

Editors
Dr. Siti Zaiton Mohd Hashim
Dr. Mohammed A. Akour
Dr. Madalina Calbureanu
Dr. Reza Sirjani

Book Title : Advances in Innovative Engineering and Technologies
Conference Title : International Conference on
Innovative Engineering and Technologies
&
Civil and Structural Engineering
ISBN : 978-0-9948937-4-1



**Canadian Arena of
Applied Scientific Research Ltd.**

Suite No : 203, 2112 Osler Street,
Regina, SK, Canada S4P 1W6
Tel : 1-306-757-9985 Fax: 1-306-546-4591
Email : info@caasr.org
www.caasr.org



May 2016

Advances in Innovative Engineering and Technologies

Advances in Innovative Engineering and Technologies



Proceedings of the
International Conference on
Innovative Engineering and Technologies (ICIET '16)
&
Civil and Structural Engineering (ICCSE '16)
5th & 6th May 2016, Kuala Lumpur, Malaysia

Organized and Published by



**Canadian Arena of
Applied Scientific Research Ltd.**



Advances in Innovative Engineering and Technologies
PROCEEDINGS OF THE INTERNATIONAL CONFERENCE
ON INNOVATIVE ENGINEERING AND TECHNOLOGIES
(CAASR-ICIET '16)
PROCEEDINGS OF THE INTERNATIONAL CONFERENCE
ON CIVIL & STRUCTURAL ENGINEERING
(CAASR-ICCSE '16)

Kuala Lumpur, Malaysia

May, 5th -6th 2016

Scientific Sponsor



© Copyright 2016, by CAASR Ltd | CANADA Press

Advances in Innovative Engineering and Technologies
Proceedings of the International Conference on Innovative Engineering and
Technologies (CAASR-ICIET `16)
Proceedings of the International Conference on Civil and Structural Engineering
(CAASR-ICCSE `16)

Kuala Lumpur, Malaysia

May 5th - 6th, 2016

Published by CAASR Ltd | CANADA Press

www.caasr.org

All the copyright of the present book belongs to the Canadian Arena of Applied Scientific Research Ltd | CANADA Press. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the Editor of the Canadian Arena of Applied Scientific Research Ltd | CANADA Press.

All the articles of the present volume were peer reviewed by no less than two independent reviewers. Acceptance was granted when both reviewers' recommendations were positive.

ISBN: 978-0-9948937-4-1

Advances in Innovative Engineering and Technologies
PROCEEDINGS OF THE INTERNATIONAL CONFERENCE
ON INNOVATIVE ENGINEERING AND TECHNOLOGIES
(CAASR-ICIET `16)
PROCEEDINGS OF THE INTERNATIONAL CONFERENCE
ON CIVIL & STRUCTURAL ENGINEERING
(CAASR-ICCSE `16)

Kuala Lumpur, Malaysia

May, 5th -6th 2016

Editors

Dr. Siti Zaiton Mohd Hashim, Universiti Teknologi Malaysia, Malaysia
Dr. Mohammed A. Akour, Yarmouk University, Jordan
Dr. Madalina Calbureanu, University of Craiova, Romania
Dr. Reza Sirjani, Cyprus International University (CIU), Turkey

Advisory Board Members

Dr. Mohammad Bin Ismail, Universiti Teknologi Malaysia
Dr. Rajesh S. Prasad, Sinhgad School of Engineering, India
Prof. Elena SCUTELNICU, University of Galati, Romania
Dr. Yasser Morgan, University of Regina, Canada
Dr. Saad Harous, United Arab Emirates University, Dubai.
Dr. B.G. Geetha, KSR College of Technology, India
Dr. A. M. J. Mohamed Zubair Rahman, Al-Ameen Engineering College, India

Scientific Committee Members - Reviewers

Dr. Mark Urban-Lurain, Michigan State University, USA
Dr. Lubomir Dimitrov, Technical University of Sofia, Bulgaria.

Dr. Mario P. De Leon, De La Salle University, Manila, Philippines
 Dr. Mohamad Ali Fulazzaky, Universiti Teknologi Malaysia
 Dr. Klimis Ntalianis, National Technical University of Athens, Greece
 Dr. Said AGOUJIL, University of Moulay Ismail Meknes, Morocco
 Dr. Roszilah Hamid, Universiti Kebangsaan Malaysia, Malaysia
 Dr. K. Latha, Anna University, Trichy, India
 Dr. Azami Zaharin, Universiti Kebangsaan Malaysia, Malaysia
 Dr. Mohammadreza Vafaei, Universiti Teknologi Malaysia, Malaysia
 Dr. Khaled Galal Ahmed, United Arab Emirates University, Al-Ain, Dubai
 Dr. Yahaya Coulibaly, Universiti Teknologi Malaysia
 Dr. J. Alamelu Mangai, Birla Institute of Technology, United Arab Emirates
 Dr. Mariyana Aida Ab Kadir, Universiti Teknologi Malaysia
 Dr. Nor Fariza Mohd Nor, Universiti Kebangsaan Malaysia, Malaysia
 Dr. Saad BAKKALI, Faculty of Sciences & Techniques, Tangier, Morocco
 Dr. Mohamed Battour, University of Malaya, Malaysia
 Dr. Mahacine AMRANI, University Abdelmalek Essaâdi, Morocco
 Dr. Yilun Shang, University of Tongji, China
 Dr. Umesh C. Pati, National Institute of Technology, Odisha, India.
 Dr. Yousra Chtouki, Alakhawayn University in Ifrane, Ifrane, Morocco
 Dr. Ahmed Najim Mohammed, University of Dubai, UAE
 Dr. Yaroslav Vykhlyuk, PHEI-Bukovinian University, Ukraine.
 Dr. S. Poonkuzhali, Rajalakshmi Engineering College, India
 Dr. Muthukumar Subramanyam, DGCT, Anna University, Tamilnadu-India.
 Dr. Ma Chau Khun, Universiti Teknologi Malaysia.
 Dr. S. APPAVU alias BALAMURUGAN, K.L.N. C.I.T, India
 Dr. Sara Bilal, International Islamic University, Malaysia
 Dr. Manoj Pandian, National Health Service, Maudsley Hospital, London
 Dr. S. Senthamarai Kannan, Kalasalingam University, India.
 Dr. Mueen Uddin, University Malaysia Pahang
 Dr. Nayla El-Kork, Khalifa University, Sharjah, UAE
 Dr. Sri Devi Ravana, University of Malaya, Malaysia.
 Dr. BalaSubramanie, Kongu Engineering College, India
 Dr. Chandrasekaran Subramaniam, SRIET, India
 Prof. Anita Dani, Higher Colleges of Technology, Sharjah, UAE
 Dr. Harjit Pal Singh, Punjab Technical University Jalandhar, Punjab, India
 Dr. K. Ramesh, Government College of Technology, India
 Dr. Ramachandran Thangavelu, Karpagam Institute of Technology, India.
 Dr. Md. Jakir Hossen, Multimedia University, Melaka, Malaysia

Preface

Book: Advances in Innovative Engineering and Technologies.

The advancement of engineering and scientific technologies has continuously influenced the way we live, the way we learn; how we form opinions, communicate, purchase product, and even explore new ideas to be more innovative and creative. Industries have to do enormous research in their product development and meet the requirements of the fast moving and competitive world. Hence, CAASR International Conference on Innovative Engineering and Technologies (CAASR-ICIET`16) & CAASR International Conference on Civil and Structural Engineering (CAASR-ICCSE`16) brings together leading academic scientists, researchers and research scholars to exchange and share their experiences and research advancements in all aspects of Innovation, Engineering and Technology.

The selected works ranging from various streams like computer science; mechanical; applications; civil and structural engineering creating diversity in innovating awesome and best solutions providing know-how for developing new materials and construction of specific equipment; moving to better standard of life and satisfaction. With the nature of interdisciplinary platform, this publication provides opportunity for researchers, practitioners and educators to present and discuss the most recent innovations, trends, and concerns, practical challenges encountered and the solutions adopted in the relevant field. The whole information concentrates a significant part of available scientific and technical capability from the academic and industrial areas and its capitalization is achieved by research in collaboration with design, as long as these two different professions have a shared responsibility. We strongly believe this publication will provide significant input and serve as a catalyst for new initiatives in science, technology, engineering, and innovation. These researches definitely enhance the knowledge on the invented methodologies and provide an opportunity for the forth coming generations to design and develop new products and facilities.

The Editors

CONTENTS

Articles...

Load- deformation of RWFC Composite Wall Panels using Finite Element- A Analytical method	001
<i>Mst. Sadia Mahzabin, Ehsan Nikbakht, Roszilah hamid, W.H.W. Badaruzzaman</i>	
Study on Effect of Proximity of Interfering Building on Wind Pressure and Velocity	014
<i>Dr.Palanisamy Thangaraj, Satheesh Kumar KRP</i>	
Steel Concrete and Profiled Steel Sheet Dry Board Composite Constructions in Civil Engineering	027
<i>W H Wan Badaruzzaman, N.E. Shanmugam</i>	
Review of Customer's Opinion for Knowledge Extraction	053
<i>Dhananjay S. Gaikwad , Poonamkumar S. Hanwate, Ms. Swati Shirke, Mrs. Shwetambari A. Chiwhane, Jayashree R. Prasad</i>	
Priority Based Traffic Classification for Efficient Routing in Wireless Multimedia Sensor Network	069
<i>Rachana Borawake-Satao, Aaradhana Deshmukh, Dr. Rajesh Prasad</i>	
Discrimination Between DDOS Attacks And Flash Crowds through Adaptive Caching Mechanism in Hybrid Networks	079
<i>K.ChandraPrabha, Dr.B.G.Geetha</i>	
A Brief Survey on Text Classification Process and Algorithms	096
<i>Nihar Ranjan , Dr. Rajesh S. Prasad</i>	
Improved JPEG Anti-forensics using Non local means de-noising	108
<i>Mrs. P. M. Shelke, Dr. Rajesh S Prasad</i>	

Parallel Bi-clustering Fuzzy C-means clustering Recommendation Algorithm	116
<i>Mrs Manisha Kumbhar , Dr. Rajesh Prasad</i>	
Energy Consumption using java Framework	124
<i>N.Karthikeyan ,Dr S.Selva Brunda</i>	
Author Identification on Data Set Scale Size : A Systematic Survey	138
<i>Kale Sunil D.</i>	
Study on Behavior of Geopolymer Beam Column Joint	152
<i>Dr. Palanisamy Thangaraj, Balasubramanian Sankaranarayanan</i>	
Modeling and Simulation of MIMO Underwater Acoustic Channel for BER Enhancement	169
<i>Alaa Abdel Bary, Mohammed Khedr, Mostafa Mansour</i>	
Feature Selection in Author Identification	182
<i>Mubin Shaukat Tamboli, Dr. Rajesh Prasad</i>	
Improved Approach for Preprocessing Text using Hadoop framework	191
<i>Ms.Swati B. Bhonde,Prof. Jayashree R. Prasad</i>	
Effective Intervention for Gestational Diabetes Mellitus Based on Sentiment Analysis	203
<i>J.Jeyalakshmi, S.Poonkuzhali, S.Sreesubha, R.Deepa</i>	
Review on Variant Algorithms of High Utility Itemset Mining	217
<i>V.Kavitha, Prof. B.G.Geetha</i>	

“Improved Approach for Preprocessing Text using Hadoop framework”

Ms.Swati B. Bhonde

Prof. Jayashree R. Prasad

Abstract

“No quality data, no quality results”. Preprocessing is most important task in Sentiment Analysis as it passes data to further stages. It focuses on transforming data into a form which can be easily & effectively used as input in many domains. In this paper, we have introduced novel algorithm for the preprocessing of the text using distributed approach. This algorithm makes use of many techniques to remove redundant information used for sentiment analysis and it will result in compressing a lengthy statement into shorter and easier one without changing its meaning. Filtered sentences enhance the performance of sentiment analysis system in terms of time & space complexity also. Further this type of data can be used for different types of research domains like natural language processing, information retrieval, text classification and text clustering

Keywords- SA- Sentiment Analysis, NLTK- Natural Language Processing Tool Kit, POS - Part Of Speech, AP-Aspect polarity.

I. Introduction

Huge amount of reviews are available on Internet which can be used to determine overall opinion or feeling of product. Very often these reviews are raw so to preprocess them is one of the challenging tasks. We need to analyze them so that they can be used in intelligent decision making system. The main objective of pre-processing is to obtain the key features or key terms from existing text documents and to enhance the relevancy between word and its associated class. Pre-Processing step is crucial in determining quality of output for next stage that is classification. It's important to select specific words which convey meaning & remove other words which do not contribute to distinguish text between documents. The pre-processing phase thus converts the original data in a text-mining ready structure. Process of text mining is depicted in the following figure 1^[1]. It includes gathering of data from different sources, preprocessing text, applying text mining techniques to learn meaning of sentence. This data is then analyzed to gain specific expected information.

The importance of preprocessing is emphasized by the fact that the quantity of training data on web is growing exponentially with the dimension of the input space. It has already been proven that the time spent on preprocessing can take from 50% up

Amrutoahini College of Engineering, Sangamner Savitribai Phule Pune University, Pune
Maharashtra- India swati.bhonde@gmail.com
Sinhgad College of Engineering, Pune Savitribai Phule Pune University, Pune
Maharashtra- India jrprasad.scoe@sinhgad.edu

to 80% of the entire classification process [2], which clearly evidences the importance of preprocessing in text classification process.

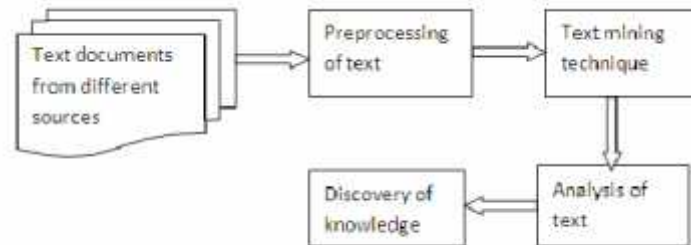


Fig. 1 Text mining process

This paper discusses the various preprocessing techniques used in the present research work. The remaining portion of the paper is organized as follows. Section II gives literature review Section III gives preprocessing steps. Section IV describes mathematical model & results followed by conclusion in section V.

II. Literature Review

Purpose of stemming is to reduce different grammatical forms or word forms of a word like its noun, adjective, verb, adverb etc Anjali Ganesh Jivani [3] has discussed different methods of stemming and their comparisons in terms of use, advantages & disadvantages. She has also focused on difference between lemmatization & stemming.

Vishal Gupta et.al [4] has evaluated the stemmer's performance in applications such as spell checker for multiple languages. This approach uses algorithm to remove suffixes using a list of frequent suffixes, He has also discussed common stemming techniques and existing stemmers for Indian languages.

K.K. Agbele [5] focused on Pervasive computing applications which are adaptable for user. He used Context-Aware Stemming algorithm, which is a modified version of the extensively used Porter's stemmer. His results have shown that the modified algorithm significantly reduces the error rate of Porter's algorithm without compromising the efficiency of Porter's algorithm.

Hassan Saif [6] has explored whether removing stop words helps or hampers the effectiveness of Twitter sentiment classification methods. He used six datasets & found that sometimes it gives negative impact on sentiment analysis system as it may happen words conveying meaning will be deleted because of adverse effect of an algorithm.

“Improved Approach for Preprocessing Text using Hadoop framework”

Implementation of sentiment analysis is needed for wide range of classification algorithms and for varying data size. There exist many possible directions for research. Work has been started in distributed & parallel domain by Lin, Jimmy, and Alek Kolcz, Liu, Bian, Jiang but literature survey shows that there is lack of system which implements all preprocessing techniques like stemming, stopword removal, emoticons handling & short informal text in one. So, system introduced here will handle all these steps. As text mining tasks are CPU-bound jobs & they consume more time, distributed framework used here significantly reduces amount of time required for preprocessing.

III. Preprocessing Steps

Following figure shows preprocessing steps used in sentiment analysis.

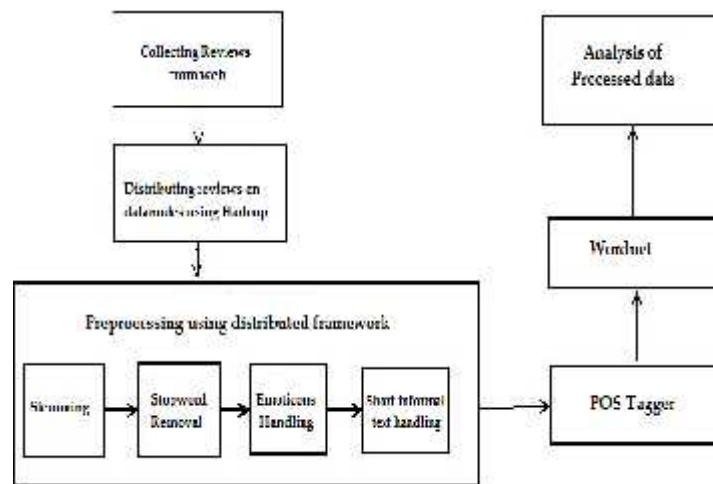


Fig 2 System Architecture

A. Review Collection:

Input or data sets used in SA plays very important role. Main source is coming from reviews from social media sites for example-in political debate, we may find peoples opinion about certain election candidate/political parties [7]. Also, we can predict results of election from posts sent by people. This raw data may contain short text which should be carefully handled and it needs special attention from programmer. Tremendous amount of work has been done in recent years on the problem of text collections in the database and information retrieval communities.

B. Stemming

This method is used to detect the root form of a word. For example, words saw, seen, seeing & see - all can be stemmed to the word “**See**” [8]. The goal here is to remove various suffixes, to reduce the number of words so that to save time & memory. This is demonstrated in Figure 3. Here morphological forms of word are treated as semantically related.

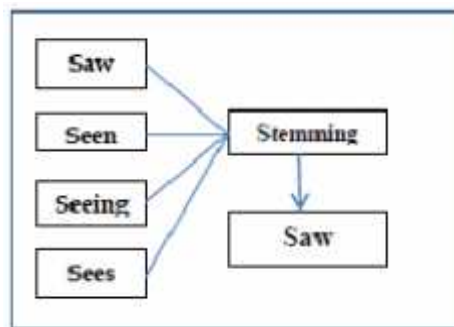


Fig. 3 Stemming example

There are two points are considered while using a stemmer:

1. Words with different meaning are kept separately.
2. Morphological forms of words are supposed to have same meaning so they are mapped to same word/stem.

Stemming is usually considered as a recall-enhancing device. For languages with relatively simple morphology, the power of stemming is less than for those with a very complex morphology. Most of the stemming experiments done so far are for English and other west European languages [9].

Usually, stemming algorithms can be classified into three groups: truncating methods, statistical methods, and mixed methods [10]. Each of these groups has a typical way of finding the stems of the word variants. These methods and the algorithms discussed in this paper are shown in the Figure 4.

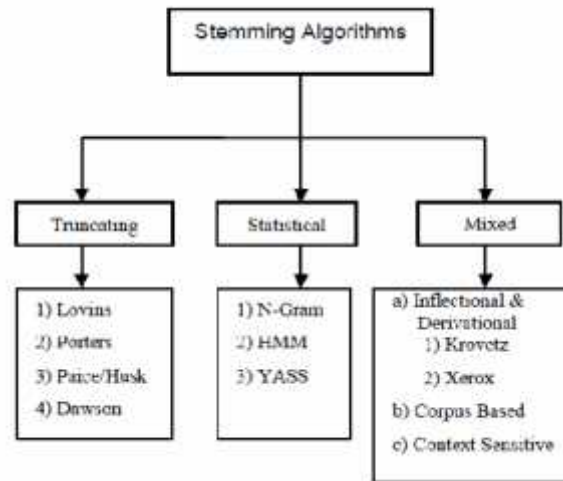


Fig 4 Types of stemming algorithms

Comparative analysis of Comparative, Statistical & Mixed type of stemming algorithms [12, 13, 14, 15]

Stemming expressively increases retrieval result for both the approaches; this in turn reduces size of index file as numbers of words to be stored & indexed are reduced. The performance of statistical stemmers is better for some rule based stemming algorithms. [16] But the main drawback that we have seen in these statistical stemmers is the poor coverage of language as stemming consumes more time this method considers only sample of documents from the corpus for this analysis and then analysis of only small collection may lead to poor coverage of the words. [17,18].

Each of these methods has their own advantages & disadvantages. In this paper, we have used Porter's algorithm.

It has five steps, and within each step, rules are repeatedly applied till one of them passes the conditions. If a rule is accepted suffix is removed accordingly and then next step is performed. The resultant stem at the end of the fifth step is returned.

The rule looks like the following:

(Condition) S1 -> S2

Where S1 and S2 are suffixes. Following are the conditions

Table 1: Conditions used in Stemmers algorithm

M	The measure of the stem
m	The stem ends with S
*S	The stem contains a vowel
V	The stem ends with a double consonant
*d	The stem ends in CVC (second C not W, X, or Y)

Example 1: (m>1) EED -> EE Condition verified: *agreed* -> *agree*

Condition not verified: *feed* -> ***feed***

Example 2: (*V*) ING -> e

Condition verified: *motoring* -> *motor*

Condition not verified: *sing* -> *sing*

C. Stop word removal

Most frequently used words in English are useless in Text mining. Such words are called Stop words. Stop words are language specific functional words which carry no specific information. It may be of the types such as pronouns, prepositions, conjunctions etc. If we remove stopwords from our sentence then also it should not affect meaning of sentence. Benefits of removing the stopwords are:-

- 1) Reduction in size of data.
- 2) Reduction in processing time.

In fact, it is possible to obtain a compression in the size of the indexing structure. List of 425 stop words is identified by researchers. Despite of these benefit, elimination of stopwords might reduce recall. For instance, consider a user who is looking for documents containing the phrase, "to be or not to be". Elimination of stopwords might leave only the term **be** making it almost impossible to properly recognize the documents which contains the phrase specified ^[1].

D. Emoticons handling

Nowadays people are becoming bad fans of using emoticons to express their emotions in all types of feedbacks & reviews. But use of emoticons makes processing of data difficult, as system is completely unaware about meaning of these emoticons. In our system we can handle these emoticons by providing specific meaning to every emoticon. Following is the list of some emoticons and their meaning.

Typical examples of emoticon synsets.^[19]

Emoticon synset	Emoticons
Happiness	:-D, =D, xD, (^ ^)
Sadness	:-[, =(
Crying	:'(, ='[, (; ;)
Boredom	- -, -.-, (> <)
Love	<3, (L)
Embarrassment	:-\$, =\$, >///<

Table 1 show different positive & negative emoticons used in reviews. Benefit of handling the emoticons is, it will correctly process emotion expressed input data ^[19]

Table 2: Emoticons used in reviews

Positive Emoticons	Negative Emoticons
o n n *_* *O* **	
:P :D :d :p	:(:(: (: :
:P :D :d :p	=(={ }):
:~) :-) :-) :-)	)' :)' :)= }=
:< > >) =)	:-{ { :-{ { :-{
=} :) (:)	:-(: :-(: :-(:
(: :) { : }	:(' {
{ : }	[:]
[: ') ') :-3	
:-3 :-x :-x :-X	
:-X :-} :-} :-]	
:-] :-)	
^ ^ ^ ^	

E. Short informal text handling.

Short informal text is one of the challenges to sentiment analysis. It is limited in length, usually spanning one sentence or less. It contains any misspellings, slang terms, and shortened forms of words.

Example: How R U?, Gr8

Okkkk, Wowww, Yup! Nops..etc

It is tedious task to process such type of data. So we need to handle it by the same way that we used while handling the emoticons. We can provide specific meaning to different types of short informal texts which user uses while giving his opinion in a review^[20]

F. POS Tagger

POS tagging is the task of labeling each word in a sentence with its appropriate parts-of-speech like noun, verb, adjective, etc. This process takes an untagged sentence as input then assigns a POS tag to words and produces tagged sentences as output. The most widely used part of speech tagset for English is PennTree bank tagset. POS and lemma of word forms are shown in Table 3 The example shown below represents the POS tagging for English sentences [21].

Example *The boy is going to the school.*

Part-of-Speech Tagging

The/DT boy/NN is/VBZ going/VBG to/TO the/DT school/NN ./.

Word	POS	Lemma
Playing	NN	Playing
Playing	VBG	Play
Walked	VBD	Walk
Pens	NNS	Pen
Training	VBG	Train
Training	NN	Training
Trains	NNS	Train
Trains	VBZ	Train

Table 3 POS & lemmas of words.

G. WordNet:

WordNet is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations.[22]

Table 4: Semantic relations in Wordnet

Semantic Relation	Syntactic Category	Examples
Synonymy (similar)	N, V, Aj, Av	pipe, tube rise, ascend sad, unhappy rapidly, speedily
Antonymy (opposite)	Aj, Av, (N, V)	wet, dry powerful, powerless friendly, unfriendly rapidly, slowly
Hyponymy (subordinate)	N	sugar maple, maple maple, tree tree, plant
Meronymy (part)	N	brim, hat gun, martini ship, fleet
Troponymy (manner)	V	march, walk whisper, speak
Entailment	V	drive, ride divorce, marry
Note: N = Nouns Aj = Adjectives V = Verbs Av = Adverbs		

H. Analysis of processed data:

Further this data is used for different natural language processing tasks such as classification, clustering, association etc.

iv. Mathematical model & results

In this system we are following the distributed approach for processing of the reviews which are collected from the web portal.

Various states carried out in this system are as follows:-

- q0 - Collecting review from web portal
- q1 - Distributing review on hadoop system
- q2 - Reviews on Datanode0.
- q3 - Review on Datanode1
- q4 & q8 - Stemming.
- q5 & q9 - Stopword removal.
- q6 & q10 - Short informal text handling, extra characters removing
- q7 & q11 - Emoticons handling
- q12 - Compress review1
- q13 - Compress review2

- q14 - Display all compressed reviews by categorized manner i.e., positive negative or neutral.

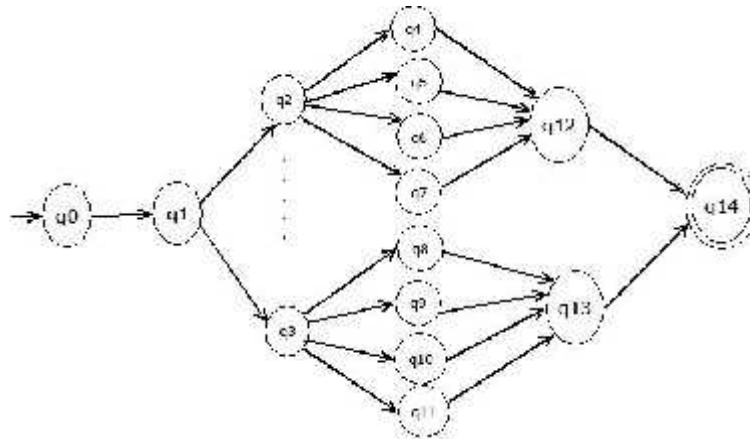


Fig. 4 State transition diagram

Sample Input:

For a given sample input results generated are shown below:

Graph shows time (in Sec) on Y axis & activities carried out on X axis.

Yesterday I brought new mobile. It has got great features. Camera quality is high but battery backup is less. Its cost is around 10K. Color of mobile is Pink. This mobile doesn't fits well into packet but its lightweight. Sound level is too good. Speed is faster but it hangs when many applications are running.

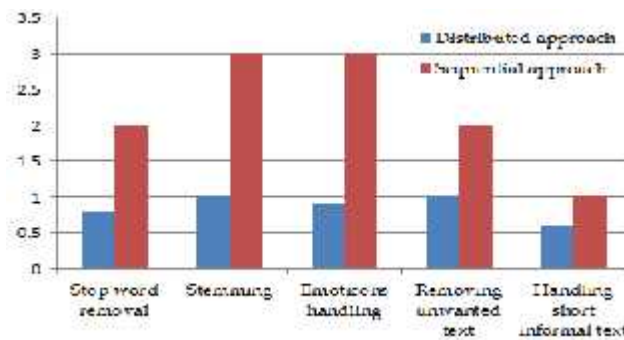


Fig. 5 Comparative results

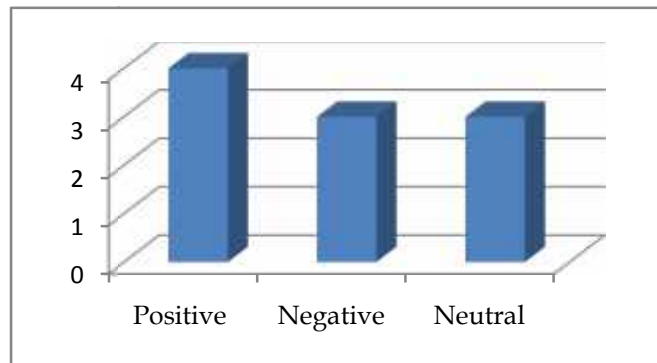


Fig. 5 Sentiment Classification

Conclusion

Pre-processing activities plays a vital role in the various applications. Therefore it is concluded that the domain specific applications gives more correct results for text mining. This paper presented pre-processing techniques using distributed approach, which significantly reduces time required in preprocessing phase. Future work will focus on use of yet more powerful stemming algorithm & comparison of distributed(Hadoop) & parallel programming(Cuda) in terms of efficiency.

Acknowledgement

I take this opportunity to extend my grateful thanks to Dr. Jayashree R. Prasad for her valuable guidance & kind support.

References:

- [1]. [1] Nikita P.Katariya *et al*, International Journal of Computer Science and Mobile Applications, Vol.3 Issue. 1, January- 2015, pg. 01-05
- [2]. [2] Katharina, M. and Martin, S. (2004) The Mining Mart Approach to Knowledge Discovery in Databases, Ning Zhong and Jiming Liu (editors), Intelligent Technologies for Information Analysis, Springer, Pp. 47-65
- [3]. [3] Anjali Ganesh Jivani , A Comparative Study of Stemming Algorithms, International Journal of Computer, Technology and Application, Volume 2, ISSN:2229-6093.
- [4]. [4] Vishal Gupta, Gurpreet Singh Lehal, A Survey of Common Stemming Techniques and Existing Stemmers for Indian Languages, Journal of Emerging Technoloigies in Web Intelligence, VOL. 5, NO. 2, MAY 2013.
- [5]. [5] Agbele, A.O. Adesina, N.A. Azeez, & A.P. Abidoye , Context-Aware Stemming Algorithm for Semantically Related Root Words, African Journal of Computing & ICT.

- [6]. [6] Hassan Saif, Miriam Fernandez, Yulan He, Harith Alani , On Stopwords, Filtering and Data Sparsity for Sentiment Analysis of Twitter.
- [7]. [7] Swati B. Bhonde, Prof. J. R. Prasad "Sentiment Analysis - Methods, Applications & Challenges", IJECCE, Vol.6, Issue 6(2015) Pp 634-640.
- [8]. [8] Saleh Alsaleem, Automated Arabic Text Categorization Using SVM and NB, International Arab Journal of e-Technology, Vol. 2, No. 2, June 2011.
- [9]. [9] Dr. S. Vijayarani, Ms. J. Ilamathi, Ms. Nithya, Preprocessing Techniques for Text Mining - An Overview
- [10]. [10] Deepika Sharma, Stemming Algorithms, A Comparative Study and their Analysis, International Journal of Applied Information Systems (IJ AIS) – ISSN : 2249-0868, Foundation of Computer Science FCS, New York, USA, Volume 4– No.3, September 2012 – www.ijais.org.
- [11]. [11] Harman Donna, How effective is suffixing? Journal of the American Society for Information Science, 1991; 42, 7-15 7.
- [12]. [12] J. B. Lovins, Development of a stemming algorithm, Mechanical Translation and Computer Linguistic., vol.11, no.1/2, pp. 22-31, 1968.
- [13]. [13] Porter M.F, An algorithm for suffix stripping, Program. 1980; 14, 130-137.
- [14]. [14] Porter M.F, Snowball: A language for stemming algorithms. 2001.
- [15]. [15] Mladenec Dunja, Automatic word lemmatization. Proceedings B of the 5th International Multi- Conference Information Society IS. 2002, 153-159.
- [16]. [16]. Deepika Sharma, Stemming Algorithms: A Comparative Study and their Analysis
- [17]. [17] Alexander Hogenboom, Daniella BalExploiting Emoticons in Sentiment Analysis, ACM 978-1-4503-1656-9/
- [18]. [18]. Svetlana Kiritchenko, Xiaodan Zhu, Sentiment Analysis of Short Informal Texts, 2014 National Research Council Canada.
- [19]. [19] Text Mining, chapter 4, "Preprocessing for English Sentence"
- [20]. [20] <https://wordnet.princeton.edu/>
- [21]. [21] V.Jude Nirmal1 and D.I. George Amalarethnam, Parallel Implementation of Big Data Pre-Processing Algorithms for Sentiment Analysis of Social Networking Data
- [22]. [22] George A. Miller, ACM, WordNet: A Lexical Database for English