acknowledgement, then let the words play the heralding role of expressing our gratitude to all those who has helped us directly or indirectly during our project.

First and foremost, we wish to express our deep and sincere feeling of gratitude to our project guide Mr.Sufal Das, Asst. Professor, Department of Information Technology, School of Technology, NEHU for giving us valuable information, providing the required support, relentless guidance at every stage of this project.

We are greatly indebted to Dr. H. K. Kalita, Head of Department of Information Technology, School of Technology, NEHU, for giving us constant guidance, co-operation and for providing us with the required infrastructure to help us complete our project work.

We are also greatly thankful to all the teachers, staff and students of Information Technology Department for their constructional comments and helpful suggestions.

Alphonsius Sungoh  
Brandon Victor Syiem  
Moidu Islam  
Pranjal Deka  
Sailen Kalita

## **ABSTRACT**

Online social networking has become extremely popular with adolescents and young adults. Ninety-three percent of college students have a Facebook account, Whereas almost all VIP persons all over the world having a Twitter account. Some other popular social networking sites are You tube, Flickr etc. The users of these sites form a social network, which provides a powerful means of sharing, organizing, and find content and contacts. The popularity of these sites provides an opportunity to study the characteristics of online social network graphs at large scale.

To analyze the large set of data the conventional methods of data analyze is not efficient, so in our project we have used the hadoop data analyzing tool. Hadoop uses distributed file system to store data. The main point of use of this hadoop tool is, it brings processing to the data, not data to processing unit.

We have used perl which is an interpreted language optimized for scanning arbitrary text files, extracting information from these files, and printing reports based on that information to create a scraper.

A scraper (or “screen-scraper”) extracts the information you want whatever you consider to be data from a given webpage. We have used scraper to collect different data from different sites of people of interest or group.

Then we have used hadoop (a distributed file system to store data) to distribute the processing of data to the clusters of computers. Hadoop uses three techniques- mapping, reducing and driver.

<b>CONTENT</b>	<b>PAGE NO</b>
CHAPTER 1-----	1
INTRODUCTION-----	1
BACKGROUD STUDY-----	3
USERS-----	3
BIG DATA -----	6
BIG DATA CONCEPT-----	6
BIG DATA CHARACTERITICS-----	6
IMPORTANCE OF BIG DATA-----	8
BIG DATA CHALLENGES-----	8
BIG DATA ANALYTICS-----	9
BIG DATA ANALYTICS SOFTWARE-----	9
PERL-----	10
SCRAPER-----	11
SCRAPER TECHNIQUES-----	12
CHAPTER 2-----	15
FORMAL DEFINITION OF THE PROBLEM-----	15
PROPOSED APPROACH-----	17
PERL-----	17
SCRAPER-----	18
HADOOP-----	18

CHAPTER 3-----	22
STEP-----	22
CHAPTER 4- -----	23
IMPLEMENTATION DETAILS-----	23
PERL -----	23
HADOOP-----	23
MAPPER-----	25
REDUCER-----	26
DRIVER-----	27
CHAPTER 5-----	28
RESULT ANALYSIS-----	28
CHAPTER 6-----	35
CONCLUSION-----	35
LIMITATIONS OF THE PROJECT-----	35
REFERENCE-----	37
PERL-----	37
HADOOP-----	37
PAPERS ON HADOOP-----	38