**Identifying Good and Bad Comments**

*GOLPERIC, Jasmine, MALAYO, Dave, MEDRANO, Jamie Paul, SEFE, Marlon*

*Department of Computer Science*

*College of Computer and Information Sciences*

*Polytechnic University of the Philippines*

*Sta. Mesa Manila*

**Chapter I**

**INTRODUCTION**

The determination of Sentiment or another indicator is another step in the process of converting unstructured content to structured content, so that the humans who interact with this ever-expanding sea of information can spot trends and patterns within the content.

A basic task in sentiment analysis is classifying the *polarity* of a given text at the document, sentence, or feature/aspect level — whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive or negative.

**Chapter II**

**RELATED WORKS**

Namrata Godbole, Manjunath Srinivasaiah and Steven Skiena conducted a research about newspapers and blogs expression opinion of news entities – people, places, things – while reporting on recent events. They presented a system that assigns scores indicating positive or

negative opinion to each distinct entity in the text corpus. ***[1]***

Tetsuya Nasukawa and Jeonghee Yi illustrated a sentiment analysis approach to extract sentiment of a specific subject on a document rather than the whole document itself, in determining its polarities. In order to improve the accuracy of the sentiment analysis, it is important to properly identify the semantic relationships between the sentiment expressions and the subject. By applying semantic analysis with a syntactic parser and sentiment lexicon, our prototype system achieved high precision (75-95%, depending on the data) in finding sentiments within Web pages and news articles. ***[2]***

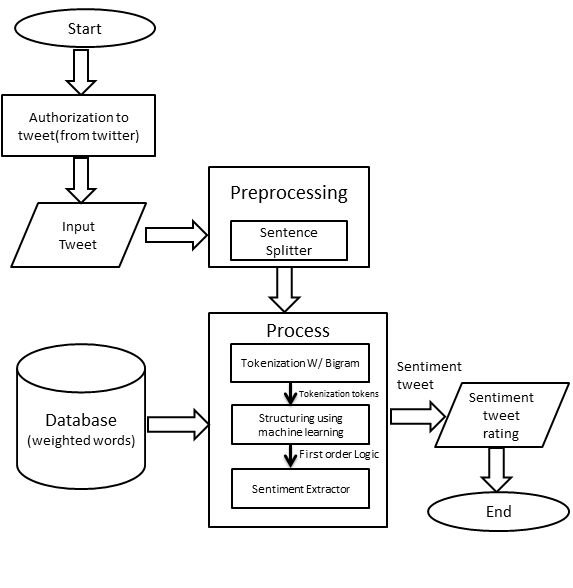
Meena Rambocas and João Gama proposed that sentiment analysis is an alternative research technique for collecting and analyzing text data on the internet, in this research they use sentiment analysis in evaluating consumer opinions. ***[3]***

Alec Go, Lei Huang, and Richa Bhayani used tweets as there domain, and the purpose of their project is to build an algorithm that can accurately classify Twitter messages as positive or negative, with respect to a query term. ***[4]***

Kuat Yessenov and Saˇsa Misailovi´c presented a paper with an empirical study of efficacy of machine learning techniques in classifying text message by semantic meaning. They used movie review comments from popular social network Digg as their data set and classify text by subjectivity/objectivity and negative/positive attitude. ***[5]***

C**hapter III**

**RESEARCH METHODOLOGY**



**Fig. 1 System Architecture**

The system works by inputting a comment that first undergoes Pre-Processing. First, the input comment will undergo Text Normalization, here shortcut words will be analyzed so that it can be transformed to its proper structure. Second is Sentence Splitting; here the converted input comment will now undergo the said process to divide it into smaller bits so that it will be ready for the next process which is Tokenizing.

When the Pre-processing stage is done it will undergo the Main process. First is Tokenization; through this, sentences will be tokenized and by the use of bigrams the program will guess an unknown word that is not included in the database then tokenize it.

Next process is structuring using machine learning where the structure of each sentence will be compared to the structure of the FOL statement. Once compared it will be easier to determine the parts of the sentence which would determine the words that contains good or bad meaning in them that will tell the polarity of the input comment.

**BIBLIOGRAPHY**

***[1]*** *Sentiment Analysis.* Retrieved from

Sentiment Analysis: Capturing Favorability Using Natural Language Processing

***[2]*** *Marketing Research.* Retrieved from

Marketing Research: The Role of Sentiment Analysis

***[3]*** *Large-Scale Sentiment Analysis for News and Blogs*

***[4]*** *Sentiment Analysis of Movie Review Comments*

***[5]*** *Pinagyamang Pluma 8.* Retrieved from

Ailene Baisa-Julian et.al.. Quezon City: Pheonix Publishing House, 2013

***[6]*** *Sentiment Analysis and Subjectivity* Retrieved from

Bing Liu, Department of Science, University of Illinois at Chicago

***[7]*** *TPOST: A Template-Based, n-gram Part-Of-Speech*

*Tagger for Tagalog*