SEMANTIC TEXT CLUSTERING USING ENHANCED VECTOR SPACE MODEL USING NEPALI LANGUAGE

A Dissertation

Submitted to
The Central Department of Computer Science and Information Technology, Institute of Science and Technology
Tribhuvan University

In Partial Fulfillment of the Requirements for the degree of
Master of Science in Computer Science and Information Technology

By
Chiranjibi Sitaula
January 2012

Under the Supervision of
Prof. Dr. Shashidhar Ram Joshi
(IOE, TU)
Co-Supervisor
Mr. Bikash Balami

Date:……..

Recommendation

We hereby recommend that the dissertation prepared under my supervision by **Mr. Chiranjibi Sitaula** entitled "**Semantic Text Clustering using Enhanced Vector Space Model using Nepali Language**" be accepted as fulfilling in part requirements for the degree of Masters of Science. In my best knowledge this is an original work in computer science.

----------	----------
Dr. Shashidhar Ram Joshi	Mr. Bikash Balami
Professor	Lecturer
Institute of Engineering, Pulchowk, Nepal	Central Department of Computer Science and IT, T.U, Nepal
(Supervisor)	(Co-Supervisor)

Tribhuvan University

Institute of Science and Technology

Central Department of Computer Science and Information Technology

We certify that we have read this dissertation work and in our opinion it is satisfactory in the scope and quality as a dissertation in the partial fulfillment for the requirement of Master of Science in Computer Science and Information Technology.

Evaluation Committee

Dr. Tanka Dhamala
Asst.Professor
Head, Central Department of
Computer Science and Information
Technology
Tribhuvan University

(External Examiner)

Date: ----------

Dr. Shashi Dhar Ram Joshi
Professor
Institute of Engineering
Pulchowk, Nepal
(Supervisor)

(Internal Examiner)

Date: ----------

Acknowledgement

This dissertation has been prepared under **Central Department of Computer Science and Information Technology** (*Tribhuwan University*), Kirtipur and the study was done for the "Semantic Text Clustering using Enhanced Vector Space Model using Nepali Language". I am very grateful to our department for giving me an enthusiastic support. I deeply extends my heartily acknowledgement to my respected teachers and dissertation advisor **Prof. Dr. Shashidhar Ram Joshi**, Head of Department of Electronics and Computer Engineering, Institute of Engineering, Pulchowk, for his wholehearted cooperation, encouragement and string guidelines throughout the preparation of this study. With this regard, I wish to extend my sincere appreciation to respected Head of Department of Central Department of Computer Science and Information Technology, **Assoc. Prof. Dr. Tanka Dhamala** for his kind help.I am very grateful and thankful to all the respected teachers of CDCSIT, TU, for granting me broad knowledge and inspirations within the time period of two years.

My special thanks goes to my co-supervisor **Mr. Bikash Balami** who directly and indirectly extended his hands in making this thesis work a success. Similarly, I would like to thank to all of my friends of my class who helped me directly and indirectly to complete my thesis.

As we know that "there is an eraser at the top of the pencil", so mistakes are done by human beings of which we are also members. Although, I don't experience any enjoyment, since I did have a very hard time while completing this project, I have done my best. There may be some errors in our project, so any suggestion regarding the mistake of this project will be always welcomed.

Abstract

The vector space model is popular method for the clustering process while doing research in the field of text mining. The main reason of its popularity is its less computational overhead and simplicity. Classical vector space model can not be used for semantic analysis purpose because it simply uses syntactic model for clustering. In order to cluster the documents in sentence level, how individual keyword plays the important roles in the text clustering, is studied in this work. For this, an Enhanced method is proposed which cand easily outperform classical vector space model due to the involvement of fuzzy set approach. The classical Vector Space Model is enriched with fuzzy set so as to form the Enhanced Vector Space Model in text clustering.In order to give the semantic text clustering, fuzzy set plays crucial role in addition to classical Vector Space Model.

TABLE OF CONTENTS

Details **Page No.**

LIST OF TABLES

TABLES Contents Page

LIST OF FIGURES

FIGURES **Contents** **Page**