DECLARATION

I hereby declare that the thesis entitled "**SMS based FAQ Retrieval using Hybrid Similarity Measure"** which is being submitted to the **Delhi Technological University**, in partial fulfillment of the requirements for the award of degree of **Master of Technology in Computer Science Engineering** is an authentic work carried out by me. The material contained in this thesis has not been submitted to any university or institution for the award of any degree.

Sonal Meena

Department of Computer Engineering

Delhi Technological University,

Delhi.

CERTIFICATE

DELHI TECHNOLOGICAL UNIVERSITY

(Govt. of National Capital Territory of Delhi)

BAWANA ROAD, DELHI-110042

Date: _____

This is to certify that the thesis entitled "**SMS based FAQ Retrieval using Similarity Measure"** submitted by **Sonal Meena (Roll Number: 2K11/CSE/15),** in partial fulfillment of the requirements for the award of degree of Master of Technology in Computer Science Engineering, is an authentic work carried out by her under my guidance. The content embodied in this thesis has not been submitted by her earlier to any institution or organization for any degree or diploma to the best of my knowledge and belief.

Project Guide

Mr. Manoj Kumar

Associate Professor

Department of Computer Engineering

Delhi Technological University, Delhi-110042

ACKNOWLEDGEMENT

I take this opportunity to express my deepest gratitude and appreciation to all those who have helped me directly or indirectly towards the successful completion of this thesis.

Foremost, I would like to express my sincere gratitude to my guide **Mr. Manoj Kumar**, **Associate Professor, Department of Computer Engineering**, **Delhi Technological University, Delhi** whose benevolent guidance, constant support, encouragement and valuable suggestions throughout the course of my work helped me successfully complete this thesis. Without his continuous support and interest, this thesis would not have been the same as presented here.

Besides my guide, I would like to thank the entire teaching and non-teaching staff in the Department of Computer Science, DTU for all their help during my course of work.

SONAL MEENA

ABSTRACT

Mobile technology gave contribution to the progress of media of communication for example: chats, emails and short message services (SMS). The Popularity, utility and simplicity of SMSes is encouraging people to access information via SMSes, accessing information via internet creates hassle, it's not necessary that internet connection is always available. So user can clarify their query, make complaint and get updates of result etc by sending SMS. Accessing information in such a manner makes information access very economic and easy for everybody from rural to metro city people.

"FAQ retrieval" means there is corpora of frequently asked questions, and user sends a query in SMS language to retrieve some information. Such systems finds best match from FAQ corpora for given user defined query written in SMS language. The main problem in SMS language is the noise associated with it. Spelling mistakes, transliteration, phonetic spellings, abbreviations and short forms create difficulties in string matching.

In proposed work, a novel approach has been presented by developing Hybrid similarities which evaluates similarity scores with the questions in the corpus for SMS query. In this way, we can further improve the accuracy of the SMS based FAQ system significantly by refining the results of the system using different hybrid similarity scores.

Table of Contents

List of Figures

List of Tables

List of Equations