

# Chapter 49

## Effectiveness of Normalization Over Processing of Textual Data Using Hybrid Approach Sentiment Analysis

**Sukhnandan Kaur Johal**

*Thapar Institute of Engineering and Technology, India*

**Rajni Mohana**

*Jaypee University of Information Technology, India*

### **ABSTRACT**

*Various natural language processing tasks are carried out to feed into computerized decision support systems. Among these, sentiment analysis is gaining more attention. The majority of sentiment analysis relies on the social media content. This web content is highly un-normalized in nature. This hinders the performance of decision support system. To enhance the performance, it is required to process data efficiently. This article proposes a novel method of normalization of web data during the pre-processing phase. It is aimed to get better results for different natural language processing tasks. This research applies this technique on data for sentiment analysis. Performance of different learning models is analysed using precision, recall, f-measure, fallout for normalize and un-normalize sentiment analysis. Results shows after normalization, some documents shift their polarity i.e. negative to positive. Experimental results show normalized data processing outperforms un-normalized data processing with better accuracy.*

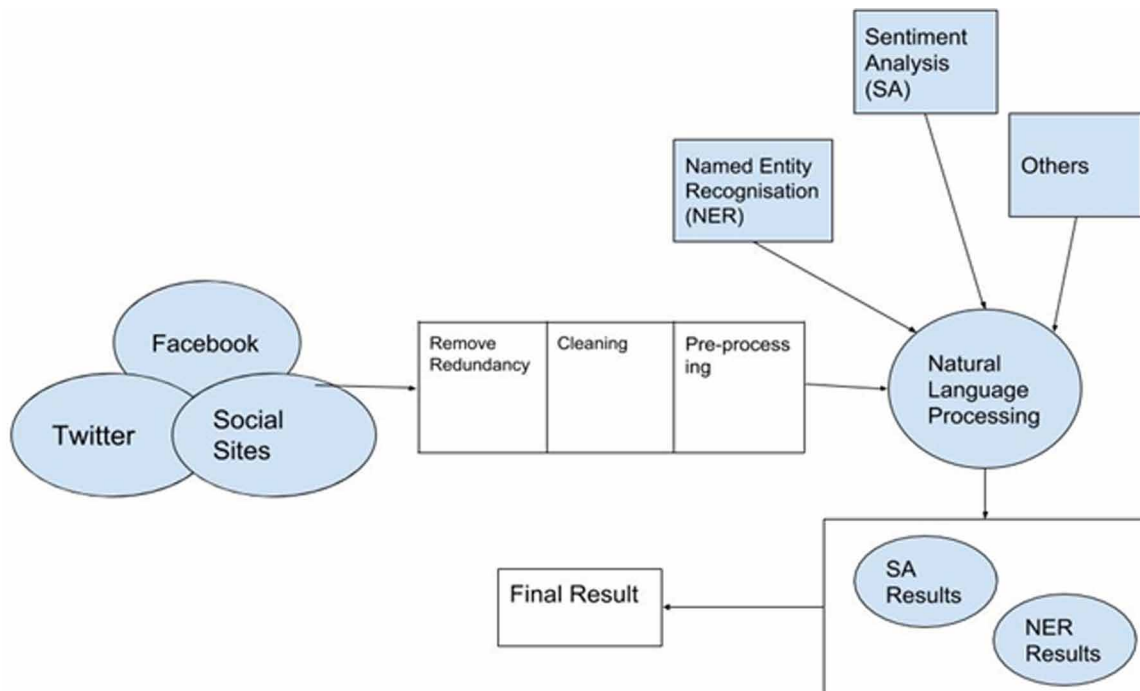
### **1. INTRODUCTION**

Natural language processing is a field of computational linguistics and artificial intelligence. It is the key to unlock various decisions using narrative web content. The automation of decision support system widely relies over the performance of natural language processors. Data available over the web sphere in various forms such as text, audio, video or pictures. Due to the arbitrary nature of the language, this data

DOI: 10.4018/978-1-6684-6303-1.ch049

is unstructured in nature. Efficiency of decision support system also gets affected by this unstructured data processing. This may sometimes hinder the performance of sentiment analyzer thus affecting the decision support system. As shown in Figure 1, initially, data is collected from the various social sites for automation of the decision support systems. Then data is pre-processed to get the structured content which includes removing the redundant content, cleaning and normalization. Later, various language processing tasks are carried out. Depending on the requirement, the results of the language processor are filtered out for the automation of decision support system. In this work, the result of sentiment analyzer (SA) is considered.

*Figure 1. Automation of decision support system*



The proliferation of web data primarily as communication medium give rise to the existence of unstructured content in the form of posts, blogs, reviews, etc. This web data is rich indicator of people’s reaction for any entity. This reaction of people is analyzed and termed as sentiment analysis in the field of natural language processing.

Classification of this web data into predefined categories, i.e. positive, negative or neutral is the task of sentiment analyzer. The web content is usually the raw data which is taken as an input by the sentiment analyzer. To reduce the performance degradation, it is necessary to pre-process data efficiently. Given the importance to minimize the human intervention in sentiment analysis and to get better results, systematized and efficient mechanisms is the need of the hour. Normalization is the basic task to handle performance degradation of various natural language processing tasks. The term normalizes in past is taken as to just make the content in a well-structured format. These days normalize has broader term in the field of natural language processing. It includes handling slangs, spell correction, finding missing

13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:  
[www.igi-global.com/chapter/effectiveness-of-normalization-over-processing-of-textual-data-using-hybrid-approach-sentiment-analysis/308527](http://www.igi-global.com/chapter/effectiveness-of-normalization-over-processing-of-textual-data-using-hybrid-approach-sentiment-analysis/308527)

## Related Content

---

### Frequent Pattern Discovery and Association Rule Mining of XML Data

Qin Ding and Gnanasekaran Sundarraj (2013). *Data Mining: Concepts, Methodologies, Tools, and Applications* (pp. 859-879).

[www.irma-international.org/chapter/frequent-pattern-discovery-association-rule/73474](http://www.irma-international.org/chapter/frequent-pattern-discovery-association-rule/73474)

### Application of Machine Training Methods to Design of New Inorganic Compounds

Nadezhda Kiselyova, Andrey Stolyarenko, Vladimir Ryazanov, Oleg Sen'ko and Alexandr Dokukin (2013). *Diagnostic Test Approaches to Machine Learning and Commonsense Reasoning Systems* (pp. 197-220).

[www.irma-international.org/chapter/application-machine-training-methods-design/69410](http://www.irma-international.org/chapter/application-machine-training-methods-design/69410)

### Some Aspects of Reliability Estimation of Loosely Coupled Web Services in Clustered Load Balancing Web Server

Abhijit Bora and Tulshi Bezboruah (2020). *Critical Approaches to Information Retrieval Research* (pp. 198-209).

[www.irma-international.org/chapter/some-aspects-of-reliability-estimation-of-loosely-coupled-web-services-in-clustered-load-balancing-web-server/237646](http://www.irma-international.org/chapter/some-aspects-of-reliability-estimation-of-loosely-coupled-web-services-in-clustered-load-balancing-web-server/237646)

### Large-Scale System for Social Media Data Warehousing: The Case of Twitter-Related Drug Abuse Events Integration

Jenhani Ferdaous and Mohamed Salah Gouider (2022). *International Journal of Data Warehousing and Mining* (pp. 1-18).

[www.irma-international.org/article/large-scale-system-for-social-media-data-warehousing/290890](http://www.irma-international.org/article/large-scale-system-for-social-media-data-warehousing/290890)

### Answer Selection in Community Question Answering Using LSTM

Saman Qureshi, Sri Khetwat Saritha and D. Kishan (2021). *New Opportunities for Sentiment Analysis and Information Processing* (pp. 153-165).

[www.irma-international.org/chapter/answer-selection-in-community-question-answering-using-lstm/286909](http://www.irma-international.org/chapter/answer-selection-in-community-question-answering-using-lstm/286909)