Digos City Sentiment Analyzer Using Lexicon Based Approach

[1]Jake B. Aboyot, [1]Christian Rey P. Binoya, [1]Catherine V. Udo, [1]Cyvil Dave T. Dasargo*

[1]Department of Technical Program, UM Digos, Philippines
Corresponding author: cyvildavedasargo@gmail.com

EXECUTIVE SUMMARY

Customers ' opinions and reviews on hotels on the web are important sources of information in travel planning. Sentiment analysis or opinion mining is an important type of text analysis that supports decision-making by extracting and analyzing opinion-oriented text. It identifies positive and negative opinions. In this study, we present a Sentiment Analyzer using Lexicon Based Approach that collects such comments from the web, creates classified and structured overviews of such comments, and facilitates access to that information. It aims to create an efficient and reliable interactive website to present the overall polarity of review sites, identify the total reviews by classifying their polarities, and finally create a module that generates a graph from the classified polarity on each hotel. Lastly, the project proves itself effective. It presents a concise sentiment profile of any hotel at any particular location, and hotels in this system can also be conveniently compared.

Keywords: *natural language processing, opinion mining, sentiment analysis, lexicon-based approach, web scraping*

INTRODUCTION

Sentiment Analysis, also known as opinion mining, refers to natural language processing, text analysis, and computational linguistics to identify and extract subjective information in source materials. Sentiment Analysis is widely applied to reviews and social media for various applications ranging from marketing to customer service. Generally speaking, sentiment analysis aims to determine a speaker's attitude or a writer concerning some topic or the overall contextual polarity of a document. The attitude may be his or her judgment or evaluation (appraisal theory), affective state (that is to say, the emotional state of the author when writing), or the intended emotional communication (that is to say, the emotional effect the author wishes to have on the reader).

According to Liu (2012), the field of study focusing people's sentiments, opinions, evaluation, appraisals, and emotions towards entities such as products, services, events, topics, and their features on the social web or at any online platform is known as sentiment analysis. Sentiment analysis is a form of natural language processing (NLP) and machine learning, and its main application is related to analyzing unstructured text and extracting key concepts from that text. Sentiment analysis refers to identifying and understanding consumers' feelings and attitudes towards products through the use of natural language processing (NLP). Oxford dictionary defines the term sentiment analysis as the process of computationally identifying and categorizing opinions expressed in a piece of text, especially to determine whether the writer's attitude towards a particular topic, product, etc., is positive, negative, or neutral.

According to Patacsil et al. (2015), the process of analyzing and summarizing opinions is known as Sentiment Analysis (SA), a type of natural language processing for tracking the moods and sentiments of the public about a particular service, product or topic. Furthermore, SA may involve building a system or automated method of collecting and examining opinions about the products made through blog posts, comments, reviews, or tweets.

Sentiment Analysis is a technique to generate data on trends or reviews from customers, guests, or people's attitudes, feelings, or opinions from products and services to current events. To process the data, we filtered out the input reviews from the people, ran the statement through a sentiment analyzer from NLP (Natural Language Processing), and indexed each review's statements by numeric sentiment score and category. In which data is calculated to sentiments scores and to identify the feelings expressed in words, whether it is" positive," "neutral," and

"negative" based on peoples' sentiments from the reviews. This project aims to create a system that can benefit the user and the company, which is an efficient and reliable interactive system that will improve customer service.

PROJECT CONTEXT

Review sites or customer feedback are vital sources of information for improving operations in the service industry, but getting an accurate and complete representation of the customer experience has always been challenging. With the recent rise of social networks and reviews posted on travel portals, the hotel industry has access to a large and rapidly growing number of online customer reviews. Still, most of those reviews consist of unstructured comments that are not good to direct analysis using traditional methods. In this study, we conduct a Sentiment Analysis of the data gathered from the comments section of selected websites (i.e., Tripadvisor) to extract subjective information and to classify polarity, whether it is positive or negative so that the hotel industry can apply this kind of information from these online data to develop actionable managerial insights and improve their operations.

PURPOSE AND DESCRIPTION

User reviews and comments on hotels on the web are important sources of information in travel planning. Therefore, knowing about these comments is important for quality control needed in hotel management. In this study, we present a Sentiment Analyzer system that collects such comments from the web and creates classified and structured overviews of such comments and facilitates access to that information. By providing such a service for hotel managers that collects user reviews for hotels from various sites on the web, analyzes and classifies the textual content of the review and concisely presents the result, and create a sentiment profile of each hotel, which can be further used to compare and select hotel at a particular location by a traveler.

OBJECTIVE OF THE STUDY

This study's main objective is to determine the polarity of comments by extracting features and components of the object of the reviews that have been commented on in each document. Specifically, it aims to:

- To develop and create an efficient and reliable interactive website to present the overall polarity of review sites.

- To identify total reviews by classifying polarity if it is positive or negative.
- To determine and implement a lexicon approach for automatic application of text into positive and negative.
- To create a module that generates a graph from the classified polarity on each hotel.

TECHNICAL BACKGROUND

Digos Doctor's Hospital Web-Based Tentative Billing System and Room Availability Updates is a system that will be able to help the staff and the patients in terms of providing the necessary and accurate information about the hospital's client–oriented needs such as billing system information and room availability update. The system is focused on these features to ensure better functionality, accessibility, and an efficiently productive website. This system aims to know their tentative bill of the patient and the real-time room availability to have an effective conclusion to the problems and issues that the hospital encountered. So, the patients themselves and their families have detailed instruction or knowledge in the given system. The researchers can solve the methods by using different existing systems that we will modify and help the hospital be a more reliable, dynamic, and positive development. It will provide proper materials to support our created system in an organized aspect. The data collection (data RAW) from Digos Doctors' Hospital creates a dummy database that will generate the data into our website. The collected data reviews should store in a database used as the input to the system.

These are the selected tools that help the developers to develop the proposed system. The significant tools that are used in developing the proposed system are as follows:

- **Sublime –** it is the source code editor that the researcher used to develop the system.
- **Xampp –** a local database server used to store all the information in the system.
- **PHP –** it is a programming language that is used to develop the system.
- **CSS –** it is a style sheet language used to design the layouts of the system.
- **ROUNDME –** is a framework that can make virtual tours

Conceptual Framework

The illustration below shows the conceptual framework of the proposed project.

Figure 1. *Conceptual Framework Diagram*
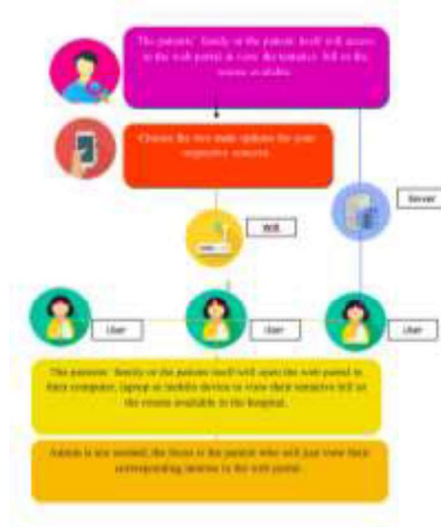
This chapter explains the support and components to be used in establishing and implementing the proposed system. This provides a discussion of the research methods, the various procedures and operations performed to conclude the subject.

Thus, this study specifies the method of research used, research design, the respondents' respondents, instruments used, the description of the study, how the activity be done, and why it should be done.

Figure 2. *RAD Diagram*
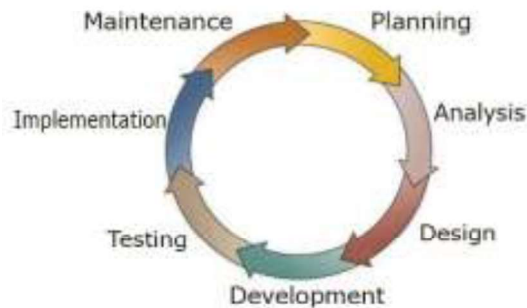
It shows in the cycle that research starts with planning, analysis, designing the system, the development of the system, integration, and testing in case of bugs, implementation, and the maintenance of the system. The table presented on the next page will help to identify each phase of the cycle.

Planning - The planning of creating and developing the website and generating different ideas are done to prepare essential documents for the proposal.

Analysis - The phase where the researchers do the analysis. The information needed is derived from the different kinds of literature studied and multiple suggestions or possible action courses taken during planning.

Design - The system is designed by the proponents and where the front-end programmer codes the different purposes and aspects of the given system.

Development - The phase where the "Digos Doctors' Hospital Web-Based Tentative Billing System and Room Availability Updates with Virtual tour" system is coded by the back-end programmer.

Testing –The phase after the development stage where the proponents try and run the given system to lessen the errors encountered. This helps proponents debug and replace some codes and integrate charges to have a successful system dry run. It will be tested again for further observation.

Implementation –The phase where the system software setup is installed into the server or the client system of the given company. This happens after the system components are met and tested successfully during the integration and integration and testing phase. Afterward, the product is handed to the client, where an orientation follows.

Maintenance –The phase where the system is being observed for its performance. The system in charge does regular maintenance to avoid errors and debugs it to enhance it for further usability and improvement.

Functional Requirements
The Functional Requirements Specification documents the operations and activities that a system must be able to perform. It could be a calculation, technical details, data manipulation and processing, and other specific functionality that define what a system is supposed to accomplish.

Interface Requirements

Admin
To log-in, the admin must fill-up the username and password field. The username field and password field accept general data entry.

Patient
The patient is not required to log in or register for identification as they can only view the result of their tentative bill in the hospital.

Regulatory/Compliance Requirements
The database will have a functional log transaction.

Non-functional Requirements
Meeting all the functionality and ability of the system would be critical for success. However, ease of use and efficiency would be adequate measures of performance as well. The users should be able to use the system effectively.

Software Requirements
- Windows Server 2008
- PHP
- Sublime Text

Hardware Requirements
- Intel Core5
- 32 GB RAM
- 2 TB Hard disks

Security Requirements
To view the tentative bill, type the admission code consists of letters, numbers, or both, and entered it in the command field. Similar to transaction code.

Reliability/Survivability Requirements
The system is completely reliable due to its real-time updates, and it can generate a tentative bill for the patients.

Maintainability Requirements
- System upgradeable
- Update functions
- Upload data

Usability Requirements
- The system allows users to access the system from the web.
- The system is user-friendly.
- The system provides guidance such as user-manual for further knowledge for the user who uses the system.
- All users will be satisfied with the usability of this system.

Design of Software Systems
This part focuses on system design and system flows and the topics included: HIPO, IPO, Manual and Proposed flow Chart, and Current and proposed Data Flow Diagrams.

Hierarchical- Input-Process-Output
HIPO is a systems analysis design aid and documentation technique from the 1970s, used to represent the modules of a system as a hierarchy and document each module.

RECOMMENDATION

This section contains the recommendation regarding the created website for Digos Doctors' Information Website with Room Availability Update. The following are the recommendation of the researchers. The proponents suggested that the researchers' system should be maintained for better usability and should have a back-up server so that all data will be protected. For future researchers, they must have additional functionalities to expand the usability and capabilities of the system.

IMPLEMENTATION PLAN

The end-users or the clients themselves constitute the system's focus created and accounted for successful system implementation. The end-user should apply during the testing, so training will need to be rolled out before that. Getting the end-users involved at this point is also a good way to get them eager about the system, as many of them may not have been involved with the project before training. Their incomparable assistance testing will help them prepare for when the system goes live. The admin may be the IT head or the IT staff and undergo orientation or training from the developers of the system.

The design made by the researchers should be maintained for user usability and efficiency. If applicable, there should be a back-up server so that all data will be protected. Thus, research information is best assured from not losing it. The proponents also recommend that the IT Department assign IT personnel who will assist and mandate the research personnel for excellent management and maintenance.

The admin is recommended to monitor and update the system to maintain it. Admin also recommends checking errors for possible outcomes that will happen. If some errors cannot be checked by the admin, they can ask the developers' help.

REFERENCES

Bulone, T. (2014). *A case study of bed management in acute care settings and the relationship to patient care* (Doctoral dissertation, Utica College).

Del Rosario, E., Del Rosario, J., Nieva, M., Tan, T., & Tangkeko, M. (2016, March). CollaborateIT: A CCS IT thesis portal with electronic document management system. In *Proc. of the DLSU Research Congress* (Vol. 4).

Nicholls, A. G., & Young, F. R. (2007). *Innovative hospital bed management using spatial technology* (Doctoral dissertation, Spatial Sciences Institute (Queensland)).

Villeneuve, C. (2011). Fujitsu's Lean Solutions Group—Lean Healthcare in Canada—. *FUJITSU Sci. Tech. J*, *47*(1), 41-48.