**CHAPTER 3**

**METHODOLOGY**

This chapter presents the methodology and procedure employed in the development of the study. Sources of data, instruments used and data gathering procedure are likewise given. This section lists and discusses specific steps and activities that were performed by the proponents to accomplish the project.

**Software Development Methodology**

The software development methodology to be utilized in the study is Extreme Programming (XP). It is a software development methodology that provides values and principles to guide the team behaviour and improves the quality of the results of the study. The researchers used this methodology because it is a pragmatic approach to program development that emphasizes results first and takes an incremental, get-something-started approach to building the product, using continual testing and revision.

**Planning**

**Designing**

Coding

Data Collection

Data Annotation

Data Preprocessing

Data Training and Evaluation

Classification using Model

Visualization

**Testing**

**Listening**

Figure 1

**Extreme Programming Methodology**

**Phase 1: Planning Phase**

This is the first stage of the Extreme Programming development life cycle. Its main task is to set goals of the entire project and certain iterative cycles. [1]

The researchers discussed the amount of Marawi siege and its humanitarian response related tweets to be collected to produce desirable classification results. The first dataset was gathered using GetOldTweets with the time frame of May 2017 to September 2018. Another tweet dataset with geolocation is also needed for the visualization.

The research proponents used a combination of Extreme Programming (XP) and Sentiment Analysis methodology for this study. The software to be implemented is composed of various modules assigned to specific tasks. The development of these modules is split equally and assigned to different proponents. Requirements and outputs of each module were identified so that a smooth transition of process form the first module, which is the collection, up to the last, which is the archive.

**Phase 2: Designing Phase**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **TASKS** | **JUNE** | | | | **JULY** | | | | **AUG** | | | | **SEPT** | | | | **OCT** | | | | | **NOV** | | | | | **DEC** | | | | | **JAN** | | | | | **FEB** | | | | | **MAR** | | | | |
| **WEEKS** | **1** | **2** | **3** | **4** | **1** | **2** | **3** | **4** | **1** | **2** | **3** | **4** | **1** | **2** | **3** | **4** | **1** | **2** | **3** | **4** | **1** | | **2** | **3** | **4** | **1** | | **2** | **3** | **4** | **1** | | **2** | **3** | **4** | **1** | | **2** | **3** | **4** | **1** | | **2** | **3** | **4** |
| Title Proposal Creation |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |
| Inquiry |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |
| Data Gathering |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |
| Preparation of the Needed Requirements |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |
| Developing Chapter 1 - 3 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |
| Submission Chapter 1 - 3 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |
| Preparation for Proposal Defense |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |
| Proposal Defense |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |
| Revision of Chapter 1 - 3 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |
| Construction of Questionnaire |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |
| Conduction of Interview |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |
| Tabulation and Interpretation of Result |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |
| Analyzing System Needs |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |
| Planning and Designing System Interface |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |
| Developing Chapter 4 - 5 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |
| System Development |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |
| Testing and Debugging System Codes |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |
| Preparation for Final Defense |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |
| Final Defense |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |
| Revision of System Based On Recommendation |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |
| Revision of Manuscript |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |
| Printing and Bookbinding |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |
| Consultation to the Advisers |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |
| Legends: |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |
| Completed |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |
| Not Yet Completed |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |

Several flowcharts were used as a guide for the system development such as Gantt Chart, Flow chart and Graphical User Interface design.

Figure 2

**Gantt Chart**

**Knowledge Requirements**

Related Literature

* Opinion Mining
* Natural Language Processing for Social Media
* Language Switching from Multiple Perspectives
* Social Role of Code Switching

Related Studies

* Acquiring and Exploiting Lexical Knowledge
* Sentiment Classification
* Estimating Filipino ISP Customer Satisfaction
* Crowdsourcing-based Annotation

Algorithms

* Naïve Bayes
* Learning Vector Quantization

Hardware Requirements

* Desktop Computer
* Intel Core i5 4thGen
* 4GB Memory

Software Requirements

* Python
* Java
* Windows 7
* WEKA

**Extreme Programming / NLP Process**

Planning Phase

* Identification of Problems
* Assigning of Tasks

Designing Phase

* Identifying Software and Hardware Requirements
* Identification of Algorithm to Be Used

Coding Phase

* Transforming Software and Hardware Requirements to Design
* Design of Software

Testing Phase

* Test Cases
* Acceptance Test
* NLP Methodology
  + Data Cleaning
  + Pre-Processing
  + Generate Topic Models
  + Topic Labeling
  + Model Visualization

Listening Phase

* Product Increment
* Revisions

**Sentiment Analysis on Marawi City Siege and Its Humanitarian Response Related Tweets**

Evaluation / Feedback

Figure 4

**Conceptual Framework**

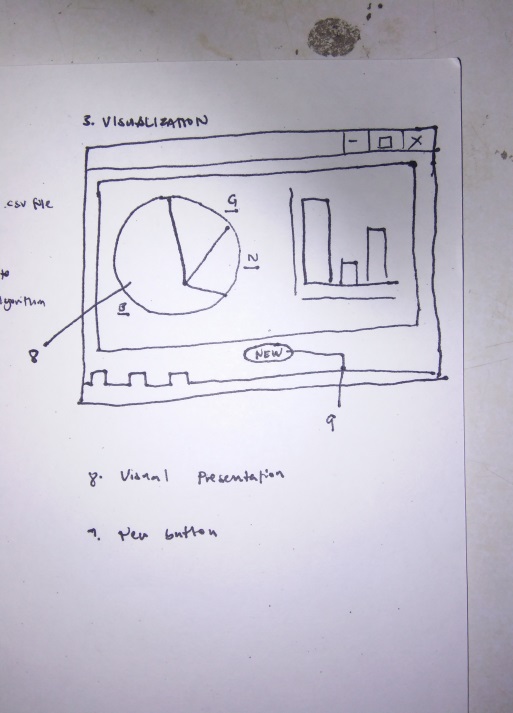
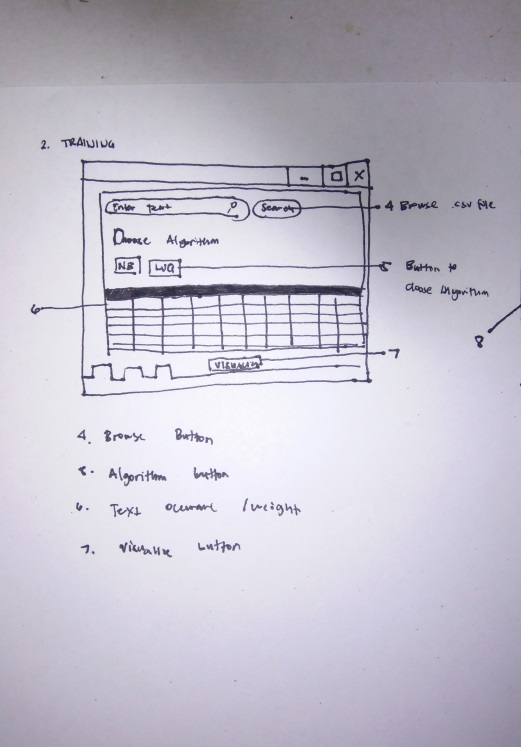
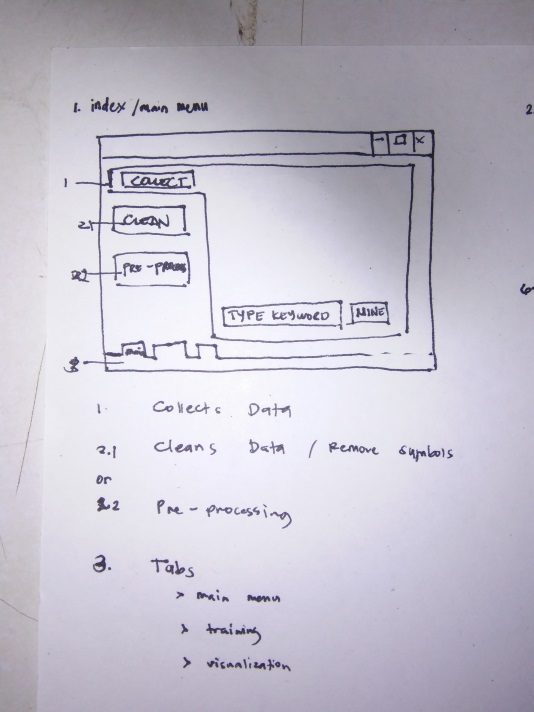


Figure 3

**Initial Graphical User Interface**

Table 1

**Hardware Requirements**

|  |  |  |
| --- | --- | --- |
| **Component** | **Minimum** | **Recommended** |
| Processor | Intel Core i5-4130 Processor (6M Cache, 4.20 GHz) | Intel Core i7-4570T Processor (8M Cache, 4.10 GHz Max Turbo Frequency) |
| RAM | 4GB | 4GB or higher |
| Hard Disk | 512 GB | 1TB |
| Internet Connection(One time connection) | Broadband Internet Connection (2mbps) | Broadband Internet Connection (4mbps) |

Table 1 shows the hardware requirements for the proposed system. The system will require a processor which is at least an Intel 5th Generation with 4.20 GHz this is to allow a better performance in collecting, cleaning, pre-processing, and generating of tweets. A minimum of 4GB of ram will be needed since almost half of the RAM is being consumed by the OS. 512GB of minimum Hard Disk space is needed to install all the software requirements since we will need to install packages other than Python. Internet connection is optional since it will only be used one time for the downloading of data in a comma-separated values format.

Table 2

**Software Requirements**

|  |  |  |
| --- | --- | --- |
| **Component** | **Minimum** | **Recommended** |
| Operating System | Windows 7 | Windows 7 or higher |
| Memory Requirement | 1GB | 2GB |
| Disk Space Requirement | At-least 500 MB | 1GB |
| Python | Version 3.6 | Version 3.6 or higher |
| Python packages (ntlk, spacy, re, numpy, pandas, ppprint, gensim, pldavis) | Latest Version | Latest Version |

The proposed system must meet the minimum software to be used as shown in Table 2. The operating system needs to run in Windows 7 or later to provide a user friendly interface for the user, and provide better compatibility on the other softwares needed. Version 3.6 or later of Python is required since all packages that needs to be installed requires a version 3.6 of Python. Latest version of Python is the minimum and recommended since it provides more compatibility and more bug fixes has been implemented. The Python packages can be installed through the pip command, however it requires an internet connection for the downloading of these packages.

**Phase 3: Coding Phase**

Coding constitutes the most important phase in the Extreme Programming life cycle. XP programming gives priority to the actual coding over all other tasks such as documentation. The researchers are tasked to translate both the initial and physical designs to actual development. The researchers will use Java as the language in the development of the system and tools will be utilized to accomplish the goals.

The methodology for Sentiment Analysis was amplified to the designing phase of Extreme Programming, as seen in Figure 1, to create a hybrid methodology of software development and sentiment analysis.

**Data Collection**

Opinions come in various ways. It could be a piece of writing or in a short form which makes the data disorganized and messy. Twitter is a microblogging website where users may post a short message with 280 characters limit; each of these posts is referred to as “tweet”. It is an outlet for everyone to let their thoughts, opinions and feelings out. In this study, the source of data comes from the Twitter platform.

**Data Annotation**

The researchers will use a supervised machine learning algorithm and human-labelled datasets that are needed during the training phase. In this case, a separate dataset containing thousands of Marawi siege response related tweets will be annotated by one of the research proponents. This data will serve as a ground truth for the training model and to allow the classifier to learn the pattern and words occurring in specific classes which are positive, negative, and neutral.

**Data Cleaning and Preprocessing**

Datasets should be cleaned and filtered to avoid inclusion of unnecessary information during training. Duplicate tweets, usernames, external links, URLs, retweets, stop words and special characters should be removed from the dataset to be used. The researchers will use GetOldTweet to gather datasets of tweets.

**Data Training and Analysis**

After the cleaning of the dataset, it will be converted into word representations and will undergo model building along with the training data. The tweets for training will be separated by their class. Datasets for supervised learning should be partitioned into two subjects: training and testing sets. The training set will serve as the grounds for learning and model building, while the test set validates the performance using evaluation metrics.

**Data Evaluation**

After the completion of sentiment analysis, the results will be tested with evaluation metrics of accuracy, precision, recall, and f1 score to validate the performance and level of correctness of the classifier.

**Data Visualization**

Visualization of data is the last and one of the important phase in Sentiment Analysis. It provides a high-level and understandable representation of the results by placing it in a visual context. Graph and map visualization will be used to visualize the classified dataset.

**Phase 4: Testing Phase**

In the testing phase, the finished system will be used to filter the collected data from the Twitter platform. Each tweet will be classified whether their contents have any relation regarding the Marawi Siege. After sorting out the tweets from the platform, they will then be grouped into three classes: positive, negative, and neutral. The results will then be displayed, showing the words commonly used in the tweets, and if the majority of them are positive, negative, or neutral.

**Phase 5: Listening Phase**

At the final stage of the life cycle the Extreme Programming, it has customer involvement through feedback during the development phase. Apart from the customer, the developer should also receive feedback from the project manager. In this phase, programmers have to take note what the customer need and should have a room for improvement. [2]

After the completion of the development of the interface application, the system will be present to the assigned programming adviser. The adviser will provide feedbacks, instructions and comments for the improvement of the overall software.

**Notes**

[1] Mckay, V. (n.d). What is Extreme Programming? (XP). Retrieved September 25, 2018, from <http://www.selectbs.com/process-maturity/what-is-extreme-programming>

[2] Extreme Programming Process Cycle. Retrieved September 25, 2018, from https://www.tutorialspoint.com/extreme\_programming/extreme\_programming\_process\_cycle.html