JSS Mahavidyapeetha

Sri Jayachamarajendra College of Engineering, Mysore – 570 006

An Autonomous Institute Affiliated to

Visvesvaraya Technological University, Belgaum

"CLASSIFICATION OF TEXTS – A THRESHOLD BASED APPROACH"

Report submitted as partial fulfilment of curriculum prescribed for the award of the degree of

BACHELOR OF ENGINEERING IN COMPUTER SCIENCE & ENGINEERING

Team members**:**

Meghana P C (USN No: 4JCO8CS048)

Pratheeksha D (USN No: 4JCO8CS069)

Shruthi Sudhanva (USN No: 4JCO8CS095)

Under the guidance of the faculty

Sri. Anilkumar K.M

Assistant Professor

Department of Computer science and Engineering,

Sri Jayachamarajendra College of Engineering,

Mysore

JSS Mahavidyapeetha

Sri Jayachamarajendra College of Engineering, Mysore – 570 006

An Autonomous Institute Affiliated to

Visvesvaraya Technological University, Belgaum

CERTIFICATE

This is to certify that the work entitled **"Classification of texts: A threshold based approach"** is a bonafide work carried out by **Meghana P C(4JC08CS048), Pratheeksha D(4JC08CS069), Shruthi Sudhanva (4JC08CS095)** in the partial fulfilment of the degree of **Bachelor of Engineering** in **Computer Science and Engineering at SJCE, affiliated to Visvesvaraya Technological University, Belgaum during the year 2012**. It is certified that all the corrections / suggestions have been incorporated in the report. The project report has been approved as it satisfies the academic requirements in respect of project work prescribed for the Bachelor of Engineering degree.

Internal Guide
Mr AnilKumar K M
Assistant Professor
Dept of Computer Science and Engg.
SJCE, Mysore

Dr C N Ravikumar
Head Of The Department
Dept of Computer Science and Engg.
SJCE, Mysore

Examiners:

1. **Name:** ……. **Signature: …….**
2. **Name: …….** **Signature: …….**
3. **Name: …….** **Signature: …….**

DECLARATION

We, **Meghana P C (4JC08CS048)** , **Pratheeksha D(4JC08CS069) , Shruthi Sudhanva(4JC08CS095)** students of the Department of Computer Science and Engineering from Sri Jayachamarajendra College of Engineering - Mysore, do hereby declare that this dissertation entitled **"Classification of texts: A threshold based approach"** has been independently carried out by us under the guidance of **Mr Anilkumar K M, Assistant Professor, Dept of Computer Science & Engineering, Sri Jayachamarajendra College Of Engineering, Mysore** in partial fulfilment of the requirements for the award of the degree of **Bachelor Of Engineering** in **Computer Science and Engineering, affiliated Visvesvaraya Technological University, Belgaum.**

We also declare that we have not submitted this dissertation to any other University for the award of any degree.

Meghana P C
Pratheeksha D
Shruthi Sudhanva

ABSTRACT

The rapid development of web and its related technologies have fuelled the popularity of the web with all sections of society. The web has been rightfully used by governments, business houses, industries, educational institutions etc., and individuals to make information available globally and also gain knowledge globally. The web is the source of many research activities and one interesting area is to mine user opinion from web on diverse topic. The study of opinions is useful for both producers and consumers of the topics. The producers can be manufacturers of digital products, automobile manufactures, movie producers, editor of news article etc., very much interested to find opinion of a user. The consumers are individual users who document their opinion and want to share it with others about the topic. Many readers of online reviews say that these reviews or opinions influence their purchasing decision. Today people of all ages and from all over the world use web for collecting opinions. There are many sites which allow users to express their opinion such as Epinions, Amazon, etc.,

The project 'Classification of texts : A threshold based approach' experiments different methods of classifying texts as opinion or factual. The experiments are carried out on collected data sets of opinion and factual text files. The project makes an inference about the best method from the results obtained. The inferred best method is used to classify user input text as factual or opinionated. The results of the experiments performed are visualized with the help of graphs. It also provides a front end for user interaction.

The first step in building opinion search engines or other applications is determining which documents are topically relevant to an opinion oriented query. And hence, determining which documents or portions of documents contain review like or opinionated material forms a significant part of opinion mining.

Our approach to subjectivity classification of texts classifies a document into factual or opinionated based on the polar word count, adjective count, count of Parts-Of-Speech patterns and presence of cues. We experiment these different techniques with varying input sizes and thresholds, calculating their corresponding accuracies. We then conclude with the optimal threshold and the best method. Our approach also tries to look at the behaviour of the classification methods considering the number of input files as a factor.

ACKNOWLEDGEMENT

Every successful completion of any undertaking would be complete only after we remember and thank the almighty, the parents, the teachers and all the personalities, who directly or indirectly helped and guided in execution of that work. The success of this work is equally attributed to all well-wishers who have encouraged and guided throughout the execution.

First and foremost we express our deepest gratitude to our internal guide **Mr Anilkumar K.M**, Assistant professor, Dept of Computer Science & Engineering, Sri Jayachamarajendra College Of Engineering, Mysore for his guidance and valuable suggestions during the course of this work.

We are thankful to our Head of the Department, **Dr C.N Ravikumar** for giving us the opportunity to undertake this project.

We are grateful to **Dr. B.G Sangameshwara, Principal SJCE, Mysore,** for all the facilities provided to us in the college for carrying out this project.

Meghana P C

Pratheeksha D

Shruthi Sudhanva

Table Of Contents

List Of Figures

List Of Tables