

**EXTRACTING AND ORGANIZING DISASTER-RELATED
PHILIPPINE COMMUNITY RESPONSES FOR AIDING
NATIONWIDE RISK REDUCTION PLANNING AND
RESPONSE**

A Thesis
Presented to
the Faculty of the College of Computer Studies
De La Salle University Manila

In Partial Fulfillment
of the Requirements for the Degree of
Master of Science in Computer Science

by

NOCON, Nicco Louis S.

Charibeth K. CHENG
Adviser

August 31, 2020

Acknowledgement

Completing this research would not have been possible without the support and guidance I have received from people and organizations throughout the years of this journey. First, I would like to gratefully acknowledge the academic support of Philippine-California Advanced Research Institutes (PCARI) through Commission on Higher Education (CHED) and Department of Science and Technology Science Education Institute (DOST-SEI). They have opened and granted opportunities that made me grow and capable of producing my outputs.

Specifically, I thank Ms. Gigi Beleno for managing my scholarship and requests. To the members of PCARI-Malasakit: Mr. Nathaniel Oco, Mr. Manolito Octaviano Jr., Mr. Matthew Go, Mr. Joseph Imperial, Ms. Alena Sipalay, and Ms. Angelica Dela Cruz, I appreciate their efforts in assisting and accommodating me with PCARI concerns and requirements.

Many thanks to my thesis panel, my adviser Dr. Charibeth Cheng, Dr. Ethel Ong, and Dr. Rachel Roxas. Extending my appreciation to Dr. Nelson Marcos and Dr. Conrado Ruiz, Jr., who have reviewed and provided valuable insights in the initial stages of my work. Without these professors' guidance and assessment, I would not have pushed my work to its full potential.

My heartfelt gratitude goes to my family and relatives. Their boundless support and patience are invaluable, and I offer this success for them, as my achievements are also theirs. To my closest friends, I thank each of you deeply, mainly for providing a comfortable, fun yet safe environment. Their efforts gave balance to my studies and social life, steering me on the right track.

Last, to God who have bestowed His grace upon me, for maintaining my good overall health and keeping me safe during the pandemic. I am indebted with His protection. Paying forward, I hope that the fruit of this research will be able to protect others too.

Abstract

Philippines is one of the most disaster-prone countries in the world. Attempting to amend its current situation, disaster risk reduction strategies were directly taken from local communities by an online participatory platform, *Malasakit*. In light of the specific needs of people, their gathered insights were unstructured. Hence, this research aims to convert unstructured data to structured form by extracting disaster-related Filipino community responses, organizing the information, and generating a report for decision makers to address, whilst assembling Filipino language processing tools. It was implemented through Part-of-Speech-based Information Extraction (IE) and Clustering techniques, namely Dice's Coefficient, Word2Vec, and FastText. As support, a Normalizer, Language Identifier, and Part-of-Speech Tagger were used and compiled in an Application Programming Interface (API). Results for extraction achieved 70.09 Precision, 80.87 Recall, 72.57 Accuracy, and 75.09 F-Measure. For clustering, all approaches joined similar ideas, yet needs to improve on capturing the right balance of relationships between words. In terms of performance, Dice provided clear relationship in clusters, while FastText covered and clustered more ideas. Notable experiments include lexicalization, which was useful in representing clusters; normalization, which produced faster runtime but lower scores in extractions; and testing pattern-based IE, which successfully processed a dataset in another domain. Products of this research are the open-source API and reports. The API can be used mainly to develop and supplement language processing tasks. Generated reports such as Excel spreadsheet and Word document can both be used as data sources for researches, but the former was made specifically for text analysis and latter for assisting decision makers in disaster strategies. Future directions involve exploring more novel approaches, including other report formats such as infographics or visualizations, commercializing and integrating the API on other applications, and studying the effects of applying the report in disaster strategies.

Keywords: Information Extraction, Clustering, Word Embeddings, Lexical Analysis, Community Responses, Disaster Risk Reduction Strategies

Contents

1	Introduction	1
1.1	Overview of the Current State of Technology	1
1.2	Research Objectives	3
1.2.1	General Objective	3
1.2.2	Specific Objectives	3
1.3	Scope and Limitations of the Research	3
1.4	Significance of the Research	4
1.5	Research Methodology	5
1.5.1	Research Design	5
1.5.2	Literature Review	5
1.5.3	Data Acquisition	5
1.5.4	Insights Processing	6
1.5.5	Experimentation	6
1.5.6	Evaluation and Analysis	6
1.5.7	Consultation	7
1.5.8	Documentation	7
2	Review of Related Literature	8

2.1	Information Extraction	8
2.1.1	FILIET: An Information Extraction System for Filipino Disaster-Related Tweets	9
2.1.2	Other IE Applications	11
2.2	Other Disaster Data Analysis	18
2.2.1	Classification	18
2.2.2	Topic and Language Modeling	19
2.3	Platforms for Sourcing Disaster Data	21
2.3.1	Malasakit	21
2.3.2	Twitter	22
2.3.3	NOAH	23
3	Theoretical Framework	28
3.1	Information Extraction	28
3.1.1	Approaches	29
3.1.2	Tasks and Subtasks	30
3.1.3	Evaluation Metrics	37
3.2	Clustering	39
3.2.1	N-grams	40
3.2.2	Sørensen-Dice Coefficient	41
3.2.3	Word Embeddings	42
3.3	Filipino Part-of-Speech (POS)	45
4	Architectural Design	48
4.1	Malasakit Community Responses	49

4.2	Preprocessing	51
4.3	Information Extraction	52
4.4	Information Organization	54
4.4.1	Information Clustering	55
4.4.2	Information Ranking	56
4.5	Report Generation	57
4.6	Report Document	58
5	Results and Discussion	60
5.1	Information Extraction	60
5.1.1	Gold Standard	61
5.1.2	Insight Phrases	61
5.1.3	Word Sets	64
5.1.4	Normalized Evaluation	66
5.2	Information Organization	66
5.2.1	General Statistics	67
5.2.2	Experiments Clustered by Dice's Coefficient	75
5.2.3	Experiments Clustered by Word2Vec	82
5.2.4	Experiments Clustered by FastText	88
5.2.5	Summary of Qualitative Experiments and Results	94
5.3	Survey on the API and Report	98
5.3.1	API Functionalities	98
5.3.2	Report Formatting	100
5.3.3	Survey Procedures	101

5.3.4	Survey Results	102
5.4	Test on News Dataset	112
5.4.1	General Statistics	112
5.4.2	Information Extraction	113
5.4.3	Information Clustering	115
6	Conclusion and Recommendations	125
A	Filipino Part-of-Speech (POS) Tagset	127
B	Application Programming Interface (API) Functions	132
C	Report Screenshots	137
D	Generated Reports	144
D.1	Sørensen-Dice Coefficient, Organized by Response Categories	144
D.2	Word2Vec, Organized by Response Categories	191
D.3	FastText, Organized by Response Categories	229
E	Informed Consent Form	254
F	Survey Forms	257
G	Survey Results and Feedback	261
H	Resource Persons	281

List of Figures

2.1 Rainfall Probability	23
2.2 Weather and Water-level	23
2.3 Flood Simulation	24
2.4 Landslide Mapping	24
2.5 Storm Surge Simulation and Hazard Map	25
3.1 Relation Network Samples	34
3.2 Temporal Relations Order	36
3.3 Vector Space Visualization Example	42
3.4 CBOW and Skip-gram Architectures	43
4.1 Architectural Diagram	48
4.2 Information Organization Options	54
4.3 Excel Report Screenshot	58
4.4 Word Report Screenshot	59
5.1 HTML Report: Front Page	110
5.2 HTML Report: Clusters	110
5.3 HTML Report: Malasakit Response List	111

5.4 Word Network or Graph ¹	111
C.1 Word Report: First Page (By Categories)	138
C.2 Word Report: First Page (All Entries)	139
C.3 Word Report: Malasakit Responses List	140
C.4 Excel Report: Response Information	141
C.5 Excel Report: Part-of-Speech (Stanford Format)	141
C.6 Excel Report: Insight Phrases	141
C.7 Excel Report: Insight Word Sets	142
C.8 Excel Report: Cluster List	143
C.9 Excel Report: Ranked Cluster List	143

List of Tables

2.1	Results in Extracting Information from Mario and Ruby Datasets	10
2.2	Comparative Summary of Information Extraction Applications	15
2.3	Comparative Summary of Other Disaster Data Analysis	20
2.4	Comparative Summary of Disaster Data Platforms	26
3.1	Table of Agreement between Two Raters	39
3.2	Word Similarity Pair Examples	43
3.3	MGNN Tagset: Main Categories	46
4.1	Sample Malasakit Responses	49
4.2	Malasakit Community Responses Codebook 4.7	50
5.1	Insight Phrases System-Gold Standard Match Count Results	61
5.2	Insight Phrases Standard Metric Results	62
5.3	Confusion Matrix of Information Extraction	64
5.4	Word Sets Match Count and Standard Metric Results	65
5.5	Report Production of Experiments	68
5.6	Report Composition of Experiments	69
5.7	Report Composition of ORC through Dice's Coefficient Clustering	71

5.8	Report Composition of ORC through Word2Vec Clustering	72
5.9	Report Composition of ORC through FastText Clustering	74
5.10	Normalization Samples	76
5.11	API Results on Usefulness	103
5.12	API Results on Ease of Use	104
5.13	API Results on Ease of Learning	104
5.14	API Results on Satisfaction	105
5.15	Report Survey Results	107
5.16	Extracted Information in News Dataset	114
5.17	Report Composition of News through Dice's Coefficient Clustering	115
5.18	Report Composition of News through Word2Vec Clustering	118
5.19	Report Composition of News through Word2Vec Clustering	121
B.1	Data Utilities Module	132
B.2	Normalization Module	133
B.3	Language Identification Module	133
B.4	Filipino Part-of-Speech Tagger Module	134
B.5	Information Extraction Module	134
B.6	Information Organization Module	134
B.7	Information Clustering Module	134
B.8	Information Ranking Module	135
B.9	Report Generation Module	135

Chapter 1

Introduction

In this chapter, an overview of the study is presented – stating the research problem, related works, proposed solution, research objectives, scope of the study, and the work's significance.

1.1 Overview of the Current State of Technology

Disasters are one of the primary reasons for disrupting the world's social and economic status. It comes in many forms, natural and man-made, such as flood, earthquake, fire, and radiation. From 2006 to 2015, the average occurrence of natural disasters is at 376.4, resulting in 69,827 deaths at the average and about US\$ 137.6 billion worth of damages. Around the world, China, USA, India, Indonesia, and Philippines are the most disaster-prone countries – where Philippines average on 18.1 counts of natural disasters annually (Guha-Sapir, Hoyois, Wallemacq, & Below, 2017).

In these occurrences, efforts have been given by numerous people and organizations to provide support for the experienced losses. There are those that provided relief programs, money and goods, and some were inspired to help and address the problem through research, even using technology for practical use. In fact, these types of situations produce a large amount of data across different sources that are usable in contributing knowledge about disasters.

One work, called *FILIET* (Regalado, Kalaw, Lu, Dela Cruz, & Garcia, 2015), took opportunity in acquiring tweets to classify and extract disaster-related contents. Since *FILIET*'s data came from an online platform, finding out relevant

information was the priority than looking for solutions that could help in disaster prevention and mitigation. As a result, the presented information, that is details about experienced disasters such as casualty, damages, and donations, was only stored in an ontology.

A more direct approach was conducted by *Malasakit* (B. M. Nonnecke et al., 2017), an online participatory tool. It collects responses from local communities with ideas and solutions on how to make the country better handle disasters, which may contain suggestions involving prevention or mitigation. Existing works analyzed the responses through classification (De La Cruz, Oco, & Roxas, 2017) and modeling (Gorro et al., 2017) techniques, providing a general representation and understanding of the responses given. Even though hints to what people want or need for their communities were presented, more can still be exploited by capturing specific points in the responses that can directly help disaster risk reduction. Doing so would present a large number of ideas that would need certain methods for arranging. Nevertheless, when these ideas are brought up, they can be further used by not only researchers, but also organizations that handles disasters in the country.

Hence, this research attempts to fill what is missing from *Malasakit* and its related works: the extraction of key insights or actionable points in community responses regarding disaster prevention or mitigation, the organization of these thoughts, and a medium that can connect local communities with respective decision makers.

In the first place, *Malasakit* responses are unstructured. This poses an opportunity to perform *Information Extraction (IE)*, a method focused on automatically extracting and providing structure to a given unstructured text (Jurafsky & Martin, 2018). A structured information has many usefulness, particularly to disseminate information to an intended target, may it be people or technology (a different program). In relation to *FILIET*, this study adopts the idea of using Part-of-Speech for Information Extraction. It is then supplemented by introducing a grouping and ranking technique that organizes the extracted ideas. Furthermore, information dissemination in a form of a list report was made as medium to relay and communicate the ideas with corresponding decision makers.

1.2 Research Objectives

1.2.1 General Objective

The objective of this research is to extract and organize disaster-related community responses and generate a report for decision makers to address.

1.2.2 Specific Objectives

This research specifically aims to:

- Extract key insights or actionable points in community responses;
- Organize similar ideas from extracted text;
- Generate report from structured information; and
- Evaluate system's performance and report's contents.

1.3 Scope and Limitations of the Research

A Part-of-Speech (POS)-based Information Extraction (IE) system was developed to identify and filter suggested details from a disaster domain data. In particular, these are keywords or phrases that contain ideas or actions to an object or place that can prevent or mitigate disasters in the Philippines.

Data extracted were from Philippine local community responses collected by *Malasakit* (B. M. Nonnecke et al., 2017), an online survey tool gathering insights on how their Barangay ‘neighborhood’ can help them prepare for disasters (more details at Chapter 4). 934 responses in text form were used in this study, focusing on responses in Filipino language, which includes words in English. These responses undergone preprocessing to format the data, fix typographical errors and expand shortcut texts. Additionally, a gold standard was manually prepared and used as guide for comparison in evaluating and analyzing the system’s extractions, containing the ideal or correct extracted information in responses.

Post-extraction, related information was grouped by the affixed response category in *Malasakit*'s data to give a general view of the information, while similar concepts were collated through clustering to reduce duplicates. Techniques investigated in this study are n-grams, Sørensen-Dice coefficient, and word embeddings. Aside from grouping, ranking of the information was part of the organization task. Beforehand, *Malasakit*'s categories were sorted based on urgency, with highest priority on those representing concrete actions that can be addressed by decision makers. Provided with this, there are two ways that can be used for information ranking, these are through frequency counts of the extracted entries and the priority arrangement of the categories.

The system includes a report generation module that produces a formal collection of information. The report was generated through a template-based technique. It is in a list format, emphasizing the solutions or actions that decision makers can act on. In detail, there would be fields pointing to an extracted information's frequency count, response category, and the solution's details like "proposed action" and "target".

Standard metrics were reviewed to find appropriate evaluation for the IE product and research's application in the real world. The metrics serve as performance basis for future related works. In addition to that, the generated report was assessed based on its usefulness of information and format.

1.4 Significance of the Research

This research contributes to Philippines' disaster risk reduction planning and management by extracting and presenting key insights provided by the Filipino community. In light with that, the information extracted extended the capabilities of *Malasakit* (B. M. Nonnecke et al., 2017) which can make use of the product of this research to advise corresponding decision makers such as organizations and government agencies about the primary needs of people in their local areas.

Furthermore, the product of this research can act as a platform for collecting and relaying solutions about disasters, a component usable not only to decision makers but also to current and future researches. In the same way, it can be used as an inspiration to conduct research that tackle disaster-related problems and provide an effective, concrete solution by connecting citizens with decision makers; especially on researches that involves dissemination as solution for a problem. Adding to this, API functionalities presented can be used to develop, extend, or supplement various applications related to the domain and field.

1.5 Research Methodology

This section enumerates the activities that led to the completion of the research. Included activities consists of Research Design, Literature Review, Data Acquisition, Insights Processing, Experimentation, Evaluation and Analysis, Consultation, and Documentation.

1.5.1 Research Design

The motivation, problem, and purpose of the research were defined in this stage. It includes formulating an idea for the solution, setting the steps necessary to address the problem, identifying resources needed, the scope and limitations in doing research, and the significance of solving the problem and producing the solution (refer to Chapter 1).

1.5.2 Literature Review

Existing works on Disaster, Information Extraction and Clustering were surveyed to expand the proponent's knowledge in the given fields (refer to Chapters 2 and 3). Various approaches in several domains and how they were evaluated were examined, finding the closest literature for this study. Ideas from these related literatures were applied, which took account but not limited to the involved components, data or resources, metrics, and underlying research gaps.

1.5.3 Data Acquisition

Data gathered came from *Malasakit's* Local Community Responses (B. M. Nonnecke et al., 2017). It captured insights given by people to improve their Barangay or community (more information available at Chapter 4). This data ensured the absence of personal identifiable information as it only includes those for extracting insights, including actual community responses or comments and annotations done by *Malasakit* pertaining to the category of user comments. The responses have been preprocessed and corrected further of misspellings, shortcut texts and spacing. Along this data, a gold standard was manually prepared by the proponent for evaluating the performance of the information extraction process. The expected (or correct) set of insights to be extracted per response were listed accordingly. This gold standard was used for evaluation and analysis.

1.5.4 Insights Processing

The architectural design (see Chapter 4) shows the processing of disaster-related insights. It is expressed by the unstructured responses going through four main software modules implemented, namely information extraction, information clustering, information ranking, and report generation modules. In addition to this, a preprocessing module was developed for applying input text correction or standardization, and language identification. Moreover, this design is also supplemented by including other resources utilized by the software such as tagger and word embeddings models, and a view of the report (see Figure 4.4 and Figure 4.3) which contains suggestion entries. Formats for the report were provided, with other designs in accessing the information considered and presented. Additional tasks included in this activity debugged and integrated these components to form a holistic software.

1.5.5 Experimentation

Having *Malasakit's* community data and developed software that processes insights, there were experimentations on both the utilized data and software components. In terms of data, there was a test on another dataset in another domain. The assumption is that since the solution processes texts, it should be able to handle inputs regardless of the domain. On the other hand, software components involved adjustments in configurations which can be through the usage of different clustering approaches, preprocessing tools, configurations, and lexicalization of entries.

1.5.6 Evaluation and Analysis

Evaluation came in two parts: system and report evaluation. In system evaluation, the Information Extraction's performance was measured through standard, quantitative metrics such as Precision, Recall, Accuracy, and F-Measure. Whereas, for the organized information a qualitative analysis was done by the proponent to bring out the characteristics (positive and negative) and coverage of the approaches and collated information. For the generated report, selected *Malasakit* members evaluated it based on the usefulness and presentation of the information. The assessment can be then used to improve the output and accessibility of the developed program. Generally, both parts have corresponding analyses that highlighted their contents, capabilities, issues, and limitations.

1.5.7 Consultation

Experts in Disaster, Research, and Computer Science fields were continuously consulted to facilitate better contents in the software, report, and document. Considering their inputs, changes were reflected in the document. The desired frequency for consultation was achieved, at least once a month with either of the experts.

1.5.8 Documentation

Throughout the research, constant documentation was done, which reflected the products of every research stages. It was written formally, covering from the tackled problem to the implemented solution, with affixed analyses and evaluations in this research.

Chapter 2

Review of Related Literature

Philippines has experienced countless disasters in the past years. These occurrences resulted into an outburst of data all over different mediums. Under this chapter are several works from researchers who were able to make use of disaster data and Information Extraction techniques, discussing their problems, solutions, sources of data, and architectures.

2.1 Information Extraction

Information Extraction (IE) is a subarea of Natural Language Processing, where it turns unstructured texts into structured texts (Jurafsky & Martin, 2018). A basic overview of the process starts with an unstructured text that will be processed by analyzing and finding features to extract information. Then, given enough features, intended information are extracted and placed on a structured formatting. Normally, components used consists of a Named-Entity Recognition (NER), Relation Extraction, Temporal Expression, Event Extraction, and Template Filling (Jurafsky & Martin, 2018). Although there are times that components could be lesser or greater than this generic architecture, that depends mainly on the problem being dealt with (how simple or complex it is). There are several ways to develop these components, implementations include rule-based, machine learning, ontology-based, or hybrid (mixed) approaches.

In this section, several IE implementations were reviewed. In these works, different approaches and architectures for solving a particular problem or domain were supplied – featuring an IE for disaster and more from other domains such as medicine, media, and business.

2.1.1 FILIET: An Information Extraction System for Filipino Disaster-Related Tweets

The research done by Regalado et al. (2015) is about *FILIET*, a Filipino Information Extraction tool for disaster-related tweets. Their main goal is to capture the most relevant disaster-related information from twitter and make use of these information to help in disaster relief efforts.

FILIET processed around 2,000 tweets about Typhoon Mario and Ruby, that were manually tagged into five categories: Caution and Advise (CA), Casualty and Damage (CD), Call for Help (CH), Donation (D), and Others (O). *FILIET* is composed of six modules, namely crawler, preprocessing, feature extraction, category classifier, rule inductor, and ontology population module.

The crawler module functions as an automated tweet collector. Twitter's Stream API and Twitter4J was used to collect tweets and was preprocessed using a shortcut text normalizer (NormAPI), tokenizer (ArkNLP), Part-of-Speech tagger (POS Lookup), and Filipino Named-entity Recognition (SOMIDIA gazetteer).

Under feature extraction, presence (binary feature indicating presence of hashtags, URLs, and retweets), tweet length, and word features (unique words without stop words, accented characters, and links) were extracted from tweets.

The process is followed by the classifier module, for which tweets with extracted features were labeled with the five categories through k-Nearest Neighbor, Random Forest, and J48 algorithms using Weka.

Provided with the labels, handcrafted pattern-based, POS-marked rules were applied to extract information like typhoon signals, suspension of classes, casualties, damages, and items donated.

Finally, the extracted information was stored into a retrievable ontology using Protégé and OWL API, that functions as a connector for information that are related with each other.

Experimenting on *FILIET* extracted multiple information such as CA's location and advice/caution, CD's location, object destroyed, object details and victim's name, CH's location and victim's name, and D's location, donated items and item details. Results for the experimentations garnered multiple values found on Table 2.1, with absolute F-Measure scores on Mario dataset's D-Resource, D-Detail, and D-Victim extracted information, and Ruby's CD-Victim, D-Detail, and D-Victim extracted information. Generalizing *FILIET*'s IE performance, the authors indicated a 0.4 F-measure score.

Table 2.1: Results in Extracting Information from Mario and Ruby Datasets

Categories	MARIO			RUBY		
	Precision	Recall	F-Measure	Precision	Recall	F-Measure
CA-Advice	0.5593	0.3388	0.4219	0.6332	0.3010	0.4080
CA-Location	0.6762	0.3352	0.4482	0.8216	0.4454	0.5777
CD-Object	0.4737	0.1125	0.1818	0.5693	0.3790	0.4550
CD-Detail	0	0	0	0.7531	0.1317	0.2247
CD-Victim	1	0.9825	0.9912	1	1	1
CD-Location	0.4700	0.0803	0.1372	0.6274	0.5142	0.5652
D-Resource	1	1	1	0.9688	0.8267	0.8921
D-Detail	1	1	1	1	1	1
D-Victim	1	1	1	1	1	1
D-Location	1	0.2602	0.4130	1	0.2778	0.4348

Provided with the scores, there are parts in this work that can be improved, namely improvement on the preprocessing modules, categories used, and extraction rules. For preprocessing modules, state-of-the-art or improved models for Filipino POS and Named-entity Recognition can be utilized. It is also stated in the recommendations that FILIET has room to implement a Lemmatizer module for better detection and handling of words.

Regarding the categories used, there can be experimentations into increasing and decreasing its scope (adding and reducing categories), to fit the contents in the tweets. In extraction rules, POS markers can be extended into capturing more patterns and also by adding specific set of rules for special and/or common cases. Additionally, an automated, adaptive approach can be implemented to ease in developing the said rules.

FILIET is certainly one of the concrete examples that uses technology to analyze disaster-related texts. However, it focused more on finding relevant information than its second goal, to make use of the extracted information. Granted, it missed the part wherein information is used by people, could be through the act of relaying the information to respective bodies or organizations such as the Barangay, Fire Stations, and more, or simply through visualizing its findings.

Not only that, information extracted consists of casualty, damages, and donations, which are details about a disaster. It may be better to couple the information with a set of solutions in order to assist disaster handling. Examples could be pointing out locations with high casualties to find ways in lessening them, finding areas with high damages to secure and fix, and organize donations from twitter users. Exceedingly important, is to find ways on how to prevent and mitigate disasters.

2.1.2 Other IE Applications

Reviewing other literature showed that there are notable IE applications distinguishable by their purpose and usage of IE, data processed, methodology, and extracted information. Below are works arranged by different types of approaches used, specifically rule-based, ontology-based, machine learning, and hybridized approach. A comparative summary is provided at Table 2.2, with added details regarding each work's data, technologies, evaluation metrics, and performance scores.

C. Cheng, Cagampang, and Lim (2016) generalized data points that resulted into a “read less, know more” information by extracting the 5Ws (*who, what, when, where, why*) from Filipino news articles. Their approach made use of grammar rules or markers and sentiment analysis. The rules used were the following:

- for *who*, a *pantukoy* (article) was used as a marker to determine people (e.g., the word/s following *si*, *sina*, *ni*, and *nila*);
- for *what*, phrases in gold standard was used;
- for *when*, names of months, days, and time formats were used;
- for *where*, adverb of place, *pang-abay sa lugar* or *panlunan* (e.g., word/s following *kina*, *nasa*, *mula sa*, and *tungo sa*) were used; and
- for *why*, conjunctions (cause) or *pangatnig na panahon* were used (e.g., word/s following *kaya* and *dahil sa*).

These kinds of rules can be effective on a specific domain and is capable of being extended further to better fit other domains and cover a larger variety of examples. Additional action to take in order to capture more information is to narrow down the details (lessening the words to process) before extraction. Solely, the concept of extracting 5Ws present valuable information that more or less provides the whole sense of a text.

Saggion, Funk, Maynard, and Bontcheva's (2007) study aspired to model information about countries, regions, and companies that can be used to determine areas where businesses are ideal to be set up. OWL and PROTON upper ontology was used, where it contains the said information, formatted with their relations, properties, and hierarchies. The General Architecture for Text Engineering (GATE) framework was also applied to develop a large language processing technology containing ANNIE, an IE system. The company intelligence IE takes

up-to-date information such as company name, their activities, employee count, names of board of directors, and more; whilst the country/region intelligence IE extracts details such as country name, official language, currency, exchange rate, foreign debt, unemployment rate, GDP, foreign investments, region area, population, and more.

Having developed an IE system that targets businesses, results are crucial. Even with quantitative basis and capability of pointing out valuable information that can get businesses confident about their decisions, without testing and analysis on user acceptance, it could not be evaluated based on its impact or effectiveness in the real world. Hence, in this type of systems, having businesses try and rate the system is as important as developing the product.

Peña and Melgar's (2015) research used Protégé ontology editor and extracted information from discussion forums. The subject of the forum containing comments from students are about course offerings. Their idea is to help improve institutions by providing information to which courses students are more interested in. With attention to data coming from people, complications may incur and affect how the system processes, for which requires a more considerate processing or handling of texts. That is why, as their study dealt with informal language, heavy text standardization was applied: filtering out offensive language, special characters, and irrelevant texts, while misspelled words were corrected and words in general are reverted to their root words (lemma). Cleaned and standardized, the information extracted includes courses, teachers, materials, comments, and words that refer to a subject.

Developing an IE to collect insights from and for students is a step towards making one for a larger population. In this study, the purpose was to listen to students and find out which courses are more interesting; modifying this purpose a little bit, an IE system can be used to gather insights in largely engaging topics such as politics and disasters – answering “what do people think or want?”.

Culotta, Bekkerman, and McCallum (2004) aimed to connect people through email and webpage contents by generating a large social network and contact information using Conditional Random Fields (CRF) IE and Graph-partitioning clustering. From personal emails and with the help of the Web, 25 fields were successfully extracted such as First Name, Middle Name, Last Name, Nick Name, Suffix, Job Title, Company Name, Address Line, Country, Mobile Phone, and many more – forming an address book. The address books are then connected to others that forms a social network, for which clustering makes sure, through thresholds, that connections are under the right groupings.

Approaching Web data as a supplement into providing and verifying more related information is evident in this work. Local and Web data were bridged together that boosted the extraction and clustering process. This gives an idea to make use of the Web as a large database of information.

The next set of works combined rule-based, ontology-based, and/or machine learning approaches to form a hybrid IE. The combination can be through the IE as a whole or mixed under its components and also the distribution can be balanced or weighing more on a particular approach.

Livelo, Ver, Chua, Yao, and Cheng's (2017) work introduced machine learning to C. Cheng et al.'s (2016) linguistic rules IE, that is intended to speed up processing and understanding of unstructured news data. Processing of the system starts with parsing and transferring news articles into a word table – with word tokens, Part-of-Speech (POS) tags, and Named-entity tags field. Possible 5W candidates were then gathered through a rule-based approach, making use of the information provided by the word table. Filtering the candidates, machine learning (*who*, *when*, *where*), rule-based weighing (*what*), and hybrid (*why*) feature extraction were performed, retaining the most suitable candidate with respect to the Ws.

In their hybrid IE implementation, they took into account other criteria and made use of machine learning to automate the decision process. The machine learning component looked into the candidate string, number of words, sentence for which the candidate belongs, numeric position of the candidate, distance of the candidate from the beginning of the sentence, frequency count, 10 word-neighbors (before and after), and POS tags of word-neighbors to find the most suitable information to be extracted. In this way, the combined approaches efforts considered a wider range of information, covering whole sentences. Despite having the ability to extract lots of information, their main issue lies in determining which is the right one to represent the whole article. Having said that, there has to be considerations outside quantitative metrics as to which information are or can be accepted (even if it is over- or under-extracted) to satisfy the goal.

Ali et al.'s (2017) IE is another that used OWL ontology, but instead was used on social robots or “an interface to convey information and to provide recommendations” in service to persons with disability. Particularly, multiple ontologies (i.e., domain, hotel, medical, city and transport) were merged for extraction and was added with a Support Vector Machines (LIBSVM) classifier for filtering out irrelevant information to form a hybrid IE. The social robot accepts three types of query related to medical drugs (specifically diabetes), hotel accommodations, and city information, then extracts precise information such as name, link, information, and polarity from the Web.

The main issue in this work is with regards to processing large data. There are ways to lessen the load such as retrieving only the top related searches from the Web or introduce a better equipped machine learning algorithm like Artificial Neural Networks to potentially produce more precise and wider scope of information. Other than that, applying hybrid IE for social robots accommodated the extraction of large data through machine learning and extraction of detailed information through ontologies.

Maternal Medical Information Extraction (MaMIE) System by Borra, Santos, Gonzales, and Reyes (2013) extracted English maternal health records to lessen the details of such records whilst maintaining vital information. Records were taken from De La Salle University Medical Center (DLSUMC), Dasmariñas, Cavite, Philippines and was processed through seven modules namely, Text Zoner, Sentence Splitter, POS Tagger, Named-Entity Recognizer (NER), Pre-parser (POS groups), Semantic Interpreter, and Template Generator. Its hybridized approach consists of a statistical NER, while the remaining are rule-based. Resulting extracted information consists of the general information, obstetrical history, family history, past medical history, social and personal history, vital signs, and clinical discussion section of a patients records, which are all stored in a database.

Based on their work, one important point in extracting medical information is aside from only using generic terminologies, their IE has to be custom fitted to include medical (e.g., diseases, medicines, and procedures) and local (e.g., locations, names, and organizations) terminologies to capture integral parts of the text. Simplistically and generally, the domain knowledge must be prioritized to better handle the texts.

Table 2.2: Comparative Summary of Information Extraction Applications

Related Literature	Resource/s	Techniques	Evaluation Scores
FILIET: An Information Extraction System for Filipino Disaster-Related Tweets (Regalado et al., 2015)	2,000 tweets	Twitter's Stream API, Twitter4J (Crawler), Protégé and Owl API (Ontology), Weka Classifier: k-Nearest Neighbor, Random Forest and J48 [IE] NormAPI (Normalizer), ArkNLP (Tokenizer), Part-of-Speech tagger (POS Lookup), and SOMIDIA Gazetteer (Filipino NER)	Multiple Values (per category), but is represented by 40% F-measure
Organizing News Articles and Editorials through Information Extraction and Sentiment Analysis (C. Cheng et al., 2016)	800 Filipino News Articles	Grammar Rules Information Extraction and Sentiment Analysis: Bag-of-words, TF-IDF weighting, Naïve Bayes, and Support Vector Machines	[IE] Kappa Statistics - What: 5.49%, Who: 24.65%, Where: 56.51%, Why: 6.82%, and When: 68.71%; and Correctness - What: 5.88%, Who: 6.06%, Where: 19.51%, Why: 50%, and When: 84.39% [Tested with Hybrid IE of Livelo et al.] Complete Matches - What: 0.00%, Who: 6.06%, Where: 19.51%, Why: 50%, and When: 84.39%

Table 2.2 continued from previous page

Related Literature	Resource/s	Techniques	Evaluation Scores
Extracting Social Networks and Contact Information from Email and the Web (Culotta et al., 2004)	1441 Total Email Messages from Two Recipients	Conditional Random Fields (CRF) Information Extraction and Graph-partitioning Clustering	[IE] Accuracy (Token): 94.50%, F-score: 80.76%, Precision: 85.73%, and Recall: 76.33%
Maternal Medical Information Extraction (MaMIE) System (Borra et al., 2013)	12 Maternal Records	Text Zoner, Sentence Splitter, POS Tagger, LingPipe Named-Entity Recognizer (NER), Pre-parser, Semantic Interpreter, and Template Generator	[IE] Precision: 87.39%, Recall: 86.74%, and F-measure: 87.06% [POS tagger] Accuracy: 82% [NER] Accuracy: 70.58%
Ontology-based information extraction for business intelligence (Saggion et al., 2007)	MUSING Document Repository, Company Web Pages, Financial News (Yahoo! Finance), World Bank, Monetary Fund, BBC, Wikipedia, and CIA World Fact Book	OWL, PROTON, GATE Framework, ANNIE, Named Entity Recognition (JAPE), OCAT	[Company Intelligence IE] Precision: 85.6%, Recall: 93.6%, and F-score: 84% [Country/Region Intelligence IE] Precision: 94%, Recall: 67%, and F-score: 81%
Ontology-based Information Extraction from Spanish Forum (Peña & Melgar, 2015)	Spanish Discussion Forum	Protégé, Freeling, Lucene, and Jena API	Precision: 76%, Recall: 75%, and F-score: 75%

Table 2.2 continued from previous page

Related Literature	Resource/s	Techniques	Evaluation Scores
A Hybrid Agent for Automatically Determining and Extracting the 5Ws of Filipino News Articles (Livelo et al., 2017)	250 Filipino News Articles	[Hybrid IE] Rule-based and Machine Learning (J48, Naïve Bayes and Support Vector Machines), TPOST, and Stanford NER	[IE] Accuracy – What: 89.20%, Who: 63.33%, Where: 59.25%, Why: 50%, and When: 71.38%; F-Measure – What: 94.29%, Who: 77.29%, Where: 45.13%, Why: 50%, and When: 71.38%; and Kappa Statistics – What: 74.40%, Who: 59.35%, Where: 71.00%, Why: 70.40%, and When: 61.25% [Tested with Rule-based IE of C. Cheng et al.] Complete Matches – What: 28%, Who: 43.84%, Where: 56.4593%, Why: 11%, and When: 59.1743%
Merged Ontology and SVM-Based Information Extraction and Recommendation System for Social Robots (Ali et al., 2017)	4902 Training Sentences and 2737 Test Sentences	Merged Ontology, SVM (LIBSVM), WordNet, GATE, GSE, Protégé, OWL, DL, SPARQL, Pellet, FACT++, and Hermit	Precision: 95%, Recall: 77%, and Accuracy (F-measure): 85%

2.2 Other Disaster Data Analysis

In this section, ways to analyze and mine disaster texts are presented. Specifically discussing text classification (or categorization), topic modeling, and word embeddings.

2.2.1 Classification

In the study of De La Cruz et al. (2017), a multi-class classifier has been developed as a module for an e-participation toolkit (B. M. Nonnecke et al., 2017). It classified community responses related to disaster preparedness and contributes to the field by adding disaster-relevant data for classifiers. There are 10 categories for classification namely, Local Government Units (LGU) accountability, Education and training, Communication, Solidarity, Early warning, Flood control, Personal preparedness, Equipment and supplies, Relief, and Infrastructure.

There were 263 responses used for this study. Each response was cleaned through regular expressions and manually categorized. These responses acted as the classifier's training data, whilst 500 untagged responses were used for testing. Additional pre-processing for the data includes the computation of Term Frequency - Inverse Document Frequency (TF-IDF) weights, a text weighing system with respect to a terminology's relevance over the whole document.

Using the training data, a Support Vector Machine classifier model was created. The classifier labeled the untagged responses and was evaluated through 10-fold cross validation and a social scientist. With 10-folds, the average accuracy of the classifier was 93.30%. However, from the verification of the social scientist, only 63.8% was deemed to be correctly classified due to the following reasons: the classifier's TF-IDF bias - words that frequently appear in a given category will automatically be labeled under that category (potentially the inability to make use of verbs), untagged responses contain misspellings, and ambiguous/broad categories.

De La Cruz et al.'s work acts as an organizer that provides automatic annotation to community responses or data. A research gap on this matter pertains to the limited and outdated categories it can use as label, for which this classifier may need to be updated. Despite this, the stated 10 categories in this paper may be used as basis for importance or urgency; since they are the ones initially established, they must be the most important categories at the moment.

2.2.2 Topic and Language Modeling

Gorro et al.'s (2017) work involves the use of word embeddings (word2vec) and bi-term topic modeling for qualitative analysis. Data analyzed consists 976 disaster risk reduction responses taken from Philippine barangay communities and their main goal is to find out the people's narratives regarding disasters and discover clues as to topics with importance. The data undergone preprocessing, removing data noise such as special characters, numbers, and unnecessary words.

Through bi-term topic modeling, a process of finding, learning and using two terms that are occurring frequently in a particular order to determine similar topics, they produced 10 topics labeled through a manual qualitative method called "open coding". Ideas revolving among the 10 topics prioritizes disaster preparedness through warnings, evacuation preparation, infrastructure and solid waste management. Evaluating the model, a word intrusion test was performed by replacing a word in a topic with an irrelevant word and letting human evaluators point out the "intruder word". The test garnered an average score of 55.71%, showing a good connection between words found on the topics.

Extending their work, word2vec was implemented to find out word similarities above 60% threshold and come up with narratives (or concepts). The narratives that were distinguished are "Community preparedness for emergency", "Helping the barangay in clean-up drive", and "Awareness through seminars and information". In evaluating the result, the cosine similarity was computed, resulting in an average of 0.9020; and a word2vec analogy was listed, for instance, 'community' + 'disaster' - 'preparedness' = 'emergency' means a community experiencing a disaster without preparedness (or being prepared) will result into an emergency, or 'community' + 'bagyo' - 'tulong' = 'disaster' meaning a community experiencing 'bagyo' (typhoon) without 'tulong' (help) will result into a disaster.

Analyzing the approaches and results, bi-term topic modeling consists of words with low distinction compared to the other produced models, which means that it can result into closely similar topics such as "Typhoon preparation in a barangay" and "Disaster preparation in every barangay". With this, one recommendation is to provide a generalization function that can merge these ideas into one automatically. Another, playing around the topics by filtering them into distinct ones to produce a variety in topics that may raise other important concerns.

The results on the other hand, can be useful in coming up with generic and specific categories that most responses are under, mainly for organizing or summarizing the data. It can also be used to find key relationships or measure importance in their given subject that are evident in community responses.

Table 2.3: Comparative Summary of Other Disaster Data Analysis

Related Literature	Resource/s	Approaches	Evaluation
A Classifier Module for Analyzing Community Responses on Disaster Preparedness (De La Cruz et al., 2017)	763 Disaster Preparedness Responses	Support Vector Machine Classifier and TF-IDF Weighting	10-fold Cross Validation: 93.30% and Social Scientist: 63.80%
Qualitative Data Analysis of Disaster Risk Reduction Suggestions Assisted by Topic Modeling and Word2vec (Gorro et al., 2017)	976 Disaster Risk Reduction Responses	Bi-term Topic Modeling and Word2Vec (Word Embeddings)	Word Intrusion Test: 55.71% and Cosine Similarity: 90.2%

2.3 Platforms for Sourcing Disaster Data

This section discusses data sources in disaster domain. It includes community data, online data, and historical data. These collections contain textual and numerical data that can be used for processing and visualizing information.

2.3.1 Malasakit

A cross-platform web survey application called *Malasakit* (B. M. Nonnecke et al., 2017; B. Nonnecke et al., 2018) was developed with a vision of making Philippine communities a part of Disaster Risk Reduction (DRR) planning and response. It was developed using HTML5, Django web framework and SQLite database, with minimal user interface design. In *Malasakit 2.0* (B. Nonnecke et al., 2018), voice recognition was added to accommodate participants with visual impairment.

It has two parts, quantitative and qualitative feedback gathering that pertains to DRR which are available in numerous languages/dialects. Quantitative feedbacks taken are about the participant's demographics (personal assessment) and his/her agreement with presented issues and DRR strategies. An example for this is the statement, “I have suffered the consequence of a typhoon or flood”, for which the participant has the choice to enter a number from 0 (strongly disagree) to 9 (strongly agree).

On the other hand, qualitative feedbacks consist of an open-ended question, “How could your Barangay help you better prepare for a disaster” for *Malasakit 1.0* and “How can your barangay better help vulnerable groups such as elderly, women, children, and persons with disabilities during typhoon or flood?” for *Malasakit 2.0*. Based on the questions, *Malasakit 1.0* is geared towards DRR strategies, while *Malasakit 2.0* is towards vulnerable groups. Apart from asking these questions, a collaborative evaluation for other participants' answers is included.

Malasakit 1.0 was tested on Philippine local communities around eight locations, namely Caloocan, Cebu, Davao, Iloilo, Legazpi, Malabon, Manila, and San Mateo City. 12 Field tests were conducted, totaling to 998 participants with over 7,157 evaluations, 2,481 collaborative evaluations, and 896 qualitative feedback entries. *Malasakit 2.0* on the other hand, gathered 1,582 evaluations, 950 collaborative evaluations, and 280 qualitative feedback entries from 261 participants around Sampaloc, Manila and Simon of Cyrene (persons with disability organization), Albay, Bicol. After gathering responses from communities, *Malasakit* allows the data to be exported in a comma-separated values file for textual analysis.

Results under quantitative feedbacks show correlations with three pairs of ideas: first, having an early warning system would help a barangay in terms of response; second, participants that are highly prepared for disasters have participated in disaster drills and clean-up drives, proper coordination in their family, and ample amount of food and water supplies to withstand a disaster; and last, support within the community entails better disaster response in the barangay.

Under qualitative, collaborative evaluation produced the top-rated suggestions for DRR strategies. Ideas presented in *Malasakit 1.0* were revolving around waste management (cleaning drainages to lessen or prevent floods) and early warning system (alerts and forecasts). In *Malasakit 2.0*, suggestions focus on bringing persons with disabilities to evacuation centers immediately and necessity of having rescue teams in each barangay.

In this work, it is evident that they have already produced an application to collect data and was able to analyze the data, presenting general ideas from local communities that are valuable information specially to improving their handling, mitigation and prevention to disaster-related issues. However, their work did not highlight specific suggestions provided by their respondents and lacks the capability to disseminate information directly to decision makers (the ones who are better equipped to providing the needs of the people). In this case, it now becomes an integral part to act on the insights given. Provided that access, it would be beneficial to connect the people with governing bodies, and be able to handle, extract, organize and analyze the responses automatically.

2.3.2 Twitter

Twitter, a social networking platform, is widely known to be a good source of information. As users call it, *tweets* can be used to post and engage with information related to any topic, may it be from personal messages to news. An added feature enables users to include *hashtags* (#) for grouping and keeping track of threads that pertain to a topic.

Researches took advantage of this platform by gathering information from tweets. Several applications looked into disaster-related contents in the hopes of presenting relevant information that can aid disaster management planning and relief efforts. One existing work, classified typhoon-related tweets to find out information under resource coordination, urgent rescue needed, urgent rescue resolution, damage reporting, missing people, and media storm coverage (Lam, Paner, Macatangay, & Delos Santos, 2014). For this research, they gathered 2,356 tweets. Another, attempted to classify tweets based on disaster experienced by a

user such as flood, earthquake, fire, landslide, and the likes (Beduya & Espinosa, 2014). In this case, they made use of at least 17,000 tweets containing a variety of disaster types.

A different approach was taken by Regalado et al. (2015), which classified relevant and irrelevant disaster tweets and extracted details such as typhoon signals, suspension of classes, casualties, damages, and items donated. Similar to typhoon classification, only 2,000 tweets were used for this study.

Altogether, their works exhibit ease in accessing and collecting twitter data but showed challenges in processing them. Moreover, contents of their research were limited to proving the feasibility of their studies, the development of their application, and discussion of issues and challenges. Distribution of information, however, was not presented.

2.3.3 NOAH

Nationwide Operational Assessment of Hazards (Project NOAH) is a disaster prevention and mitigation program in the Philippines that enables government agencies “to provide a 6 hr lead-time warning to vulnerable communities against impending floods and to use advanced technology to enhance current geo-hazard vulnerability maps” (Lagmay, Racoma, Aracan, Alconis-Ayco, & Saddi, 2017). They developed a Web-based Disaster Geographic Information System (Web-GIS) for information dissemination, which is a good example of visualization tools for disasters. The Web-GIS consists of six visualization components ranging from land to sky technologies that uses state-of-the-art technologies and historical data to collect and generate near-real-time information.

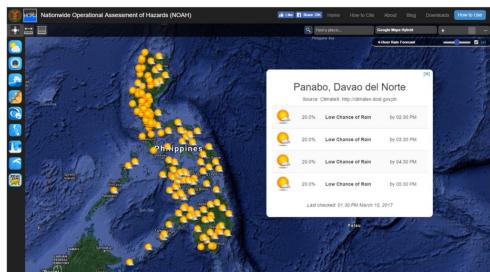


Figure 2.1: Rainfall Probability

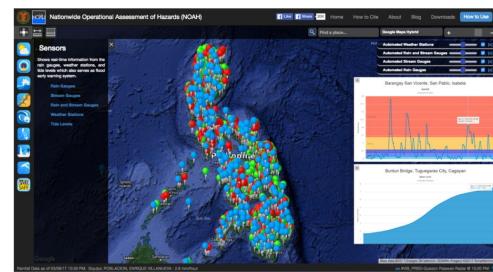


Figure 2.2: Weather and Water-level

First is an estimation of rainfall probability (see Figure 2.1) made for disaster preparedness. It has an accuracy of 82.68% and forecasts are updated every hour up until the next 4 hours. It displays icons that represents 0-20%, 20-30%, 30-40%, 40-60%, and 60-100%.

Second are weather and water-level sensors (see Figure 2.2). It displays color-coded pins such as light blue for light rainfall intensity, blue for moderate, dark blue for heavy, orange for intense, and red for torrential rain. In addition to this, the visualization includes graphs providing rainfall intensity values.

Third, Light Detection and Ranging (LiDAR) and radar-derived topography. These two provide high-resolution topography maps of the Philippines in a 1:5000 and 1:10,000 scale, respectively. Provided the scales, it enables the other components such as the flood and hazard maps to view and identify the danger in a community scale instead of regional.

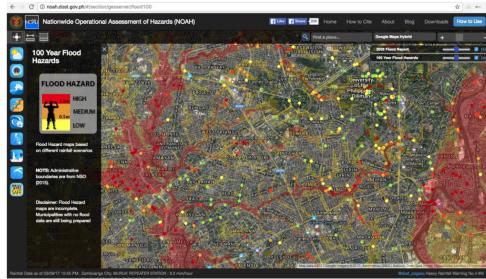


Figure 2.3: Flood Simulation



Figure 2.4: Landslide Mapping

Fourth, 1- and 2-dimensional flood simulations and flood events crowd sourcing (see Figure 2.3) for local emergency responses and infrastructure overhaul. Simulations provide flood hazard levels in no flood (green), low (yellow), moderate (orange), and high (red) metrics based on the typical height of a Filipino (167.64 cm), modelled from Manny Pacquiao, a well-known boxer. Data used came from 5- to 100-year rainfall returns, major rivers prone to extreme flooding by the Department of Public Works and Highways (DPWH), and citizen reports.

Next, landslide inventory, simulations and monitoring (see Figure 2.4) for disaster awareness and preparedness. The inventory has 96.07% accuracy, which means its entries mostly contain areas that are prone to landslides. These areas are mapped by placing three colors: red for the most dangerous sites, orange to be most likely susceptible to landslides, and yellow as the least susceptible. Areas without color impose safety and are good for infrastructure.

Last, storm surge simulation and hazard maps (see Figure 2.5) that are classified with advisory levels 1, 2, 3, and 4, measured by 2-, 3-, 4-, and 5-meter storm height. Specifically, it gives details on “... predicted flow depths, velocities, discharge hydrographs, dynamic and static pressure, specific energy, and area of inundation.”

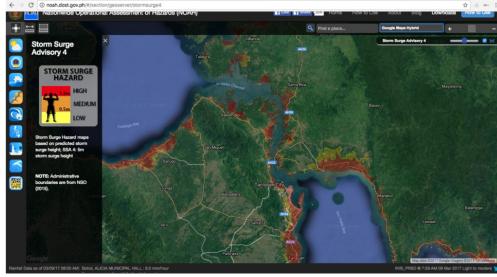


Figure 2.5: Storm Surge Simulation and Hazard Map

Generally, this work presents a number of visualizations that can be useful in wide disaster-related applications. There can still be a lot of changes and improvements to features, interface, accuracy and performance. Since the country experiences a lot of disasters, one idea is to keep on feeding more real data to make better warning and simulations, as accuracy and performance should be the priority. Specifically for this research, this work contributes by providing ideas on how visualizations could be applied on various disaster data or information. Ideally, displaying information on a web-based platform.

Table 2.4: Comparative Summary of Disaster Data Platforms

Platforms	Data and Resource/s	Technologies	Related Literature
Malasakit	<p>Malasakit 1.0 Community Responses: 7,157 evaluations, 2,481 collaborative evaluations, and 896 qualitative feedback entries</p> <p>Malasakit 2.0 Community Responses: 1,582 evaluations, 950 collaborative evaluations, and 280 qualitative feedback entries</p>	<p>HTML5, Django Web Framework, SQLite Database and Twilio audio-based interactive voice response</p>	<ul style="list-style-type: none"> • Malasakit 1.0: A Participatory Online Platform for Crowdsourcing Disaster Risk Reduction Strategies in the Philippines (B. M. Nonnecke et al., 2017) • Malasakit 2.0: A Participatory Online Platform with Feature Phone Integration and Voice Recognition for Crowdsourcing Disaster Risk Reduction Strategies in the Philippines (B. Nonnecke et al., 2018) • A Classifier Module for Analyzing Community Responses on Disaster Preparedness (De La Cruz et al., 2017) • Qualitative Data Analysis of Disaster Risk Reduction Suggestions Assisted by Topic Modeling and Word2vec (Gorro et al., 2017)
Twitter	Disaster Tweets (author collection varies)	<p>Tweet Crawler, Preprocessing Techniques</p>	<ul style="list-style-type: none"> • Classifying Typhoon Related Tweets (Lam et al., 2014) • Disaster-Related Participant Tweet Identification using SVM (Beduya & Espinosa, 2014) • FILIET: An Information Extraction System for Filipino Disaster-Related Tweets (Regalado et al., 2015)

Table 2.4 continued from previous page

Platforms	Data and Resource/s	Technologies	Related Literature
NOAH	Historical Data: Rainfall, Weather, Water-levels, Storm Surge, Hazard, Flood, Landslide, and Topography Maps	Web-GIS and various data collection technologies (e.g., weather stations, rain gauges, LiDAR, etc.)	<ul style="list-style-type: none"> • Disseminating Near-real-time Hazards Information and Flood Maps in the Philippines through Web-GIS (Lagmay et al., 2017)

Chapter 3

Theoretical Framework

This chapter discusses theories and concepts related to the study, which consists of topics related to Information Extraction, Clustering, and Filipino Part-of-Speech.

3.1 Information Extraction

Information Extraction (IE) automatically “turns the unstructured information embedded in [natural language] texts into structured data” (Jurafsky & Martin, 2018). It has been widely used by various fields such as Data Science, Media, Finance, Law, and Medicine that makes use of the structured information for analyses and dissemination. Based from Grishman (1997), the IE process has three parts. First, extracting “facts” from a given text; next, integrating facts to produce a larger (or new) facts; and last, transferring these new facts into a certain format. In this section, approaches, tasks and subtasks related to developing IE systems are discussed.

To give an example, input and output of IE using a news article was provided by Jurafsky and Martin (2018):

Citing high fuel prices, United Airlines said Friday it has increased fares by \$6 per round trip on flights to some cities also served by lowercost carriers. American Airlines, a unit of AMR Corp., immediately matched the move, spokesman Tim Wagner said. United, a unit of UAL Corp., said the increase took effect Thursday and applies to most routes where it competes against discount carriers, such as Chicago to Dallas and Denver to San Francisco

information like lead airline, new fare, effective date, and follower airline can be extracted from the article, resulting in ‘United Airlines’, ‘\$6’, ‘Thursday’, and ‘American Airlines’, respectively.

3.1.1 Approaches

Stated by Appelt (1999), “there are two basic approaches to the design of IE [Information Extraction] systems, which we label as the *Knowledge Engineering* [Rule-based] Approach and the *Automatic Training* [Machine-learning] Approach.” Knowledge engineering revolves around formalizing rules from data (specifically a corpus), whereas automatic training runs an algorithm on an annotated data in order to learn patterns or its set of rules. Alternatively to these two, an ontology, containing concepts and relationships on a particular domain can also be used for IE. In developing IE systems, it is important to note that not all components or sub-tasks has to follow the same paradigm, which even allowed a combination of rule-based, machine learning and/or ontology in a single task to produce an entirely new *hybrid* approach.

Knowledge Engineering (Rule-based)

By their definition, researchers ideal process is pointed towards automatic training since knowledge engineering requires iterative rules modification to have a high-performance IE system. Despite that, knowledge engineering should not be taken lightly as hand crafting rules can tackle problems that would be difficult to automatically process or learn, that is including patterns that have yet to be encountered by the system.

Advantages notable to knowledge engineering suggests ease in implementation and control – one can develop good extraction rules with “right level of generality”. At the same time, disadvantages require repetitions in testing and debugging, as well as being limited to the engineer’s skills and knowledge in the domain or one’s ability and speed in crafting rules. Having pointed out advantages and disadvantages, knowledge engineering is best used when linguistic resources and rule writers (engineers) are available, resources are scarce or expensive, IE specifications has a great chance of changing (especially when frequent), and highest possible IE system performance is needed.

Automatic Training (Machine Learning)

Automation of the process done by knowledge engineering through analysis of data takes away strain on the engineer. In this process, human expertise is not necessarily required, since it focuses on deriving rules within the data. As long as annotators abide on the general knowledge (providing proper annotations like indicating person or organization names), automatically trained IE systems can be developed. Furthermore, in deriving rules, automated training covers all of the data – running through each entry in rule generation. A disadvantage in this approach however, indicates dependency of the system with the data, limiting its knowledge to entries used in training it. Another, automatic training requires acquiring large volumes of data. Considering challenges in annotating, guidelines has to be created to prevent confusions as to person names that can be company names or non-profit organizations that can be confused with companies. It is then important to note the consistency of annotations in the data.

Moreover, mistakes on the data or changes in specifications may require heavy modifications (applied to a lot of entries) and retraining, instead of directly modifying or adding rules. One example is initially having specifications on extracting location names such as countries, cities, and states. If the specification is changed by adding the scope of extracting landforms (i.e., mountains, hills, plateaus, and plains), entries in the data will have to be reannotated to accommodate the change.

Provided with advantages and disadvantages in automatic training, it is best to be used when linguistic resources are unavailable, raw texts are available, training data is large and easily (affordable) acquired, IE specifications are constant or stable, highest possible IE system performance is optional.

3.1.2 Tasks and Subtasks

In general, there are five tasks, components or steps in developing an IE system, namely *named-entity recognition (NER)*, *relation extraction*, *temporal expression*, *event extraction*, and *template filling* (Jurafsky & Martin, 2018). Furthermore, Grishman (1997) added *lexical analysis* before performing the NER process and included processes such as *coreference resolution*.

An IE system can be shaped based on a particular problem by adding or reducing the tasks. There are instances where preprocessors are included; in fact, *FILIET* (Regalado et al., 2015) applied preprocessors to clean disaster-related tweets and only utilized NER and lexical analysis for rules crafting and IE.

Preprocessing

Before undergoing Information Extraction, data must be prepared by reducing noise or dirt in data to prevent loss of performance. There are four preprocessing for texts that can be used, namely *Cleaning*, *Tokenization*, *Sentence Splitting*, and *Standardization/Normalization*.

Cleaning removes unwanted characters or words that are present in the data. For example, data from social media can contain emoticons (expression symbols) or Uniform Resource Locators (URLs) that may hinder machine processing, which the solution is by omitting these characters.

Tokenization simply separates words in the data from punctuations or symbols (normally separated through whitespaces) forming *tokens*. One example using the sentence “Good day, how may I help you?”, the tokens would be ‘Good’, ‘day’, ‘,’ , ‘how’, ‘may’, ‘I’, ‘help’, ‘you’, ‘?’.

Sentence Splitting separates contents in a paragraph into individual sentences to avoid processing large chunk of texts. It is done through regular expressions, using punctuations or end of line (EOL) as markers. There are some instances that instead of sentences, a huge chunk of text is split into paragraphs.

Standardization or *Normalization* transform words into a given, consistent format. An instance of this is reducing a calendar date (e.g., January 1, 2018) into MM/DD/YYYY (i.e., 01/01/2018) format or vice-versa. Another, expands or corrects a shortened variation (e.g., abbreviated, truncated, or phonetically substituted) or typographical error in words (e.g., from gr8 to great), which normally exists in social media and informal domains.

Lexical Analysis

In analyzing lexicons, *Part-of-Speech (POS) Tagging* and *Morphological Analysis* are the main processes for structuring the input data. These processes provide additional details or *features* to words that will help in the overall language analysis, determining the right (or necessary) information to extract.

POS Tagging labels words with their corresponding POS such as Noun, Pronoun, Adjective, Adverb, Determiner, etc. These tags have certain formats for which standards are abbreviated like *Penn Treebank* by Marcus, Marcinkiewicz, and Santorini (1993) for English and *MGNN Tagset* by Nocon and Borra (2016) for Filipino (see Appendix A). A POS tagger can be implemented through rule-based,

machine learning (commonly statistical or sequence models), and hybridized approaches. Primary challenge for any of the approaches is on how to handle lexical ambiguity, that is a word having multiple POS candidates. Regardless, implementing this component and using POS tags as features can surface language patterns that will be helpful in rule creation, learning, or machine process of IE systems.

Morphological Analysis pertains to finding out a word's structure or formation. In this process, root words are discovered and variations (inflected, derived, and compounded words) are explored. Two specific techniques adhere to this idea, namely *Stemming* and *Lemmatization*, which both processes the root word.

Stemming formulates the structure of a word by chopping off characters (features) outside of the root word, specifically the affixes attached to the inflected word. For example, the Filipino term *magtatago* ‘will hide’, the prefix *mag-* and partial reduplication *ta-*, will be chopped off to get *tago* ‘hide’ as the root word. Moreover, this process is often visualized through character branches to show the structure of the word and how the root word was extracted. Issues encountered in stemming include over-stemming and under-stemming.

Lemmatization on the other hand, extracts the root word through the use of vocabulary or dictionary. A vocabulary often contains the lemma and its variations or forms. Some may provide additional information such as morphological information or descriptions (e.g., grammatical tense, point-of-view, POS, and more). Issues encountered include being limited to the number of entries present in a vocabulary (out-of-vocabulary), not capturing word features or root word outside the vocabulary.

Differentiating the two through an example, stemming the word ‘occupied’ may result into the root ‘occupi’ by removing *-ed* prefix, whilst lemmatizing the word may result into ‘occupy’ or ‘occupied’ if it does not exist in the vocabulary. Furthermore, stemming does not require much of morphological analysis as it can reduce affixes in a word without other heuristics; thus, can be faster compared to lemmatization but may need intervention as to fixing broken root words.

Named-entity Recognition (NER)

NERs as defined by Jurafsky and Martin (2018), is the process of detecting and labeling entities in a text, where entities are “... roughly speaking, anything that can be referred to with a proper name”. Common entities that are labeled include a person, location, organization, date, and time. For NER, there is a standard format in labeling such as PER for person, ORG for organization, LOC for lo-

cation, etc. A factor needed to take note of is the ambiguity of an entity. Some entities can have multiple labels that has to be considered for accurately providing the right type. NERs can be developed through lists, rules, and supervised machine learning (normally using sequence models). Terminologies related to these techniques include *IOB tagging*, *word shapes*, and *gazetteer*.

IOB Tagging pertains to features that labels a word as inside (I), outside (O), and beginning (B) of an entity. For example, the phrase “United Airlines said Friday it has increased fares...” is labeled as B, I, O, B, O, O, O, and O, where ‘United’ is the beginning of the entity and any word part of it is an inside like ‘Airlines’; as to ‘Friday’, it is another entity and is labeled as B.

Word Shapes, primarily used for labeling unknown words, are features that corresponds to letter type or case patterns. Formatting for word shapes retain punctuations or symbols, and replaces uppercase letters with ‘X’ marker, lowercase letters with ‘x’, and numbers with ‘d’. One example, the flight “AirAsia Z2-420” will have the corresponding feature of “XxxXxxx Xd-ddd”. Additionally, shortened word shapes can be used as features by reducing consecutive duplicate shapes, turning shapes from the previous example into “XxXx Xd-d”.

A *Gazetteer* is a resource used by NERs, containing a list of entities such as person, location (i.e., countries, cities, landmarks, shops, and more), product (e.g., biological, mineral, commercial, and more), and organization (with extensions or abbreviations) names. It can also include dates such as January, February, Sunday, Monday, and the like.

Coreference Resolution

Coreference Resolution or Analysis as stated by Grishman (1997) “... has the task of resolving anaphoric [an expression depending on another expression] references by pronouns [e.g., he, she, it, this, that, etc.] and definite noun phrases.” In the sense, coreference resolution finds and links the entity referred to with the corresponding pronouns and noun phrases (applicable also on abbreviated mentions like ‘Burger King’ and ‘BK’). For example, given the sentence “The neighbors love Malcolm, they made a cake for him.”, through coreference resolution, the word ‘they’ should be linked with the neighbors, whereas ‘him’ should be linked with Malcolm. A more complicated example involves noun phrases; for instance, “The kid fought off the robber at the store.” followed by the sentence “The hero deserves his praise.”, the kid and hero should be linked with each other.

Relation Extraction

As the term suggests, relation extraction is the process of finding and identifying connection/s (semantic relations) between extracted entities. It can be developed through hand-written patterns, supervised, semi-supervised and unsupervised machine learning. Generally, relations are in binary, meaning two entities are joined together by a single relation.

Illustrated by Jurafsky and Martin (2018) at Figure 3.1 are sample relations to particular entities (in blue), where samples (in orange) of generic ones are *part-of* and *employs* relationship. Furthermore, there are several domains that established relation sets or a list of binary relations such as the Unified Medical Language System (UMLS) which provided relations between substances, organisms, function, structures and many more, and Wikipedia's infoboxes which are structured facts in the form of *category = "value"* (e.g., state = "California" or president = "John L. Hennessy").

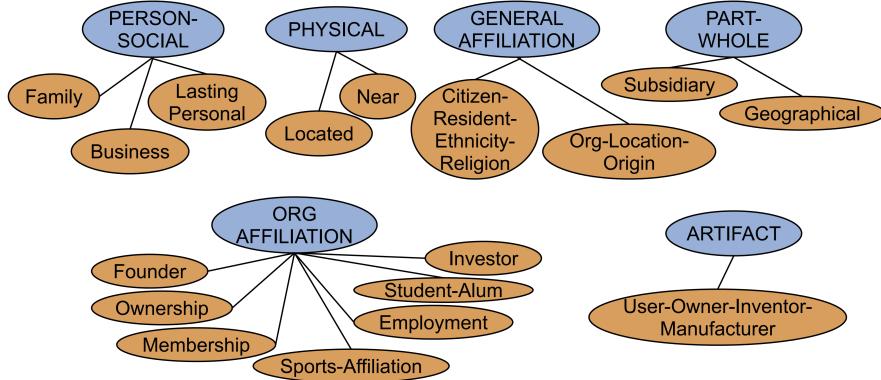


Figure 3.1: Relation Network Samples

A concrete example in applying relation extraction given “American Airlines, a unit of AMR Corp., immediately matched the move, spokesman Tim Wagner said.”, indicates entities such as ‘American Airlines’, ‘AMR Corp.’, and ‘Tim Wagner’, that are organization, organization, and person entities, respectively. Using the entities, relations can be found such as: ‘American Airlines’ is a part of ‘AMR Corp.’ and ‘Tim Wagner’ is affiliated with the organization ‘American Airlines’. In line with this, the relations can also be interpreted as ‘American Airlines’ is a unit of ‘AMR Corp.’ and ‘Tim Wagner’ works for ‘American Airlines’.

In pairing entities with a relation, there are formats that can be followed. One format is a model-based view wherein it contains domain, classes, and relations. The domain (D) is a set of entities that are represented as variables, illustrating the set show $D = a, b, c, \dots$; classes are entities grouped per type like Organization

(Org) = a, b and Person (Pers) = c; and relations are represented in a format like *PartOf* (relation) = $\{a,b\}$ (a is a part of b) and OrgAff = $\{b,c\}$ (c is affiliated with b). Another format is called *Resource Description Framework (RDF)* triple where entities and relations are formatted through a subject-predicate-object expression. An example for this is the *DBpedia* resource with over 2 Billion RDF triples.

In the light of relational resources, *WordNet* offers a hierarchical representation of relations between words. It has a “... is-a [or Instance-of] or hypernym relation between classes, Giraffe is-a ruminant is-a ungulate is-a mammal is-a vertebrate is-a animal...” (Jurafsky & Martin, 2018). Provided with resource such as WordNet, data for relation extraction can be increased.

Temporal Expression

Temporal Expression extraction pertains to capturing time and date entities. There are three types of temporal expression, namely absolute, relative and duration. *Absolute temporal expressions* refer to time and dates “... that can be mapped directly to calendar dates, times of day, or both” (e.g., ‘January 1, 2018’ and ‘12:00PM’), whereas *relative temporal expressions* refer to “... particular times through some other reference point” (e.g., yesterday, next semester, and last quarter) and *duration* “... denote spans of time at varying levels of granularity (seconds, minutes, days, weeks, centuries, etc.)” (Jurafsky & Martin, 2018).

Temporal expressions can be found through *lexical triggers* which are nouns, proper nouns, adjectives, and adverbs pertaining to time and date. Examples for this are the words ‘morning’, ‘January’, ‘recent’, and ‘hourly’, each with respect to the part-of-speech given. It can be automatically detected or recognized through rule-based approaches, particularly making use of patterns (automata or regular expressions). Another, uses sequence-labeling by implementing the *IOB scheme*, where *B* indicates the beginning of a temporal expression, *I* is any word or number that follows it but has to be a part of the temporal expression, and *O* are words that are not temporal expressions.

As temporal expressions are extracted, these entities have to undergo normalization. Temporal expressions are standardized into a specific point in time (relative are turned into absolute) or duration under a certain format. A format used is ISO 8601 standard for encoding temporal values. For this format, dates are represented through YYYY-MM-DD, weeks are in YYYY-Wnn (*n* refers to week number), durations in Pnx (*n* refers to the length and *x* refers to the unit), clock times in HH:MM:SS, financial quarters in Qn (*n* refers to quarter number), and many more other representations within the standard.

Going deeper than formatting, normalization enables computation of relative and duration temporal expressions. Through a temporal anchor, a reference point like document date or publish date, can be used to add and subtract days from terms such as tomorrow or yesterday. Calculating dates however are not as simple as it is, since there are ambiguities that may be present. For instance, the phrase “the weekend” can refer to the weekend that passed or the coming weekend.

Event Extraction

An event as defined by Jurafsky and Martin (2018) “... is any expression denoting an event or state that can be assigned to a particular point, or interval, in time.” Usually in English, verbs like ‘increasing’ and noun phrases ‘the increase’ contain events. However, there are also times that verbs do not indicate one. One instance is “took effect”, for which when the event took effect is indicated and not the event exactly. With this in mind, a lot of words and features have to be considered. That is why machine learning has been the primary approach in detecting and modeling, as such process can handle and take into account features like character affixes, nominalization suffix (e.g., *-tion*), part-of-speech, light verbs (e.g., make, take, have, etc.), subject syntactic category, morphological stem, verb root, and WordNet hypernyms.

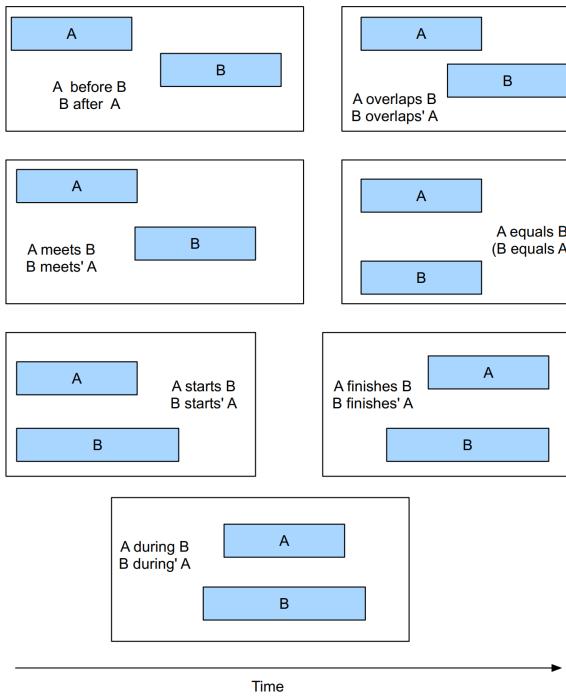


Figure 3.2: Temporal Relations Order

Having events and temporal expressions extracted, the information together can be used to make a timeline of events. For this, events are to be determined whether they are pointing to the same event and their order. It involves the process of using binary relation detection and classification techniques, identifying which temporal relations shown at Figure 3.2 match the events.

Template Filling

In an IE System, visualization of the output is an essential task. It may be done through simple, static labels or through an advanced technique by finding appropriate labels per information. Template filling refers to the task of finding “... documents that invoke particular scripts [templates] and then fill the slots in the associated templates with fillers [information] extracted from the text.” (Jurafsky & Martin, 2018).

Basically in this task, the appropriate template has to be selected and how the output is to be presented is formulated, in order to fill in the template with the extracted information under the right labels. Performing advanced template filling mentioned statistical approaches which involves template recognition and role-filler extraction.

3.1.3 Evaluation Metrics

There are various ways in building an Information Extraction (IE) system. Standard metrics are used to measure the overall performance for proper evaluation and comparison – differentiating each IE system from each other. Metrics primarily used are *Accuracy*, *Precision*, *Recall*, *F-Measure*, and *Kappa Statistic*.

Accuracy is the “closeness of agreement between a measured quantity value and a true quantity value of a measurand” (BIPM, IFCC, IUPAC, & ISO, 2012). In this case, the measured value is represented through a given system’s output, whilst the true value is an expected output, normally given by respective annotators or experts as the *gold standard*.

Comparing two values for instance, if the result is the same (or above a threshold), the score will increase, whereas having different values (or below the threshold) will not increase nor decrease (unless penalty is applied) the score. Normally, the equation for accuracy counts the number of correct instances divided by the total number of instances (multiplied by a hundred to get the percentage value).

Provided with the system's output and true value (gold standard) in an Information Retrieval setting, **Precision** or **Positive Predictive Value** is computed through the number of relevant documents (correctly) retrieved by the system over the total number of documents retrieved (Hripcsak & Rothschild, 2012). **Recall** or **Sensitivity** on the other hand, are relevant documents that has been retrieved by the system among relevant documents, or relevant documents retrieved over total number of relevant documents. An example formula for the two metrics are:

$$Precision = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|} \quad (3.1)$$

$$Recall = \frac{|\{\text{relevant Documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{relevant documents}\}|} \quad (3.2)$$

Precision (Equation 3.1) provides the idea as to the number of documents retrieved by the system which are relevant, but are only limited to it; thus, does not provide if all relevant documents has been retrieved. Recall's (Equation 3.2) idea is also limited, as it represents the relevant documents that has been successfully retrieved but does not represent the rest of the documents retrieved that are irrelevant (Dietrich, Heller, & Yang, 2015).

F-Measure, also known as **F-Score** or **F1 score** (Equation 3.3), is the harmonic mean of Precision and Recall. The 1 in F1 is a value for β , a weight for precision or recall.

$$F = \frac{(1 + \beta^2) \times precision \times recall}{(\beta^2 \times precision) + recall}; F1 = \frac{2 \times precision \times recall}{precision + recall} \quad (3.3)$$

Kappa Statistic or **Kappa Coefficient (κ)** measures the agreement between observers or raters, annotating or "...evaluating the same thing" (Viera & Garrett, 2005). It is a measurement for consistency, since multi-observer/raters often provide variations into their assessments. The Kappa statistic ranges between values -1 and 1. A kappa with the value of 1 pertains to an absolute agreement, 0 for agreement by chance, and negative values for agreement less than chance (possibly no agreement).

In calculating the Kappa statistic (Equation 3.4), the formula is represented by "the difference between how much agreement is actually present ('observed' agreement [p_o]) compared to how much agreement would be expected to be present by chance alone ('expected' agreement [p_e])" (Viera & Garrett, 2005).

$$\kappa = \frac{p_o - p_e}{1 - p_e} \quad (3.4)$$

A table of agreement is shown at Table 3.1 with given values for a sample calculation for Kappa Statistic. Raters A and B were asked to evaluate which among N (100) lectures were useful or not.

Table 3.1: Table of Agreement between Two Raters

Rater A \ B	Yes	No
Yes	(a) 15	(b) 05
No	(c) 10	(d) 70

Values (a) and (d) are the ones that both raters agree on, which are also the values needed for computing the observed agreement:

$$p_o = \frac{a + d}{N} = \frac{15 + 70}{100} = \frac{85}{100} = 0.85$$

On the contrary, values (c) and (d) are the ones which raters disagreed on and is used for computing the expected agreement by adding the probability of raters answering yes and no:

$$p_e = \left(\frac{a+b}{N} \cdot \frac{a+c}{N} \right) + \left(\frac{c+d}{N} \cdot \frac{b+d}{N} \right) = \left(\frac{20}{100} \cdot \frac{25}{100} \right) + \left(\frac{80}{100} \cdot \frac{75}{100} \right) = 0.65$$

Given the two variables, the Kappa Statistic can be computed:

$$\kappa = \frac{p_o - p_e}{1 - p_e} = \frac{0.85 - 0.65}{1 - 0.65} = 0.57$$

3.2 Clustering

Clustering as defined by Gan, Ma, and Wu (2007), is “a way to create groups of objects, or clusters, in such a way that objects in one cluster are very similar and objects in different clusters are quite distinct.” Strongly established, the main criterion for grouping is through similarity or association. Clustering was firstly introduced by Driver and Kroeber (1932) for anthropology, where ethnic groups were compared and clustered based on their cultural similarities. It has then been applied to numerous fields such as pattern recognition, artificial intelligence, information technology, image processing, and many more.

In general, data objects may come in a form of text, numbers, or a combination of the both. Under this section, ways to cluster words are presented. Techniques cover string similarity, as well as, semantic similarity; particularly, n-grams, Sørensen-Dice Coefficient, and word embeddings.

3.2.1 N-grams

Defined by Jurafsky and Martin (2018), “an *n-gram* is a sequence of N [number of] words”. In the phrase “all of a sudden...”, a 1-gram or *unigram* consists of the sequence with individual words like ‘all’, ‘of’, ‘a’, ‘sudden’, a 2-gram or *bigrad* is a sequence of “all of”, “of a”, “a sudden”, and so on. In some cases, n-grams can be a sequence of characters instead of words.

In the context of language modeling or the process of assigning probabilities to succeeding words, word sequences, and even whole sentences (Jurafsky & Martin, 2018), n-gram models can compute for succeeding word predictions using the conditional probability of a word w given a word history h . It is able to do so by checking the number of times h has been followed by w in the corpus – a collection of text. For example, the same phrase above will produce the equation:

$$P(w|\text{all of a sudden}) = \frac{\text{Count}(\text{all of a sudden } w)}{\text{Count}(\text{all of a sudden})}$$

However in this process, it would be difficult to run through and count instances of a sequence given large data. Imagine computing for the probability of an entire sequence alone, it will require counting the number of times a sequence came up to the corpus divided by the number of words with the same length of the sequence. Having said that, there are other ways to compute probabilities of sequences such as a *Markov* assumption.

The idea for *Markov*, instead of looking at the whole history to compute for the probability, an approximation of the history can be used by only looking at the last few words; particularly, the $N1$ previous neighboring word/s. With the same example and using a bigram, instead of computing for $P(w|\text{all of a sudden})$, it can be reduced and approximated using $P(w|\text{sudden})$. Generalizing this equation with respect to the n-gram, approximating the conditional probability of the next word in a sequence (Equation 3.5) and the whole sequence (using *chain rule of probability*, Equation 3.6) are:

$$P(w_n|w_1^{n-1}) \approx P(w_n|w_{n-N+1}^{n-1}) \quad (3.5)$$

$$P(w_1^n) \approx \prod_{k=1}^n P(w_k|w_{k-1}) \quad (3.6)$$

Furthermore, additional tasks done to the approximation is applying smoothing, normalizing the n-gram's probabilities into relative frequency (values between 0 and 1), and using log probabilities instead of raw probabilities for computations.

In the context of clustering, n-grams can be used in both semantic and string similarity. For semantic, n-gram language models can be used to provide a sense to which the words relate to each other by grouping together sequences that are close in terms of usage counts or probability values. In fact, semantic clustering can be done on not just words but also sentences and even in between dialects and languages.

For example, in terms of word order, “all of a sudden I notice three guys standing on the sidewalk” will make more sense than producing the same words in the order “on guys all I of notice sidewalk three a sudden standing the”. It means that there is a particular structure in the sentence that can be observed and followed, usable to compare and pinpoint words, phrases, or sentences that can be grouped together. Another example are “Hey, how are you?” and “Hello, how are you?” Assumingly, both should have the same or at least close probability values as it ends with the same phrase. In this case, the two sentences can be under the same cluster, possibly labeled as “greeting”, and can be further used to find other sentences related to them.

For string similarity, the concept of using n-grams alone (without the probabilities) is applied, normally adding a similarity or distance measure as criterion for comparing and grouping similar or dissimilar texts. One common process is to dissect two words with their character n-grams and compare each n-grams to compute its similarity to each other.

3.2.2 Sørensen-Dice Coefficient

Sørensen-Dice Coefficient or *Dice's Coefficient* (Dice, 1945) is used to measure orthographic similarity between words. It uses character n-grams to represent the structure of the words and calculates the similarity or distance score by counting the shared or matching n-grams. Specifically, Dice's coefficient is computed by counting the matched n-grams and multiplying it by 2, then dividing the produced value with the combined total number of n-grams present between two words. Formally, the equation for similarity, given two words x and y is:

$$S_{\text{SørensenDice}} = \frac{2 \times |ngrams(x) \cap ngrams(y)|}{|ngrams(x)| + |ngrams(y)|} \quad (3.7)$$

where $ngrams$ is a function containing a set of character n-grams (provided an n) of a text. The similarity score is bounded between zero and one and using 1 Sørensen-Dice can be used to transform the value into a distance score (how far the two words are from each other).

Applying this function on ‘compliment’ and ‘complement’ given an n of 2 (or bigram) produces the sets {co, om, mp, pl, li, im, me, en, nt} and {co, om, mp, pl, le, em, me, en, nt}, respectively. Substituting the values to the Equation 3.7 results in 0.778 similarity or 0.222 distance score.

Dice’s coefficient can be used in language modeling by using the n-grams and values to discover similarities between words or even in languages. Another application is clustering, wherein words are grouped together based on their orthographic similarities.

3.2.3 Word Embeddings

Word embeddings is a type of language modeling that “embeds” words into a vector space. Through these vectors, the meaning of a word is represented, and similar ones are embedded near each other (e.g., synonyms, tenses, gender-related, and country-capital relationships). Thus, this vector space model can be used to find syntactic and semantic relationships among words.

A visualization of the multi-dimensional vector space projected in 2D space using T-distributed Stochastic Neighbor Embedding (t-SNE) is shown at Figure 3.3 (Jurafsky & Martin, 2018). It exhibits clustered words based on sentiment polarity (i.e., neutral, positive, and negative) and at Table 3.2 shows more kinds of word similarity relationships from Mikolov, Chen, Corrado, and Dean (2013).

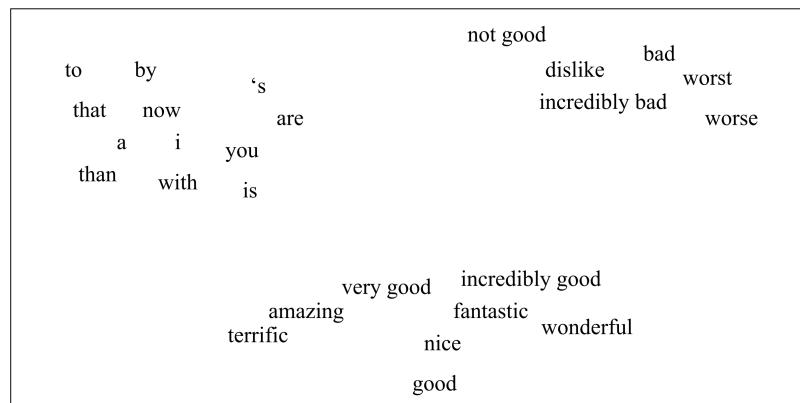


Figure 3.3: Vector Space Visualization Example

Table 3.2: Word Similarity Pair Examples

Relationship	Example 1	Example 2	Example 3
France - Paris	Italy: Rome	Japan: Tokyo	Florida: Tallahassee
big - bigger	small: larger	cold: colder	quick: quicker
Miami - Florida	Baltimore: Maryland	Dallas: Texas	Kona: Hawaii
Einstein - scientist	Messi: midfielder	Mozart: violinist	Picasso: painter
Sarkozy - France	Berlusconi: Italy	Merkel: Germany	Koizumi: Japan
copper - Cu	zinc: Zn	gold: Au	uranium: plutonium
Berlusconi - Silvio	Sarkozy: Nicolas	Putin: Medvedev	Obama: Barack
Microsoft - Windows	Google: Android	IBM: Linux	Apple: iPhone
Microsoft - Ballmer	Google: Yahoo	IBM: McNealy	Apple: Jobs
Japan - sushi	Germany: bratwurst	France: tapas	USA: pizza

Aside from explicit word similarities, another property of word embedding enables word offsets, applying algebraic operations and cosine distance in the vector space. A famous example by Mikolov, Yih, and Zweig (2013) showed gender-related offsetting like $\text{vector}(\text{"King"}) - \text{vector}(\text{"Man"}) + \text{vector}(\text{"Woman"}) = \text{vector}(\text{"Queen"})$. Another is corresponding to a word’s tense, like $\text{vector}(\text{"biggest"}) - \text{vector}(\text{"big"}) + \text{vector}(\text{"small"}) = \text{vector}(\text{"smallest"})$.

A renowned work using word embeddings for word similarity tasks was done and documented by Mikolov, Chen, et al. (2013), under the *Word2Vec* package. They established two log-linear models that captures context similarity in words from a large set of data, namely Continuous Bag-of-Words (CBOW) and Continuous Skip-gram model (architecture found at Figure 3.4). In a nutshell, CBOW’s main task “predicts the current word based on the context”, whilst Skip-gram “predicts surrounding words given the current word” (Mikolov, Chen, et al., 2013).

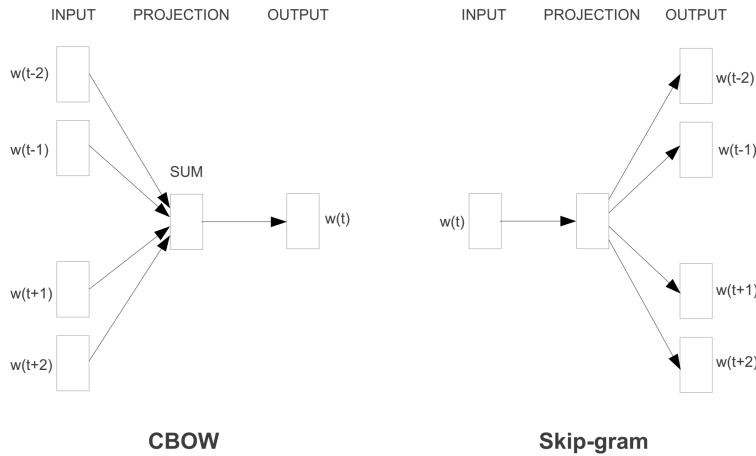


Figure 3.4: CBOW and Skip-gram Architectures

In CBOW’s architecture, the hidden layer is omitted. With this, a bag-of-words containing all word vectors are averaged and shared into the projected layer. As shown in its architecture, note that these vectors include history (i.e., $w(t-1)$, $w(t-2),\dots$) and future (i.e., $w(t+1)$, $w(t+2),\dots$) words which constitutes the context and predicts the middle or current word that is $w(t)$. Oppositely, Skip-gram’s architecture, from an input word $w(t)$ through a continuous projection layer, produces words before and after it.

An experiment comparing different models through semantic-syntactic word relationship test, resulted into CBOW and Skip-gram performing better. CBOW achieved better results in terms of syntactic tasks, while Skip-gram is better on semantics tasks. Overall, CBOW and Skip-gram’s simpler architecture produced high quality vectors.

Another well-known word embedding model incorporates global corpus statistics in its vectors, called *Global Vectors* or *GloVe* (Pennington, Socher, & Manning, 2014). Instead of just learning from neighboring words, it adds on the global statistics that can capture frequent repetitions or patterns in the data. Specifically, this statistic utilizes ratios of co-occurrence probabilities.

In comparison with Word2Vec, evaluation methods include word analogy, word similarity, and named-entity recognition tasks. In these experiments, GloVe performed better in most of the tasks, even outperforming Word2Vec’s CBOW on word similarity utilizing only less than half of the data size.

A straightforward improvement by Mikolov, Grave, Bojanowski, Puhrsch, and Joulin (2017) involves using these existing algorithms to accommodate training large datasets such as Common Crawl (630 Billion words), Wikipedia Meta-pages (9.2 Billion words), and Statmt.org News (4.2 Billion words). With this attempt, their work developed new high-quality pre-trained word and phrase representations. Learning from their implementation and analysis, it is fairly important to note the removal of duplicates in data and addition of position-dependent weights and subword features in the architectures.

Expounding on subword models, one notable work was created by Bojanowski, Grave, Joulin, and Mikolov (2017) called *FastText*. It was made to be fast, efficient, and able to handle languages with rich morphology and unknown words. As an extension of Word2Vec, specifically its skip-gram model, FastText makes use of bag-of-character n-grams for word representations.

In particular, vectors are represented by each character n-grams (ranging from three to six), and words are represented by summing up these vectors. A given example to visualize this represented the word ‘where’ with n-gram of three:

`<wh, whe, her, ere, re>`, and the whole sequence `<where>`.

Notice the special boundary symbols `<` and `>` used in the representation. These symbols signify if the character sequence acts as a prefix or suffix to a word. Having said that, the representation of ‘her’ and whole sequence pertaining to the pronoun `<her>` are different from each other.

Evaluating FastText, results showed that it outperforms other models in word similarity and analogy tasks. Even in comparison with other subword models, FastText was able to provide better, yet comparable results. In other experiments, it has been discovered that FastText does not rely on large training data, in fact was able to thrive given small datasets.

Provided with these three main developments in word embeddings, Word2Vec, GloVe, and FastText has been available in public as pre-trained word embedding models and representations. These pre-trained models were used as baseline to start researches in several Natural Language Processing tasks and there are also attempts in refining these models.

3.3 Filipino Part-of-Speech (POS)

Part-of-Speech (POS) is a categorization of words based on its grammatical function. *POS Tagging* is the process of labelling words with their corresponding POS, wherein it is guided by a collection of tags, called a *tagset*. At this time of writing, one of Filipino’s latest is called *MGNN tagset* by Nocon and Borra (2016) (see Appendix A), a version sourced from C. K. Cheng and Rabo’s (2004) tagalog tagset and English’s Penn Treebank (Marcus et al., 1993).

MGNN tagset itemizes Filipino POS and their tag counterparts. A short listing of the counterparts with examples are provided on Table 3.3. In this sample, only the main categories are provided. The actual tagset contains subcategories (subcategories count beside the tag name) that concatenates more characters at the end of the main category’s tag name. Instances of these concatenations are NNP for proper nouns, NNPA for proper noun abbreviation, PRP for personal pronouns, DTC for common noun determiner, CCP for ligatures (*pang-angkop*), CCU for preposition (*pang-ukol*), VBTS for past tense verb, JJD for describing adjective, RBJ for interjections (*sambitla*), and PMP for period punctuation. In addition to this, MGNN tagset also considers compound tags (combination of tags) and were recorded as seen on Nocon and Borra’s (2016) data. An example is the “magandang babae” ‘beautiful girl’ for which *magandang* is tagged as JJD_CCP, because *maganda* is a describing adjective that ends with *-ng* ligature.

Table 3.3: MGNN Tagset: Main Categories

POS	Tag Name	Examples
Noun (Pangngalan)	NN (4)	tao, Pilipino, Dra., Bb., km, cm
Pronoun (Panghalip)	PR (10)	ako, kami, ito, sino, dito, ganyan
Determiner (Pantukoy)	DT (4)	ang, si, ni, kay, sina,
Lexical Marker	LM (1)	ay
Conjunctions (Pang-ugnay)	CC (6)	maging, pero, kasi, ng, at, -ng
Verb (Pandiwa)	VB (14)	mag-, pwede, mayroon, -um-
Adjective (Pang-uri)	JJ (6)	maganda, pareho, mas, pinaka-,
Adverb (Pang-abay)	RB (15)	kung, dahil sa, huwag, oo, hoy
Cardinal Number (Bilang)	CD (1)	1, una, tatlo, II
Topicless (Walang Paksa)	TS (1)	Umuulan., Alas dos na.
Foreign Words	FW (1)	English, Spanish, Latin
Punctuation (Pananda)	PM (6)	‘.’, ‘!’, ‘?’ ‘;’, ‘;’, ‘#’, ‘+’, ‘(‘
Compound Tags	tag1....tagN (161)	pinakamagandang (JJCS_ JJD_CCP), ako'y (PRS_LM)

Provided with a tagset, POS tagging can be automatically done. In Miguel and Roxas' (2007) research, a comparative analysis between different Tagalog POS taggers was done. These taggers followed C. K. Cheng and Rabo's tagset and was implemented in various kinds of approaches namely template-based n-gram POS tagger (TPOST), memory-based POS tagger (MBPOST), supervised probabilistic POS tagger (PTPOST), and rule-based POS tagger (Tag-Alog).

Among these taggers, the highest performing was PTPOST with 78.3%, followed by MBPOST (77%), Tag-Alog (72.5%), and TPOST (70%). Based on the results, PTPOST's lexical and contextual probability values proved to be more effective than the other approaches which used tagging rules, sentence and lexicon collections, and word-tag database.

Revived after almost a decade, a new set of POS taggers were developed, this time for Filipino and follows the MGNN tagset. The goal was to move forward, devise and apply new approaches for POS taggers to update "... data contents, software usability, performance and availability" (Nocon & Borra, 2016).

The new set was started by Nocon and Borra (2016), which introduced the use of Statistical Machine Translation for POS tagging (SMTPOST). It was then followed by a Hybridized POS tagger (HPOST) that combined SMT, dictionary-based, regular expression-based approaches (Go, Nocon, & Borra, 2017) and Filipino Stanford POS Tagger (FSPOST) that uses maximum entropy cyclic dependency network (Go & Nocon, 2017).

Comparing the three taggers, FSPOST garnered the highest performance score with 96.15% accuracy which is considered to be the state-of-the-art in Filipino. The other two, although lower than FSPOST, got scores higher than the PTPOST, with 89.11% for SMTPOST and 91.63% for HPOST. Upon analysis, the key to FSPOST's high performance was due to an extensive use of features (e.g., word and tag context, affixes, and word shapes), lexicalization, bidirectional inference, and smoothing.

Chapter 4

Architectural Design

This chapter discusses in detail the design and implementation of the Filipino text analysis tool for disasters. Its overall task is to extract valuable information or insights from a group of text regarding disasters, organize that information, and generate a report out of it. It is packaged as an open-source Python Application Programming Interface (API)¹ or library with a collection of functions that can be usable for building software applications (see Appendix B for list). The discussion mainly includes a description of the tasks involved, content about the resources and components used, and high-level view of the algorithm. Seen at Figure 4.1 is an illustration of the tool's overall design.

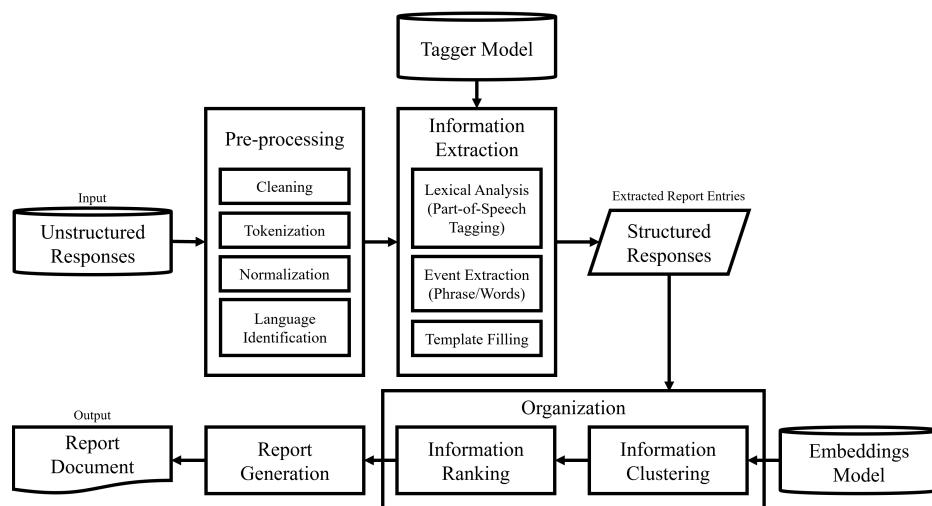


Figure 4.1: Architectural Diagram

¹Github Repository: <https://github.com/noconoccin/Filipino-Text-Analysis-Tool>
Google Drive Link: <https://bit.ly/3f56oAC>

4.1 Malasakit Community Responses

Disaster-related community responses gathered by *Malasakit* (B. M. Nonnecke et al., 2017), a cross-platform online participatory tool, were exported into an Excel (.xlsx) file. The purpose of using this data is to give the local community a chance to have a say in disaster-related issues experienced in their barangay and be part of the disaster risk reduction strategies planned by the government for the country.

The file contains a single sheet with a list of qualitative responses. Under the list are 934 instances gathered in Filipino and English, taken from different parts of the Philippines such as Caloocan, Cebu, Davao, Iloilo, Legazpi, Malabon, Manila, and San Mateo City. Each entry is comprised of two features (or columns) namely, *response* and *tag*. The *response* feature contains the ideas and was the main target for extracting and organizing information, while the *tag* (or response category) feature represents a wider, general view of the response given which was used for ranking (see Table 4.1 for sample entries).

Table 4.1: Sample Malasakit Responses

Response	Tag
Linisin ang mga baradong kanal ‘Clean up the clogged canals’	Infrastructure Maintenance and Management
Magsagawa ng programa, seminar, information drive upang magbigay ng kaalaman sa kuminidad ‘Conduct programs, seminars, and information drives to provide knowledge to the community’	Information Campaign and Capacity Building
Magkaroon ng komunikasyon kung saan magkikita sa panahon ng kalamidad. ‘Have communication on where to meet during calamities’	Preparedness for Emergency
Provide first aid kits	Disaster Relief
Prepare a place for evacuees	Community-wide Logistic Support for Disaster Response

The *response* feature (qualitative responses) contains the participant’s answer to an open-ended question, “How could your Barangay help you better prepare for a disaster”. These responses have been preprocessed which removed punctuations like commas and periods, and words are lowercased. In spite of the preprocessing, it still contains words that are misspelled (e.g., kawpa, bagy0, maayus, bahaa, esatesa, etc.), shortened (e.g., brgy, LGU, kpag, pwd, magtulong2, etc.), improperly joined (e.g. ‘drillsawareness’, 2months, etc.), and invalid suggestions (e.g. none, asdf, 3638484, blank, etc.). Some were corrected by running other preprocessing tasks such as tokenization and normalization (word standardization).

Table 4.2: Malasakit Community Responses Codebook 4.7

Category	Definition
Early Warning System	Mentions any form of communication and procedure that notifies community members of a potential hazard either face to face and/or through the means of technology (e.g., SMS, door-to-door, warning system, etc.).
Information Campaign and Capacity Building	Mentions any process of educating the community through orientations, seminars, trainings, simulations, and drills that reduce vulnerabilities and help the community to cope with hazards.
Infrastructure Maintenance and Management	Mentions a procedure to prevent and mitigate disaster hazards (e.g. maintenance of critical infrastructure like canals, river, drainage and other utilities, waste management).
Preparedness for Emergency	Mentions any plans or actions of an individual, family, or community to explicitly prepare prior to a disaster, either tangible or intangible steps.
Community-wide Logistic Support for Disaster Response	Mentions any items, infrastructure, and equipment that will respond to the needs of the community before and during the typhoon.
Disaster Relief	Mentions any act that provides goods, medicine, financial assistance, to the victims of disaster for their rehabilitation.
Local Government Accountability	Mentions specific task that assigns responsibility to the local government unit (barangay council) in their role in policy-making and policy implementation, and coordination with different agencies towards disaster risk reduction management.
Filipino values	Mentions any act that shows Filipino cultural values in times of disaster (e.g., pagkakaisa, pagtutulungan, bayanihan spirit, kapitbahayan, volunteerism, etc.).
Others	Mentions any statements not classified under the above categories (e.g., build organization, corruption, active participation, coordination, etc.).

On the other hand, the *tag* feature contains either one of the following nine *response categories*: Early Warning System, Information Campaign and Capacity Building, Infrastructure Maintenance and Management, Preparedness for Emergency, Community-wide Logistic Support for Disaster Response, Disaster Relief, Local Government Accountability, Filipino Values, and Others. Category descriptions from *Malasakit (Codebook Version 4.7 for the Malasakit qualitative responses, 2017)* can be found at Table 4.2.

As part of the tool, data utility scripts were developed to be able to read data resources like the unstructured responses and write outputs. After reading, data were placed on a *MalasakitResponse* object that consists of attributes that describes a particular response such as response ID, response text, tag, part-of-speech, insights, and more. Provided with this data structure, it may undergo the major tasks such as data preprocessing, information extraction, information organization, and report generation.

4.2 Preprocessing

Since *Malasakit* responses has already been preprocessed, this module is optional. Nevertheless, the collected data can still be tokenized, cleaned and normalized to fix responses with typographical errors and shortcut texts. Starting with tokenization, it separates words in a sentence to form ‘tokens’, while cleaning removes unwanted characters for processing such as punctuation.

Under normalization, this process follows a standardized format in texts by converting them into their original, proper forms. It involves two subprocesses: one joins unmerged prefixes (e.g., *mag karoon* into *magkaroon* ‘have’) and another corrects typographical and shortcut texts (e.g., *bkt* into *bakit* ‘why’).

For the first subprocess, an extendable prefix list from Oco and Borra (2011) was used to find and check all unmerged prefixes in the text. Once found, prefixes are joined with the next word. Based on the previous example, *mag-* is joined with *karoon*. Note that in this prefix list, those that can stand alone as a word (e.g., *i-*, *na-*, *nang-*, *magsisi-*, *pang-*, *papa-*, etc.) were removed to prevent inappropriate merging. Whereas for the second subprocess, a *Filipino Normalizer* (Nocon, Kho, & Arroyo, 2018) was utilized to automatically process the text. It covers *textspeak* (shortcuts), *swardspeak* (gay-lingo), *conyo* (upper middle class English-Tagalog colloquialism), and *datkilab* (metathesis or reversed words) styles, built from social media sources. It was implemented using Google’s Tensorflow and Statmt’s Moses but only utilized the Moses version as it is the one with the best performance.

In addition to these preprocessing tasks, a language identification module was implemented, with the intention of directing the right natural language to the Part-of-Speech tagger. An off-the-shelf tool called *LangID* or *langid.py* (Lui & Baldwin, 2012) was utilized. This tool was made using Numpy’s Naïve Bayes classifier and trained on a multi-domain language identification corpus that is comprised of government documents, software documentation, newswire, online encyclopedia and an internet crawled text. It was developed to be able to label 97 languages (follows ISO 639-1 language code in labeling) and perform well across different domains.

Applying *LangID* to the preprocessing tasks, its model was filtered to focus on two languages: English (en) and Tagalog (tl). This way, the model will have fewer choices and was deemed appropriate to consider languages that are only present in the current data. Aside from providing its language, a confidence value is part of *LangID*’s output.

4.3 Information Extraction

The responses were automatically extracted with information through Part-of-speech-based approach. In detail, responses are labeled by the *Filipino Stanford Part-of-Speech Tagger (FSPOST)* of Go and Nocon (2017). It was built from Wikipedia data using Maximum Entropy Cyclic Dependency Network, which makes use of features such as word and tag contexts, word affixes, and word shapes to determine the appropriate tag for a word. Specifically, FSPOST makes use of the following features to understand Filipino’s morphology:

- *Naacl2003unknowns* – pertains to the use of word shapes and suffixes of a word from NAACL 2003;
- *Wordshapes (-1,1)* – pertains to word shapes of the words before, currently tagged and after;
- *Left5words* – pertains to the two words before, two words after, and two tags before the word to be tagged;
- *Distsim* – or distribution similarity of words, uses the same set of features found in most tagger models of Stanford POS Tagger;
- *Pref6* – takes prefixes with lengths 1 to 6;
- *Inf2* – takes infixes with lengths 1 to 2; and
- *engNNCasFW* – a user-added feature that replaces words in the English dictionary that are tagged as common nouns NNC into foreign words FW.

Training through these features as parameters produces the tagger model, a file distributable and usable to users. Originally, FSPOST was programmed on Java programming language but a Python implementation was done through Natural Language Toolkit or NLTK (Loper & Bird, 2002). This transition still enables the usage of the tagger model without any complications. Furthermore, using NLTK facilitates switching of languages in tagging. Once the language identifier labels a response, NLTK can be used to switch from using NLTK Part-of-Speech Tagger for English or FSPOST-NLTK for Tagalog/Filipino.

Equipped with FSPOST, “events” wherein described as verb phrases or verb-noun words that contain actionable suggestions given by the community are extracted. It is done through a pattern of finding a verb and traversing through words until a noun is found (extends if there is a comma or conjunction *at* ‘and’ that is followed by another noun). A subprocess of event extraction records only the verbs and nouns in the events, forming word sets. Part-of-Speech tags to take note of are the following: VB for actions, NN and FW for nouns, and CCA and PMC for extensions such as conjunctions and commas.

It is important to highlight that the whole extraction focuses on capturing the insights regardless of the number of sentences in a response – meaning there are no limits of having a single solution per user entry and at the same time a response may not contain any insights. Regarding responses with unnormalized typographical errors or colloquialisms, and Tagalog-English code-switching, the type of tagger to use relies on the language identifier’s heuristics and its indicated label. Thus, there are instances that FSPOST encounters and tags a Filipino response with English words, as they are labeled by the identifier as ‘tl’; at the same time, NLTK tagger for English encounters and tags an English response with Filipino words, as they are labeled as ‘en’.

With regards to creating patterns on English and Filipino, NLTK uses Penn Treebank which is the basis for MGNN Tagset or the one used by FSPOST. Given that, patterns were generalized into the first two characters such as NN for nouns, VB for verbs, JJ for adjectives, etc. That said, regardless of having code-switching, both NLTK taggers would still be able to provide Part-of-Speech labels and perform pattern matching extractions.

After extraction, template filling transforms the unstructured responses into structured texts, acting as entries to the report. It works by having a set of information and plugging them into corresponding fields. Each entry will provide details about a response’s suggested action that can be found under the “proposed action” field and target of the action under “target” field. More details are discussed at the report generation module.

4.4 Information Organization

Organizing the information is comprised of two modules namely, Information Clustering and Information Ranking. There are two ways that it can be set up. One is by organizing all entries or the information entirely, clustering similar entries and being ranked based solely on frequency. Another is organizing the information by response categories, clustering similar entries under each category, ranking them by category priority and locally based on frequency.

A sample illustration can be found at Figure 4.2. The left side exhibits entries from 1 to N that are organized by descending frequencies and without being sorted by the categories, while the right side are firstly organized by categories (Information Campaign and Capacity Building is more prioritized than Others) and then are followed by entries 1 to N (again in decreasing frequency) that reside within each categories. The final report that is generated for the tool follow this illustration's template, specifically by default sorts responses per categories first then frequency. More information about the prioritized categories are discussed at the information ranking module section.

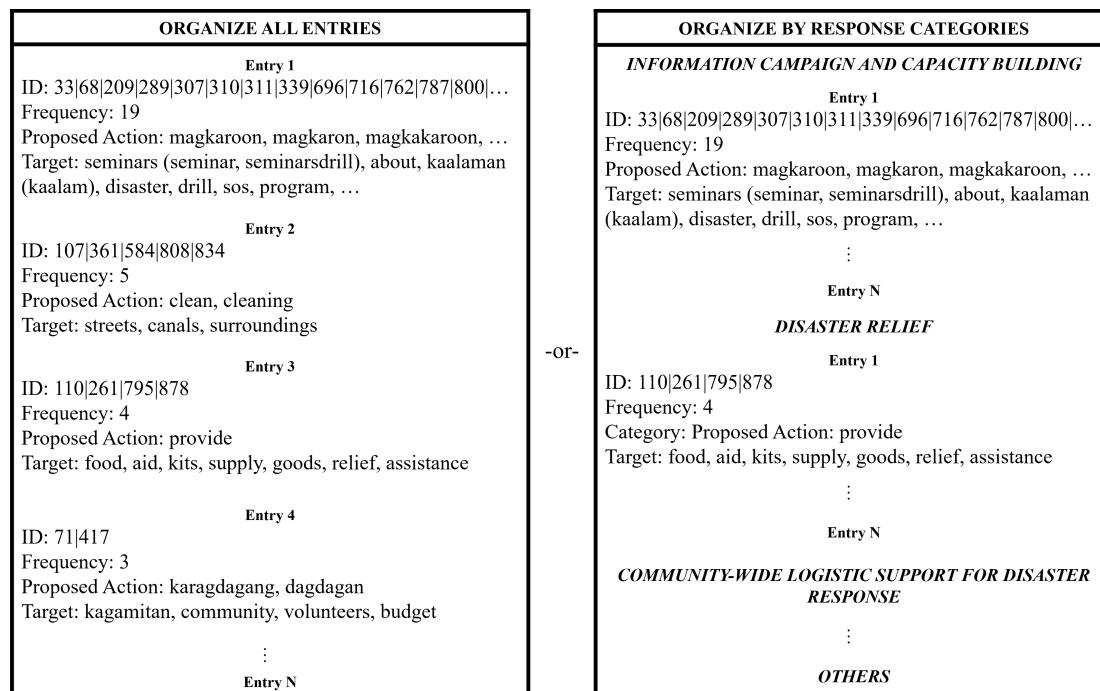


Figure 4.2: Information Organization Options

4.4.1 Information Clustering

Entries with similar contents are collated into a cluster. Those with exact word duplicates are filtered, retaining a single instance. Frequency counts are increased accordingly, and extra information are appended. In order to cluster entries properly, string (Sørensen-Dice Coefficient) and semantic (word embeddings) clustering were implemented, mainly to capture word similarities in the information.

For string similarity, an online available Python library called *Strsim*² was used. The collection contains string similarity measures, including Sørensen-Dice Coefficient. For word embeddings, pre-trained Tagalog Wikipedia models³ were used as resources and *Gensim* Library (Řehůřek & Sojka, 2010) for implementing Word2Vec (Mikolov, Yih, & Zweig, 2013) and FastText (Bojanowski et al., 2017).

The process of clustering starts by collecting all of the extracted insights from the structured information. The set of insights are then compared to one another to find words that are similar orthographically or semantically. Specifically, the process compares two types of string pairs. The first one compares the proposed actions or verbs (to join similar actions) and the second compares the targets or nouns (to remove duplicates within the clusters).

Applying a clustering technique on a pair of strings result into a computed number that represents their similarity. This similarity value is checked whether it is lower, higher, or equal to a similarity threshold. A value lower than or equal the threshold means the pair are not similar to each other, thus there would be no clustering involved. On the other hand, a value higher than the threshold means the pair are similar and should be joined into a single cluster, appending their information such as IDs (combined by vertical bar), frequencies, actions, and targets. In cases that a pair is exactly similar, only one instance of the word will remain, but their information is still appended. Note here that an insight can only belong to a cluster and if a cluster has been established, it could not be joined with others, but can only receive more members.

Provided with two types of string pair, their joining conventions are also a bit different from each other. Combining actions are done through commas (e.g., “magkaroon, pagkakaroon, nagkakaroon” ‘have’), while appended target words are through commas and similar target words through parentheses (e.g., ‘signal, supplies (food, equipment, boats), roadway, storage’). At the end of this process, a list of clusters with attached frequency counts can be used for ranking.

²Strsim codebase: <https://github.com/luozhouyang/python-string-similarity>.

³Pre-trained Tagalog Wikipedia Word Embeddings models: <https://github.com/Kyubyong/wordvectors>.

4.4.2 Information Ranking

The clustered information was ranked based on urgency. Since, information organization accommodates two options, ranking the information can use frequency counts in decreasing order, the given response categories as basis for prioritization, or a combination of both. Moreover, the prioritization ranking for categories was arranged beforehand based on its characteristic of being actionable by decision makers which are organizations or government units specific to handling national disasters.

Having said that, those that are not actionable by decision makers was the least priority such as entries under Preparedness for Emergency and Local Government Accountability which pertains to community effort and actions made by oneself, and miscellaneous categories such as those that mentions Filipino values and Others that involve topics outside the provided categories like corruption. Arranging the categories that prioritizes those that can be done before, after, and during disasters will make words tackled (or extracted) be in line with concrete actions to be considered or implemented by decision makers.

The currently applied priority ranking for the Malasakit Codebook 4.7 categories are as follows: Information Campaign and Capacity Building, Disaster Relief, Community-wide Logistic support for disaster response, Infrastructure Maintenance and Management, Early Warning System, Preparedness for emergency, Local government accountability, Filipino values, and Others. It has been stored in a list that can be modified for future changes.

Describing the process, after receiving the clustered information, if all entries were to be organized, the whole list of clusters will be sorted by frequency. Meaning, each of the entries are compared to one another where the highest counts are placed on top of the list. Otherwise, organizing entries by categories searches for and transfers the groups into a new cluster list based on the priority ranking. To elaborate, the group under Information Campaign and Capacity Building will be searched first in the cluster list and once found is transferred into a new list. The same process is done on the next group, Disaster Relief, then followed by the rest of the priority list.

Note that information organization's process on category ranking, clusters and ranks the information per category. So, once a category has been processed, it has already undergone frequency ranking, and once the whole list has been processed, it will then undergo prioritization raking or reordering of the category groups.

4.5 Report Generation

The ranked information was transformed into a report, generated in a list format. The report that is included in this tool was written in a Microsoft Word and Excel file. In this module, its main task is to add details and design the entries on those formats. Formatting the report considered alternatives such as displaying solutions per entry or per category, which can be toggled through the information organization module.

The report generation process follows a template-based approach, as to fill in predefined fields with the appropriate information. Since the goal here is to produce a report that fit the needs of decision makers and *Malasakit* team, the templates were designed to display information that will most likely be useful to them.

The Word document template was designed for reading and marking items. It is in a two-column format, ideally done to compress and reduce the page length of the contents. It consists three main parts: introduction, insights list, and *Malasakit* response list. The introduction is comprised of the report's title, timestamp, and brief introduction to the contents. The insights list is comprised of the ranked information, displaying the categories as heading (if organized by response categories), and lists each entry with the following fields: response ID/s, frequency count, proposed action, and target. The *Malasakit* response list is comprised of a table for reference, showing the response ID and text. Adding to the labels, for every page in this document show a footer with this research's title, proponent's name, and year.

The Excel file template was designed for analysis, disclosing more information related to the responses. It consists four main parts or sheets: response information, word set insights, cluster list, and ranked cluster list. In each row, response information is comprised of the two protected columns: raw response text and response category. It is then followed by language ('tl' or 'en'), normalized response text, part-of-speech (Stanford notation), response ID, and extracted insight phrases. The second sheet contains response IDs with extracted insights in word set format, where one row contains a single word set (proposed action/s followed by target/s). The next sheet contains the list of clustered information, with a cluster number heading and under it are its response IDs, frequency count, proposed action/s and target/s. The last sheet is comprised of the ranked version of the previous sheet.

4.6 Report Document

The final output for this software would be the generated report document which can be relayed to or used by decision makers in handling disasters. It presents the structured and organized key insights provided by the community in Filipino and English languages.

There are two reports that will be generated in each run of the tool. One is the official report intended for decision makers in a form of a Word document (.docx). Two is a report beneficial for users, researchers, and developers much like the Malasakit team, by overwriting and adding details to Malasakit's Excel file (.xlsx). Screenshots of the excel and word report are shown at Figure 4.4 and Figure 4.3, respectively, while more can be seen at Appendix C and Appendix D. Giving back to Malasakit, the contents of all reports, the API modules, its input and outputs, are freely accessible.

A	B	C	D	E	F
1	INFORMATION CAMPAIGN AND CAPACITY BUILDING				
2	Cluster 1				
3	13 46 208 236 243 307 311 598 603 19 maging, palaging, laging		kapitbahay, tao, disaster, mamamayan, awa		
4	Cluster 2				
5	26 208 209 339 507 576 649 762 766 18 dapat		barangay, seminar, oras, prepared, kalamide		
6	Cluster 3				
7	33 68 209 307 310 311 339 696 716 16 magkaroon, magkakaroon		seminars (seminar, seminarsdrill), about, ka		
8	Cluster 4				
9	80 121 132 149 156 169 215 222 234 15 be		advantage, possibilities, calamities, barangay		
10	Cluster 5				
11	130 215 245 249 259 266 340 390 79 12 have		assembly, disaster, drill, place, representativ		
12	Cluster 6				
13	207 238 269 289 332 343 344 382 39 12 conducting, magconduct, conduct, pagseminars (seminar), drills, community, assen				
14	Cluster 7				
15	101 104 169 287 333 343 347 396 58 10 preparing, prepare		typhoon, disaster, seminar, times		
16	Cluster 8				
17	169 222 303 331 343 358 614 627 65 10 help		prepare, families, disasters (disaster), outcor		
18	Cluster 9				
19	26 362 764 813 852 883 903 926 9 may		pangangailangan, seminar, meetings, kalami		
20	Cluster 10				
21	101 104 179 218 337 347 350 386 40 9 inform, informs, informing		consequence, people, neighbor, subordinate		
22	Cluster 11				
23	152 200 216 303 594 627 798 876 92 9 giving, living		disaster, drill, knowledge, seminars, leaflets,		
24	Cluster 12				
25	344 382 390 398 485 700 842 846 86 9 regarding		prevention, disaster (disasters), awareness, p		
26	Cluster 13				

Figure 4.3: Excel Report Screenshot

FILIPINO TEXT ANALYSIS TOOL REPORT

Mar-24-2020 20:24:54

The information below were extracted and organized automatically.

INFORMATION CAMPAIGN AND CAPACITY BUILDING

Entry 1

ID/s:

26|208|209|339|507|576|580|649|762|766|
811|813|836|852|854|885

Frequency: 20

Proposed action: dapat

Target: barangay, seminar, oras, prepared, kalamidad, sakuna, posters, paalala, drill, like, for, example, about, paraan, encourage, before, during, and, after, of, the, calamity, my, weekly, mg, roong

Entry 2

ID/s:

13|46|208|236|243|307|311|598|603|618|6
31|651|758|760|766|854|903|925

Frequency: 19

Proposed action: maging, palaging, laging

Target: kapitbahay, tao, disaster, mamamayan, aware, kabbarangay, kabaro, kalamidad, daring, beforeafter, and, during, pra, bagyo, trahedy, gawim, paalala

Entry 3

ID/s:

33|68|209|289|307|310|311|339|696|716|7
62|787|800|847|885|926|930

Frequency: 19

Proposed action: magkaroon, magkaron, magkakaroon, karoon

Target: seminars (seminar, seminarsdrill), about, kaalaman (kaalam), disaster, drill, sos, program, regarding, and, iinvite, emergency, kits, training, progma, pra, disciplina, mamamayan, weekly, meeting, organisasyon, orientation, barangay

Entry 4

ID/s:

80|121|132|149|156|169|215|222|234|340|
585|654|794|831|929

Frequency: 15

Proposed action: be

Target: advantage, possibilities, calamities, dont, times (time), drills, typhoon, disaster, preparedness (prepare), programs, officials, instructions, orientation, effects, duty

Entry 5

ID/s:

130|215|245|249|259|266|340|390|794|838

Frequency: 12

Proposed action: have

Target: assebly, disaster, drill, place, representative, check, emergency, needs, training (taraining), barangay, officials, seminar, meeting

Entry 6

ID/s:

169|222|303|331|343|358|385|614|627|654
|661

Frequency: 11

Proposed action: help

Target: prepare, families, disasters (disaster), outcomes, community, idea, seminars, people, signage, calamity

Entry 7

ID/s:

101|104|169|287|333|343|347|396|583|644

Frequency: 10

Proposed action: preparing, prepare

Target: typhoon, disaster, seminar, times

Entry 8

Extracting and Organizing Disaster-related Philippine Community Responses for Aiding Nationwide Risk Reduction Planning and Response (N. Nocon, 2020)

Figure 4.4: Word Report Screenshot

Chapter 5

Results and Discussion

This chapter contain results and analyses on the different experiments done. Experiments include those with regards to the utilized data or models, and software modules. Specifically, different clustering approaches, various word embeddings models, preprocessing tools, and lexicalization of entries was implemented. In addition, the software was tested on a different domain text. Evaluation was done through quantitative and qualitative analysis, for Information Extraction and organizational tasks, respectively.

5.1 Information Extraction

In evaluating the system, the Information Extraction's performance was measured through system-gold standard match counts (see Table 5.1) and standard, quantitative metrics such as Precision, Recall, Accuracy and F-Measure (see Table 5.2). Moreover, there were two tests, comparing insight phrases to know the performance in extracting insights and comparing word sets (see Table 5.4) to know the performance in extracting action/verbs and target/nouns. Sample format of the two are as follows:

- General format for Insight Phrases: [sentence number, insight 1, insight 2, ..., insight N]
 - [1, magkaisa dapat ang mga tao]
 - [4, paglilinis ng kanal, pagtatapon ng basura, mag ikot ikot ang mga tanod]

- General format for Word Sets: [sentence number, action/verb, target/nouns]
 - [1, magkaisa, tao]
 - [64, pagbibigay, pagkain, tubig]

5.1.1 Gold Standard

A gold standard was manually prepared by the proponent for this dataset and checked by a third person. It is mainly used as basis for comparison in evaluating the system's extractions (Information Extraction process). Under this are the ideal, expected, or correct insight extractions that corresponds to a Malasakit response.

There are two kinds of gold standard: original and its normalized version. Under each provides three parts: one, containing the Malasakit responses; two, with expected insight phrases to be extracted; and three, with expected word sets to be extracted. Given the sentence “paglilinis ng kanal wastong pagtatapon ng basura at kailangan mag ikot ikot ang mga tanod upang bantayan mga gamit ng tao”, an example for the gold standard phrases are the following: “paglilinis ng kanal”, “pagtatapon ng basura”, and “mag ikot ikot ang mga tanod”; and for gold standard word sets: “paglilinis, kanal”, “pagtatapon, basura”, and “mag ikot ikot, tanod”.

In 934 Malasakit responses, the gold standard has 934 rows and at most six columns (including response IDs) for insight phrases, whilst 1,238 rows or sets and at most nine columns (including response IDs) for word sets. Provided, there were responses with at most five insights, and responses with seven targets (nouns) in word sets.

5.1.2 Insight Phrases

Table 5.1: Insight Phrases System-Gold Standard Match Count Results

Insight Phrases %		
Measure	Original	Normalized
Complete Match	19.59	18.23
Over-extraction	8.35	8.99
Under-extraction	25.48	24.98
Overlapping-extraction	0.3	0.66
Complete Mismatch	46.27	47.13

Table 5.2: Insight Phrases Standard Metric Results

Insight Phrases % ¹				
Measure	Original _{ic}	Original _{wc}	Normalized _{ic}	Normalized _{wc}
True Positive	53.73	41.35	52.87	40.24
False Positive	29.69	17.65	29.39	17.46
False Negative	16.59	9.78	17.74	10.6
True Negative	0	31.22	0	31.70
Precision	64.41	70.09	64.27	69.74
Recall	76.41	80.87	74.87	79.15
Accuracy	53.73	72.57	52.87	71.94
F-Measure	69.90	75.09	69.17	74.14

Complete Matches and True Positives (TP) are statistical measures that counts the number of extractions that matches the gold standard (GS) or are considered as actual insights. In addition, there are partial matches that are also considered as insights which are the following: Over-extractions (OE), Under-extractions (UE), and Overlapping-extractions (OVE).

OE is a type of partial matches with more words in extractions than GS contents. Sample of OE are instances of GS entries without Auxiliary verbs at the start of the insight such as *dapat*, *kailangan*, and *be* (e.g., *dapat makiisa ang komunidad / makiisa ang komunidad*). Another, OE includes a few words after conjunctions ‘at’ or ‘and’ making it longer. Unlike in GS, some ideas were separated into two entries. An example is in GS, “linisin ang kanal” and “itapon ang basura” are separate entries, while “linisin ang kanal at itapon ang basura” is a single extraction entry. Adding to the causes of OE, forcing to search for a noun to end an insight added unnecessary words that could have stopped on an adjective or pronoun. Instances are, “be more aware and let the constituents / be more aware” and “help each other when the time / help each other”.

UE are partial matches with less words in extractions than GS contents. With the same cause in ending a noun for an insight, the effect is different. Instead of more words, necessary details are missing. The phrase “make the barangay gym” is an UE which should be “make the barangay gym ready for evacuation” to complete the insight. Despite this, there are some instances that UE can stand without the additional detail such as “ayusin ang drainage”, “do announcements”, and “conducting seminars”, which could have been “ayusin ang drainage ng barangay”, “do announcements via megaphones”, and “conducting seminars for disaster preparedness and occassional drills”.

¹For the metrics, values in parenthesis are based on word counts (wc), while those outside of it are based on insight counts (ic).

OVE are partial matches that are equal in length or wherein certain words overlaps with the GS. Sample for this is “kaylangan maging aware” and “maging aware sa balita”, where they overlap with “maging aware”. There are instances that OVE entries only differ in spelling, which by adding a normalizer failed to overwrite while in GS the annotator enforced the correct spelling, or the other way around (e.g., provide first aid *kita* / provide first aid *kits*).

Complete Mismatch are entries that does not match with either the system or GS. Under this measure belong the sum of False Positives (FP) which are extracted insights that are not actual insights or not in the GS, and False Negatives (FN) which are not extracted insights that are actual insights or in the GS.

In FP, majority of the instances counted were deemed to be either unusable or insufficient suggestions (as they were missing details) but were still extracted as it indicates an action/verb and a target/noun. Instances of FP are the following: “mentioned in the survey”, “magkaroon ang mga tao [of what?]”, and “allow the residents [to?]”.

In FN, results were similar to OE. As some instances with multiple ideas and conjunctions such as ‘at’ were joined by the system; unlike in GS, the entries were separated. In this system extraction, “magkaroon ng komunikasyon at ikutin ng mga council members”, only the first insight was counted as a match with the GS counterpart “magkaroon ng komunikasyon”, while the second could not be matched with the same extracted entry and is counted as FN. The restriction to match only once was placed so that the automated evaluation would not be prone to partial matches error.

In addition, there are a great number of annotated insights that started with a noun, adjective, or adverb. Examples include “paggamit ng public address system”, “announcement of a disaster”, “proper information dissemination”, and “regularly clean canals”. Moreover, there were also insights with typographical errors in the extraction’s end that did not match with normalized GS as annotations in it are correctly spelled (e.g., maayos na waste *dsiposal* [disposal]).

True Negatives (TN) are instances wherein the system did not extract as it is not an actual insight or not in the GS. For insight counts, TN is valued zero (0). It is due to the information available (in the insight/phrase level) focusing on those insights that were extracted and which should be extracted by the system, and not those that were not extracted. However, using word counts, the true negative can be computed by the total word count of the corpus. Specifically using the equation,

$$TrueNegative_{wc} = Total_{wc} - (TP_{wc} + FP_{wc} + FN_{wc}),$$

the words that were not extracted by the system and is also not an actual insight can be counted. Provided, statistical measures were computed through insight counts and word counts. For reference, a matrix has been provided in Table 5.3 as an alternative view for representing the TP, FP, FN, and TN statistical measures.

Table 5.3: Confusion Matrix of Information Extraction

	Description
True Positive (TP)	Extracted, Insight
False Positive (FP)	Extracted, Not Insight
False Negative (FN)	Not Extracted, Insight
True Negative (TN)	Not Extracted, Not Insight

Precision is the percentage of extractions that are insights and Recall is the percentage of insights that were extracted. Currently, Precision and Recall of the IE is 64.41 and 76.41, respectively. With word counts, values are higher with 70.09 and 80.87, respectively. To increase the values for these two, FP and FN are aimed to be low in value, while TP should be higher.

The 53.73 (or 72.57 word count) Accuracy is another value that represents the percentage of correct extraction, meaning more than half are considered to match with the GS. The 69.90 (or 75.09 word count) F-Measure score is the harmonic mean of Precision and Recall which can also be interpreted as the overall performance for extracting insight phrases.

5.1.3 Word Sets

Exact matches (EM) pertains to matches that has the same action/verb and target/noun with GS. Similar to CM and TP, it is ideal to have higher value for this measure. Partial matches (PM) label pertains to matches that has the same action and almost the same target noun with GS.

Instances of PM mostly differ in a few words, more or less than the actual GS annotations like [magkaroon, early, warning] / [magkaroon, early, warning, system] and [pagbibigay, assistance, goods] / [pagbibigay, humanitarian, assistance, goods]. There are some instances that could have been EM, only to differ in spelling like [improving, *imformation*, dissemenation] / [improving, information, dissemenation].

Table 5.4: Word Sets Match Count and Standard Metric Results

Word Sets %		
Measure	Original	Normalized
Exact Match	18.75	18.22
Partial Match	25.6	26.43
Action Match	8.22	9.32
Target Match	17.47	16.08
Crossover Match	4.02	2.48
No Match for Gold Standard	25.94	27.46
No Match for System	37.86	37.97
True Positive	51	50.2
False Positive	31.09	30.75
False Negative	17.88	19.02
True Negative	0	0
Precision	62.14	62.03
Recall	74.06	72.54
Accuracy	51.03	50.24
F-Measure	67.58	66.88

Under partial matches, action matches (AM) pertains to only the action field matches with GS. Another is target matches (TM) which pertains to only the target field matches with GS. AM and TM are comprised of entries with either insufficient or incorrect action/target, mismatches with GS' typographical corrections (e.g., *pgbaha* / *pagbaha* and *baranggay* / *barangay*), and GS' missed corrections (e.g., *paguuusap* / *pag uusap*, *nagpeperform* / *nagpe perform*, and *pagconduct* / *pag conduct*).

Providing an instance to insufficient or incorrect action/target, examples that exists in the data are like [conducting, lot] / [conducting, seminar, activities, disaster, awareness] and [think, meeting] / [organize, meeting, home, owners]. Similar to OE, TM also includes entries that recorded auxiliary verbs that differed with GS' verbs (e.g., [kailangan, ka-barangay] / [magtulungan, ka-barangay]).

Adding to the partial matches, there are crossover matches (COM) that is when action matches the target field of GS or vice-versa. Examples for this are the pairs: [mabilis, pagbibigay] / [pagbibigay, pagkain, tubig] and [do, train, deaf, emergency, responders] / [train, deaf, emergency, responders], where the action is designated on the first field while the rest indicates the targets/nouns.

No matches for System (NMS) is when an extraction does not match with any of the GS entries. No matches for GS (NMGS) is when a GS entry does not match with any extractions. NMS and NMGS' contents are generally the same with FP and FN. Examples for entries that were not considered as a suggestion by GS are the following: [maiwasan, pagbaha], [pagpapadala, kunting], and [putting, pockets]. For reasons that they contain verbs but were used as a justification to the real suggestion, failed to include a proper target, or unusable to act as a solution. On the other hand, NMGS samples mostly have actions that were tagged with a different Part-of-Speech, which definitely could not be extracted due to the verb to noun pattern rule.

Based on the results, less than 20% are EM, while PM has the highest correct instances with more than 25%. Represented with a combined value of 74.06%, extracting action and target fields were effective provided a straightforward Part-of-Speech pattern-based design. Perfectly extracting both fields, however, still needs work. Overall, word sets' Precision, Recall, Accuracy, and F-Measure metric values exceeded half, garnering 62.14, 74.06, 51.03, and 67.58, respectively.

5.1.4 Normalized Evaluation

A normalized version of the insight phrases and word sets were also evaluated. Generally, the original version gained better scores compared to it. The cause of this is mostly on spelling mismatches, specifically over-normalization (e.g., *kits* to *kita* and “ayusin ang mga” to “ayusin ang Metro Rail mga”), annotation normalizations (e.g., *kagamitin* / *kagamitan* and *pundo* / *pondo*), and unnormalized typos (e.g., *plansan*, *taraining*, and *alertoat*). Despite this, numerous words in the data were still standardized (e.g., *di* to *hindi*, *san* to *saan*, and *meron* to *mayroon*).

5.2 Information Organization

Quantitative and qualitative analysis were conducted on organized information to bring out the positive and negative characteristics, as well as, the coverage of clustering approaches and collated information. Each of the experiments was clustered through Dice Coefficient (Dice, 1945), Word2Vec (Mikolov, Chen, et al., 2013), and FastText (Bojanowski et al., 2017). Furthermore, Lexicalization was applied as sub-clusters by providing a word that can represent an entire word group. The word group is attached beside the representation enclosed with parentheses.

Experiments were generated, analyzing Malasakit Responses that were organized and ranked with all of the entries, and another grouped by response categories. Extending the test of all entries, a normalized version was also evaluated. To determine the impact of the 50% default clustering threshold, a value that indicates if words are considered similar or not, adjustments such as decreasing to 20% and increasing to 80% were made.

In this type of analysis, the following were aimed to be discovered: the set of words that are similar to each other, top community suggestions, response category with highest number of suggestions, effectivity of a normalizer, more appropriate clustering approach or threshold, and many more. Three reports were provided in this document, which are results from each clustering approaches that were organized by response categories. Generated reports can be found at Appendix D.

5.2.1 General Statistics

A statistical view has been provided that encompasses quantifiable values in the reports. The tables include the produced runtime and number of pages of the report, as well as, the number of entries and frequency (total responses = 1,392) a report has. In each table, experiments are abbreviated into the following:

- Organize All Entries (OAE), which is the base of the experiments,
- Normalized (N), the normalized version of the evaluation,
- Threshold (T), which indicates a threshold adjustment from 50 to an attached number beside the abbreviation, and
- Organize by Response Category (ORC), an alternative organizational format.

Under production of the report (shown at Table 5.5), runtime for the entirety of the program was included (from reading the file up to generating the report, together with the sum of the execution time of each modules/tasks). Majority of the experiments clustered in Dice's coefficient were deemed to be the fastest and lightest (least resources used) per experiments, with an average runtime of 10 minutes and four seconds. Oppositely, the slowest overall are experiments that made use of FastText clustering with an average runtime of 20 minutes and 17 seconds.

Table 5.5: Report Production of Experiments

Exp.	Approach	Runtime	HH:MM:SS	Total pp.	Insight pp.
OAE	Dice	604.42	0:10:04	60	27
	Word2Vec	667.37	0:11:07	60	27
	FastText	937.81	0:15:38	45	12
OAE-N	Dice	551.97	0:09:12	60	27
	Word2Vec	623.75	0:10:24	61	28
	FastText	926.11	0:15:26	45	12
OAE-T20	Dice	622.92	0:10:23	49	16
	Word2Vec	666.10	0:11:06	58	25
	FastText	677.15	0:11:17	38	5
OAE-T80	Dice	614.51	0:10:15	68	35
	Word2Vec	669.66	0:11:10	67	34
	FastText	1999.28	0:33:19	65	32
ORC	Dice	625.57	0:10:26	79	46
	Word2Vec	660.49	0:11:00	71	38
	FastText	1546.92	0:25:47	57	24

The quickest experiment to have been completed was OAE-N using Dice's coefficient. It was completed at nine minutes and 12 seconds and clustered insights worth 27 pages. Adding 33 pages for the Malasakit Response List as reference for the user, the total number of pages for this experiment is 60. On the other hand, the longest runtime was OAE-T80 using FastText word embeddings clustering. It took more than 30 minutes, in exactly 33 minutes and 19 seconds, and was able to cluster 32 pages worth of insights. Noticeably, OAE-N was faster than the base OAE. It is due to the fact that even though there are more components in OAE-N, post-normalization lessened the total words and may have helped in tagging; as comparing the runtimes, the two experiments extremely differed in tagging time.

Generally, the clustering approach with the least number of insight pages are FastText experiments, which indicates more words joined. As a matter of fact, the lowest number of pages was clustered by FastText with only five pages; given 1,392 total responses, most of them belonged into a cluster. Compared to FastText, Dice's coefficient and Word2Vec clustering have a large gap of insight pages in between experiments. Note that insight pages were provided but does not represent which among the clustering approaches produced the best set of insights.

Table 5.6: Report Composition of Experiments

Exp.	Approach	Entries	CEC %	CIC %	Highest Freq.
OAE	Dice	363	45.73	85.85	80
	Word2Vec	375	29.87	81.11	257
	FastText	114	57.02	96.48	490
OAE-N	Dice	365	45.48	85.70	82
	Word2Vec	380	28.16	80.39	255
	FastText	112	61.61	96.91	378
OAE-T20	Dice	179	63.13	95.26	127
	Word2Vec	339	25.96	81.97	657
	FastText	6	83.33	99.93	806
OAE-T80	Dice	506	33.00	75.65	79
	Word2Vec	489	30.88	75.72	104
	FastText	455	35.60	78.95	137
ORC	Dice	666	31.38	67.17	26
	Word2Vec	526	27.19	72.49	95
	FastText	295	47.12	88.79	159

Under Table 5.6, the composition of the report, that is the total number of entries, percentage of cluster entry counts (CEC) or number of entries that have actions merged (frequency greater than 1), percentage of cluster insight counts (CIC) or number of responses that have been clustered, and highest frequency count are presented. With regards to highest frequency of those organized per response category, the highest value found among all categories were placed – with an in-depth version that can be seen at Tables 5.7, 5.8, and 5.9.

Expanding on the definitions of CEC and CIC values, both basically represent the clustering capability in the experiment, but differ on perspective, where CEC signifies how diverse/dense the ideas of the clusters are and CIC signifies how much insights were covered. Note that low value for CEC means ideas given were more focused, while high gives off diverse ideas. On the other hand, for CIC, low or high value is in terms of how much insights were clustered. These two are computed using Equations 5.1 and 5.2:

$$CEC\% = 100 * \frac{\text{clustered entries}}{\text{total entries}} \quad (5.1)$$

$$CIC\% = 100 * \frac{\text{clustered entries}_{frequency}}{\text{total entries}_{frequency}} \quad (5.2)$$

For OAE experiments, the clustering approach with the lowest amount of clusters is Word2Vec, with 29.87% CEC and 81.11% CIC. Conversely, FastText produced better results, with 57.02% CEC and 96.48% CIC. Connecting the total entries with cluster percentage, it has been evident that as the total entries shrinks, the higher the cluster percentage values are.

The highest frequency value, however, does not relate into cluster percentage, but depends more on the type of clustering approach used. Aside from the highest amount of cluster percentage, FastText also produced the highest frequency with 490 responses under one cluster. The second, although produced the lowest cluster percentage among the three approaches, is Word2Vec with 257 responses. The last approach clustered at the most 80 responses in one cluster.

The normalization factor in OAE-N shifted a number of texts that should or should not have been merged. Comparing to the base OAE, OAE-N experiments that made use of Dice and Word2Vec increased the values of the total entries and decreased cluster percentages, while its effect on FastText decreased its total entries and increased cluster percentages. The normalizer's effect on an entry's highest frequency increased values for Dice, while decreased values for word embeddings approaches.

Comparing the base OAE and threshold adjustment experiments, the number of entries of OAE-T20 and OAE-T80 significantly decreased and increased the values, respectively. The effect is reversed for highest frequency where in OAE-T20 the values increased, with values more or less doubled, and in OAE-T20 values were decreased by more or less half of the original.

For CEC percentage of OAE-T20, there were less clusters made by Word2Vec, whereas the other two approaches gained more. For CIC percentage, all approaches achieved higher values. In fact, under this experiment, FastText achieved the overall highest cluster percentages with 83.33% CEC and clustering almost all of the responses with 99.93% CIC. Conversely, OAE-T80 produced more clusters using Word2Vec, and less on the other two approaches. Moreover, OAE-T80 has lesser CIC values than OAE.

From the given statistics, the gist of the adjustment is when the threshold is reduced, there would be more responses clustered, whereas increasing it would entail a more selective clustering process. Noticeably, there are changes in the quality of entries and clusters inside the report. More information about analysis of the insights produced were designated on another subsection.

With regards to ORC experiments, entries were spread out which resulted into total entries values that more or less doubled in count. The highest count is 666 for Dice, followed by 526 for Word2Vec and 295 for FastText. In terms of cluster percentages, the approach with the most amount is FastText, followed by Word2Vec and Dice. In terms of highest frequency, the results follow the order of OAE experiments, which evidently showed FastText as being able to produce the highest amount of frequencies compared to the other two approaches.

For ORC experiments, the statistics per category has been provided at Table 5.7 using Dice, Table 5.8 using Word2Vec, and Table 5.9 using FastText. Values provided include the total number of entries, cluster percentages, highest frequency in an entry, and total frequency count of the category.

Counting the total frequency of responses a category has, the top five are as follows: Information Campaign and Capacity Building (393), Early Warning System (259), Preparedness for Emergency (198), Infrastructure Maintenance and Management (169), and Filipino Values (114). The lowest among categories is Disaster Relief with only 34 responses included.

Table 5.7: Report Composition of ORC through Dice's Coefficient Clustering

Category	Total Entries	CEC %	CIC %	Highest Freq.	Total Freq.
Information Campaign and Capacity Building	152	36.84	75.57	20	393
Disaster Relief	24	16.67	41.18	5	34
Community-wide Logistic Support for Disaster Response	49	30.61	57.50	8	80
Infrastructure Maintenance and Management	84	36.90	68.64	25	169
Early Warning System	109	38.53	74.13	16	259
Preparedness for Emergency	84	28.57	69.70	26	198
Local Government Accountability	57	19.30	41.03	7	78
Filipino Values	54	27.78	65.79	12	114
Others	53	20.75	37.31	4	67

Applying Dice's Coefficient to ORC experiment, the top five categories with the highest number of total entries are ordered from highest to lowest and as follows: Information Campaign and Capacity Building (152), Early Warning System (109), Infrastructure Maintenance and Management / Preparedness for Emer-

gency (84), Local Government Accountability (57), and Filipino Values (54). The lowest count for the total entries is Disaster Relief with 24 entries. Counting the total entries signifies which categories have more variety of suggested actions attached to it.

In terms of highest frequency, it signifies which actions and categories have been highly suggested by the community. It can also be looked at which actions and disaster strategies are more important or urgent to the people. Numerically, the top five categories are as follows: Preparedness for Emergency (26), Infrastructure Maintenance and Management (25), Information Campaign and Capacity Building (20), Early Warning System (16), and Filipino Values (12). Again, the lowest count for this is Disaster Relief with 5 responses under one cluster.

In terms of cluster percentages, CEC signifies how dense or focused the suggestions of people are, particularly to the proposed actions needed in a community, while CIC signifies the amount of suggestions that has been clustered together. Statistics showed that less than half of the categories got CEC values greater than 30%, while more than half got CIC values greater than 65%. The category with highest CEC percentage is under Early Warning System with 38.53%, while the lowest is Disaster Relief with 16.67%. On the other hand, the category with highest CIC percentage is under Information Campaign and Capacity Building with 75.57% and lowest is Others with 37.31%.

Table 5.8: Report Composition of ORC through Word2Vec Clustering

Category	Total Entries	CEC %	CIC %	Highest Freq.	Total Freq.
Information Campaign and Capacity Building	130	22.31	74.30	95	393
Disaster Relief	16	31.25	67.65	9	34
Community-wide Logistic Support for Disaster Response	33	27.27	70.00	33	80
Infrastructure Maintenance and Management	70	27.14	69.82	33	169
Early Warning System	93	29.03	74.52	48	259
Preparedness for Emergency	63	33.33	78.79	43	198
Local Government Accountability	40	17.50	57.69	18	78
Filipino Values	48	35.42	72.81	20	114
Others	33	27.27	64.18	16	67

Applying Word2Vec, the top five categories with the highest number of total entries are as follows: Information Campaign and Capacity Building (130), Early Warning System (93), Infrastructure Maintenance and Management (70), Preparedness for Emergency (63), and Filipino Values (48). The lowest count for the total entries is Disaster Relief with 16 entries.

In terms of highest frequency, the top five categories are as follows: Information Campaign and Capacity Building (95), Early Warning System (48), Preparedness for Emergency (43), Infrastructure Maintenance and Management / Community-wide Logistic Support for Disaster Response (33), and Filipino Values (20). The lowest count for this is Disaster Relief with 9 responses under one cluster.

The highest in CEC with 35.42% is under Filipino Values. On the other hand, in CIC percentages, the highest is Preparedness for Emergency with 78.79%. The lowest for both CEC and CIC percentage was Local Government Accountability with 17.5% and 57.69%, respectively. Among all categories, the results show that a third got CEC values greater than 30% and five out of nine got CIC values greater than 70%.

Comparing Word2Vec statistics with Dice's coefficient, the total number of entries shifted but the categories with highest and lowest values remained the same. In highest frequency, categories were rearranged as values increased for all categories. Information Campaign and Capacity Building shifted up to first, having one cluster in the hundreds unlike Dice with only 20. Provided, entries were found by Word2Vec to be similar with others, thus were merged accordingly.

In spite the higher frequencies and supported by the values in CEC percentages, Dice was still able to have more clustered entries as compared to Word2Vec, where five out of nine categories have higher CEC percentage. Meanwhile, basing on CIC, Word2Vec was able to join more similar words with eight out of nine categories have higher CIC values than Dice. It is important to note, however, that results between cluster percentages does not entail which approach produced clusters that are better in quality (with regards to word similarity).

Applying FastText, the top five categories with the highest number of total entries are as follows: Information Campaign and Capacity Building (57), Early Warning System (44), Infrastructure Maintenance and Management (39) / Preparedness for Emergency (39), Filipino Values (30), and Others (26). The lowest count for the total entries is Disaster Relief with 14 entries.

In terms of highest frequency, the top five categories are as follows: Information Campaign and Capacity Building (159), Early Warning System (108), Preparedness for Emergency (40), Community-wide Logistics and Support for Disaster Response (40), and Infrastructure Maintenance and Management (36). The lowest count for this is Disaster Relief with 11 responses under one cluster.

Table 5.9: Report Composition of ORC through FastText Clustering

Category	Total Entries	CEC %	CIC %	Highest Freq.	Total Freq.
Information Campaign and Capacity Building	57	52.63	93.13	159	393
Disaster Relief	14	28.57	70.59	11	34
Community-wide Logistic Support for Disaster Response	21	42.86	85.00	40	80
Infrastructure Maintenance and Management	39	38.46	85.80	36	169
Early Warning System	44	52.27	91.89	108	259
Preparedness for Emergency	39	51.28	90.40	40	198
Local Government Accountability	25	44.00	82.05	27	78
Filipino Values	30	46.67	85.96	25	114
Others	26	50.00	80.60	17	67

In cluster percentages, less than half of the categories got CEC values greater than or equal to 50%; At the same time, CIC values of most categories are above 80%, for which a third are above 90%. The category with highest CEC percentage is under Information Campaign and Capacity Building with 52.63%, and Disaster Relief as lowest with 28.57%. Similar to CEC results, the highest CIC percentage is under Information Campaign and Capacity Building with 93.13% and lowest is Others with 80.6%.

Comparing FastText with Dice's and Word2Vec, total entries were reduced significantly which means more responses have been joined and majority of the frequencies were increased. It is evident since almost all of the cluster percentages and highest frequency values were in favor of FastText. The only instances that were not is CEC percentage of Word2Vec under Disaster Relief and highest frequency under Preparedness for Emergency.

5.2.2 Experiments Clustered by Dice's Coefficient

In the base OAE experiment, the most frequent suggestion clustered ideas that Barangays should be (dapat) active, aware, alert, prepared, and updated. It relates to ideally having communication (komunikasyon), programs (programa), seminars/drills, announcements (anunsiyo), evacuation, and food (pagkain).

Generally, highly used verbs pertains to *procurement* (e.g., magkaroon, pagbibigay, and provide), *dissemination* (e.g., pagiinform and ma-inform), *sanitation* (e.g., maglinis, linisin, and linisan), *preparedness* (e.g., ihanda, maiwasan, and mkaiwas), and *solidarity* (e.g., magtulong, makipagtulongan, and helping).

Procurement mentioned the following items in need for the Barangay: early warning system, medical kits, flashlight, garbage cans (basurahan), shelter, and grocery/supply. Topics for dissemination suggested information about disasters such as typhoons, floods, storms, and the consequences that comes with them. Sanitation points mostly towards cleaning the surroundings (kapaligiran), specifically sewers (kanal/imburnal) and areas near the households (harap/daan).

Mainly to avert or avoid ramifications such as floods (pagbaha), spread of tragedies (trahedyo), and getting trapped, community- and self-preparedness by alarms and designating places such as schools (paaralan) for evacuations was suggested. For solidarity, it has been recommended for families (magkakapamilya) and Barangay to continuously help (magtulongan) each other.

Aside from these suggestions, there are other low frequency ideas that can be deemed useful such as putting up leaflets or signages (karatula), utilizing media and officials, enforce laws, teach first aid, and more.

With regards to the approach, clustering words orthographically expectedly joined words not farther than prefixes and suffixes. Even without implementing a normalizer, since shortcuts or typographical errors are not far from the original's form/structure, comparing orthographically was still able to cluster some words together (e.g., mkaiwas, magtulong2, baranggay, and taraining).

As a matter of fact, implementing Lexicalization in all experiments, where one word is designated to represent a cluster (particularly in each of the Target/Noun field), orthographic clustering was successful in joining words related to each other. Instances are tenses and unstandardized variants like seminars [seminar, seminarsdrill], floods [flooding], and basura [basurahan]. In spite this, there are instances with close string distance but are not exactly related. Examples for this are evacuation-elevation and pagkain-pagkalap.

Normalized Evaluation

In OAE-N experiment, A Filipino Colloquialism Translator (Nocon et al., 2018) was implemented as a normalizer. Sample normalized words in the results are presented on Table 5.10. Under this list contains corrections on shortcut texts, typographical errors, and Filipino colloquialisms.

Table 5.10: Normalization Samples

Original – Normalized		
andyan – nandiyán	khit – kahít	palang – pa lang
anu – ano	kng – kung	pano/pnu – paano
anung – anong	kona – ko na	pg – pag
aq – ako	konting – kaunting	pls – please
atleast – at least	kpag – kapag	pmunta – pumunta
bgay – bagay	kya – kaya	pra – para
bgo – bago	lage/lgi – lagi	pwd – puwede
bhay – bahay	meron – mayroon	q – ko
bng – bang	mg – mag	qng – kong
cla – sila	mg – mga	s – sa
di – hindi	mging – maging	sakin – sa akin
dont – do not	mki – maki	san – saan
dpat – dapat	my – may	sma – sama
facebook – Facebook	n – na	tas – tapos
ganun/ganon – ganoon	naten – natin	tsaka – at saka
im – i am	nla – nila	tv – television
kana – ka na	nmin – namin	wag – huwag
kase – kasi	npo – na po	xa – siya
kelangan – kailangan	p – pa	yung – iyong

Even though the coverage of the normalizer was sufficient to correct the responses, not all were normalized correctly. Since the normalizer was built to be heavily dependent on a statistical model, some instances resulted undesirable insertions and replacements. An instance of insertion is evident on this phrase “as early as possible”, which included an extra ‘it’ in between ‘as’ and ‘possible’. Another is, “ayusin ang” which has “metro rail” succeeding it.

Instances for replacement were evident on words that are considered as colloquialisms or interlingual homographs – words that exists on both English and Filipino language. In the English phrase “seminar for pre or post...”, the prefix *pre-* was found to be a shortcut for the colloquialism *pare* (buddy). In this

case, *pre* is also an interlingual homograph. Other mistakes are ‘kits’ in “medical kits” were considered as a typographical error for *kita* (you or salary), to into *ito* (this), non into *noon* (previously), and my into *may* (there, have, or with). Unfortunately, some repercussions of these mistakes removed words from clusters.

With incorrect normalizations, there are also those that were not normalized. There were merged words that was not covered by the normalizer such as ‘atpagbigay’, ‘sumunodkapag’, ‘sakunamaaari’, and ‘dahilbansa’. Another is a typographical error variant that made use of letter ‘q’ as a ‘g’ such as ‘paqdating’ and ‘paqaabiso’. Moreover, there are also shortcut variants, specifically the omission of vowels, that were not present in the statistical model such as ‘ngbbgay’, ‘dhlans’, ‘magki2ta’, ‘mlman’, ‘gwen’, and ‘magsgwa’.

Observing the effect of the normalizer as compared to the base OAE, frequency counts fluctuated into increasing and decreasing values, shifting the order of clusters, but did not affect the order of the top five. What has been evident from the normalized version is there are words that were previously included in the clusters that are shortcuts which was then removed, and new or corrected words appeared.

A few of these which were removed are ‘pra’, ‘cla’, ‘xa’, and ‘lage’, most of which are included in Table 5.10. Given this, there have been a number of instances that shortcuts as such were tagged as nouns and post-normalization were tagged otherwise.

As part of the normalization task, joining Filipino/Tagalog prefixes with their separated root words caused a set of words to appear as insights, and in action and target fields. In this instance, the prefix *mag-* was joined with *karoon* which resulted into *magkaroon*, a valid member of a proposed action cluster. Provided with more examples, joined prefixes found on the experiment are the following: *magplano*, *nagpeperform*, *pagpapadala*, *pagpapaalala*, *paguusap*, *pagbigay*, *paganunsyo*, and many more.

Provided with the changes, highly used verbs remained the same, which suggested ideas regarding procurement (e.g., *magkaroon*, *pagbibigay*, and providing), dissemination (e.g., *pagiinform*, *ma-inform*, and *inform*), sanitation (e.g., *maglinis*, *linisin*, and *linisan*), preparedness (e.g., *prepare*, *preparing*, *ihanda*, *maghanda*, *maiwasan*, and *mkaiwas*), and solidarity (e.g., *magtulong*, *makipagtulongan*, *help*, and *helping*).

Threshold Adjustments

Decreasing and increasing the similarity threshold from 50% to 20% and 80% changed the members of clusters. Reducing the threshold signifies a loose acceptance in determining if distance between two words are similar. The assumption is since the condition is looser, there may be more related (variants) words captured in the clusters. Increasing the threshold on the other hand, tightens it, thus producing harder but more closely similar clusters. The assumption is since the conditions is stricter, unrelated words will be filtered, meaning producing more accurate or closely related words in the clusters.

In OAE-T20, the most frequent suggestion is a cluster that consists of inflected verbs with the prefix *mag-* or suffix *-ing*. From the word *maging* (be), members of the clusters include the following: *magturo* (teach), *magsgwa* (magsagawa / perform), *magtayo* (build), fixing, giving, telling, messaging, calling, helping, knowing, and more.

Pointing out the highly used verbs, the set for OAE-T20 are not the same with OAE that provided per cluster a theme such as procurement, dissemination, sanitation, and more. The verbs were mostly grouped through their similarities with either the affixes or root words. Aside from the example with the most frequent suggestion, others include groups such as *papadala* (deliver) with *papaalam* (inform), and *naipapataasan* (raise), and *magtulungan* (help) with *pagtulungan* and *tutulongan*.

With this information, lexicalization also included target words to have more clusters or members. The members were joined not through the meaning of a word but in orthographic means. Instances grouped house with household, *kita* with *balita*, *kalamidad* (calamity) with *kalamid* and *kaalaman* (knowledge), seminar with *seminar* and seminars, and many more.

Despite this, there still were instances that the groupings coincidentally made sense semantically. Examples for these are *komunikasyon* (communication) with *organisasyon* (organization) and *impormasyon* (information), *ka-barangay* with *kabarangay* and *barangay*, and information with dissemination. In addition, the low threshold value enabled the clustering process to capture variants with far distances that were not clustered in the base OAE (e.g., *pagkakaiisa* with *kaisa*, and *kanal* with *imburnal*).

In comparing OAE-T80 with the base OAE, instead of words being clustered together, there are instances that initially clustered are separated, even if the difference is just a letter. Since Dice's coefficient runs through the strings with a set of three letters or trigram, values for the comparisons differ based on letter

order or positioning, and string length. Backed by the formula, Dice's coefficient computes the similarities by the common trigrams as numerator, and trigram counts of the two strings as denominator.

Visualizing this, the word *perform* and *performs* resulted into a similarity score above the 80% threshold, as the ‘s’ character was present at the end of the word (works also if at the start) and comparing both through trigram differs at only the last trigram ‘rms’. For the pair *barangay* and *baragay*, however, their similarity score is below the threshold as the ‘n’ letter difference is position in the middle, causing it to have more trigram differences as compared to the previous example.

Elaborating on the string length, the lesser the size, the least likely it would be to have a higher score given that a pair only differs one letter. Provided with a progression of the word *magikot* (roam) from its prefix *mag-*, which has a length from three to six, and comparing it with the last character replaced with ‘x’ (i.e., *mag-max*, *magi-magx*, ..., *magikot-magikox*); the resulting values for string similarity increases, starting with 0 and following with 0.5, 0.67, 0.75, and so on, so forth.

Negative samples for this restriction was unable to cluster the following pairs: kit-kits, training-taraining, and basura-basurahan. Positive samples, on the other hand, separated the following pairs: evacuation-elevation, drill-drills, and seminar-seminars. Moreover, there were others that the separation produced better clusters; before under one cluster are *pagbaha* (flood), *pagbabaha*, and *pagbara* (clog), then after threshold increase created two clusters with pairs pagbaha-pagbabaha and pagbara-pagbaba.

Generally, for this experiment, it proves that the assumption of increasing or decreasing the threshold could filter out, include, or basically place words in more appropriate clusters. Most importantly, this threshold controls how general and specific the contents of the clusters could be. Although, it has to be considered that excessive amounts could decline the quality of clusters. Hence, the value has been returned to a balanced threshold value of 50%.

Clustered per Response Categories

The ORC experiments divided and spread out all the entries for which categories the responses are under. There are nine response categories in total: Information Campaign and Capacity Building, Disaster Relief, Community-wide Logistic Support for Disaster Response, Infrastructure Maintenance and Management, Early Warning System, Preparedness for Emergency, Local Government Accountability, Filipino Values, and Others, where each have their own highly suggested ideas.

Analyzing the clusters in this experiment, target/noun clusters were considered similar with OAE. So, the quality of the clusters was dependent on grouping actions/verbs. Since, Dice's coefficient groups together words orthographically, variants properly belonged with each other and pertain to a single action; with the assertion of response categories, entries can be related to them, as to the needs or actions to be done under that category.

ORC using Dice's coefficient and under Information Campaign and Capacity Building, focuses on actions that the community must have (dapat), which includes the necessary items or programs, preparations, support, and logistics involved in information dissemination. Key insights highlighted conducting programs such as seminars, drills, and assemblies. As well as the use of infographic materials such as signages, posters, and leaflets. Suggested content consists of reminders or tips on what to do before, during, and after the calamity, while the target for these are family members.

Disaster Relief contain clusters that mentioned receiving (pagbibigay) or providing assistance, goods, food or grocery (pagkain), medicine (medisina), and evacuation option. As a matter of fact, the highly clustered verbs pertain to the same idea, but orthographic clustering was not able to merge them all into one.

Similarly, Community-wide Logistic Support for Disaster Response contain clusters that suggests having safety gears, sirens, and shelter for the operations. It has also been suggested to add more budget, equipment (kagamitan) and volunteers. Expounding on the equipment, several responses grouped together medical kits and flashlight. Moreover, there was another cluster that pointed out boats and storage facility are a must have.

For Infrastructure Maintenance and Management, most of the suggestions were about cleaning up (maglinis) the surroundings (kapaligiran), specifically the streets (daanan), sewers (kanal), rivers (ilog), and garbage waste (basura). Cleaning these areas in the minds of the community would ensure prevention (maiwasan) of floods (pambahay) and clogs (pagbara). Furthermore, in avoidance to throwing trash everywhere, one of the top suggestions mentioned proper place (lugar) or containers (tapunan) for garbage.

As stated for Early Warning System, clusters contain responses with insights that involves information dissemination. There should be (dapat) proper communication (komunikasyon), alert, news (balita) and updates, and radio. Grouping the information (pagiinform), it should be about the disaster, specifically the typhoon, and announced to the public, people (tao), citizens, and residents. Another top cluster had the same idea, that is to alert if there are (may) calamities (kalamidad), catastrophe (sakuna), disaster, typhoon, communication, support,

and assembly. However, given the same reason, the two clusters have different verbs nor even a variant of each other for them to be merged.

Under Preparedness for Emergency, clusters were in a form of a reminder, with verbs grouped such as maging (become), pagiging, magiging, and palaging (always). Target clusters for this set is to alert and attentive (atentibo) with nature (kalikasan), news (balita), and officials (opisyal). Another cluster suggests to be reminded of being aware and updated with the same target or nouns specified previously. Ideas as to where people could be reminded is through watching the weather forecast or news in television. In addition to this, other things to prepare are the following: news, evacuation plan, officials, management, unit, and the community.

For Local Government Accountability, these are the target ideas that the government should be accountable for: disasters, evacuation, posters, case, and society. It is also expected for the government to be ready (kahandaan), cooperative (mgsma/magsama and mgusap/magusap), and able to show (hrpin/harapin) and send (maiparting/maiparating) help (tulong). In light with this, there was a suggestion that the government should be (dapat) active.

Regarding Filipino Values, clusters pointed out two main ideas and these are to help and cooperate with each other. In terms of the first idea, help should be observable in preparing (paghahanda), families (magkakapamilya), and community (komunidad); while the other pointed out being cooperative with duties (tungkulin), plans (plano), and preparation. Other than the two, separate entries have verbs that mentioned tidiness (malinis), readiness (handa), equality (pantay-pantay), and kindheartedness (magmalasakit).

Under Others category, there were mixed ideas to improve disaster prevention and mitigation. There are a few clusters that called out corruption, which mentioned the act of putting valuables in pockets. Some endorsed their satisfaction with the decision makers, encouraging to continue (ipagpatuloy) their work or activities (gawain). There are also suggestions about minor and major entities in disasters, that there has to be communication between deaf, citizen, and responders.

Overall, using Dice's coefficient grouped together variants, but there are others more that could have been included in the clusters. Causes could be far string similarity distances or connection between words are through their meanings. A few of these are magkaisa-makiisa, handa-mapaghandaan, maglinis-clean, fixing-fix, nagkakabit-naglalagay, and create-make.

5.2.3 Experiments Clustered by Word2Vec

In this experiment, responses were clustered semantically through Word2Vec, which makes use of vector distances to determine if words are related to each other. In OAE, the most frequent suggestion clustered verbs that seem to be the essentials in disaster response. Sample verbs are be, put, have, do, provide, make, create, know, help, cover, check, give, avoid, visit, support, contain, update, and more. From this verb collection, target nouns were training, news, garbage, evacuation, drills, flood, devices, supplies, and more. It is apparent that there are a lot of suggestions in this cluster pertains to objects that are a “must have” in the barangay. A few of these are boats, equipment, facilities, electricity, kits, gears, megaphones, goods, signage, storage, and dumpster.

One notable difference between Dice’s coefficient and Word2Vec is vectors are positioned based on usage, so words that operate the same way are close together. Having said, in clusters, there were synonyms present such as make-create, help-support, have-contain, and provide-give. Moreover, there were also tense variants such as make-making, be-been, and need-needed. Lexicalizing the given targets, there was a large cluster under the topic of ‘training’ which consists of the following words: community, disaster, times, typhoon, plan, things, programs, government, technology, effects, management, events, case, info, society, emergency, food, help, assembly, warning, sign, advance, areas, week, gym, days, question, people, map, part, place, center, check, needs, information, aid, class, house, project, drive, household, proper, object, council, supply, heart, relief, and systems. On it are different topics and participants that can be the focus of trainings.

Listing other major clusters advises actions regarding *logistics* (e.g., magkaroon, magamit, maglagay, magbigay, ayusin, panatilihin, tanggalin, ilabas, malaman, gawin, magbibigay, mabigyan, handa, gagawin, matulungan, ibigay, malaman, bigyan, iwasan, ipagpatuloy, ilagay, hikayatin, makakatulong, and magkakaroon), *dissemination* (e.g., inform), *sanitation* (e.g., linisin), *preparation* (e.g., prepare), and *solidarity* (e.g., tumulong, gumawa, magsagawa). Under logistics, aside from generally covering it, there are other top clusters that indicated movement (e.g., dumating, pumunta, and lumikas), procurement (e.g., giving), and expectations (e.g., maayos, dapat, mabawasan, epektibong, and inaasahan).

Under these topics, noun for logistics indicated sub-clusters about supplies (e.g., signal, gamut, komunikasyon, basura, and pagkain), areas (e.g., tahanan, lugar, bahay, and paaralan), information (e.g., detalye, plano, problema, kaalam, ideya), and situation (e.g., sitwasyon, kalinisan, kalagayan). In spite of these related sub-clusters, there were also lexicalized instances that does not make sense such as joining food, warning, peace, report, notice, and government.

Dissemination topics recommended to point out the consequences of disasters to people, citizens, residents, neighbor, subordinate, and barangay. Sanitation then mentioned cleaning up garbage from sewage, river, and streets. Next, in terms of preparation, a place for seminars and alarms was let out. Last is for solidarity that pointed out helping the organization or community in activities or programs such as seminars, drills, orientations, and the likes.

Encapsulating the Word2Vec approach, there were a number of samples that proved to be sensible, with cluster members having clear relationship with each other. Providing further examples, notable instances include bagyo-baha, month-week, balita-ulat, sakit-kapansanan, organisasyon-pagpupulong, typhoon-storm, bagyo-lindol, people-public-community, harap-paligid, napinsala-tinamaan, and mataas-malaki. Extending this, it includes orthographic variants as instances like barangay-baranggay, kapaligiran-paligid, kabataang-bata, nakarating-pagdating, and pagtulong-tulong were grouped together.

However, relatedness does not always mean it produces the ideal groupings. There were clusters that produced antonyms like tao-bagay and sakuna-sanhi. Furthermore, Word2Vec missed words that should have been clustered together like prepare-preparing, giving-pagbibigay, maglinis-clean, announce-inform, and linisin-maglinis; and although under one cluster already because of merging their verbs, there are instances that should be lexicalized like kanal-imburnal, barangay-brgy, typhoon-flood, and gamit-bagay.

Normalized Evaluation

The underlying successes and issues of normalization discussed on Dice's coefficient is present in OAE-N experiment for Word2Vec. Some clusters had increased values, and some decreased because of it. As a result, there were shifts in the insights as to the verbs and nouns extracted, affecting the generated clusters.

Even with normalization, the shift did not prevent Word2Vec in overlooking words as cluster members. The assumption was, with normalization, more verbs will rise and be placed more appropriately on the vector space, while nouns will be correctly pointed out by the part-of-speech tagger and information extraction's pattern matching. However, the fact that most normalizations affected verbs inflected with prefixes; at the most, it can only add a few responses with the similar prefix or root word into some clusters and unfortunately was not able to significantly manipulate the vector's usages or positions in the space.

Concerning cluster composition, the changes with some of the groups from the most frequent clusters alone are provided. From ‘training’ sub-cluster, the keywords ‘events’, ‘info’, ‘week’, ‘world’, ‘aid’, ‘household’, and ‘relief’ moved into the ‘news’ sub-cluster. Another, from *sitwasyon* sub-cluster containing *kalinisan* and *kalagayan*, the members *kaligtasan*, *epektos*, *kalikasan*, and *ideya* was added. Then, the *samahan* sub-cluster was disbanded and removed. From a verb cluster regarding logistics, *magkaroon* was clustered with *pakakaroon* and *nagkakaroon*, *gumawa* with *magsagawa*, *magdagdag* with *dagdagan*, and *iwasan* with *maiwasan*. Observing the rest, here are notable examples that new clusters were made: *kaalaman* with *problema*, *komunikasyon*, *kalinisan*, *gawain*, *paraan*, *kaligtasan*, *karanasan* with *epektos*, and *gamit* with *kagamitang*.

Threshold Adjustments

In OAE-T20, the most frequent suggestion hoarded and absorbed verbs from other clusters. Under this cluster, ideas relating to logistics were grouped together. Initially with the theme of essentials in disaster response, most of its contents were transferred to come up with a collective idea towards logistics such as concerns with its movement and expectations. In this new cluster, suggested actions such as *magkaroon*, *tumulong*, *maayos*, *dapat*, *maiwasan*, *maglagay*, *magbigay*, *panatilihin*, *mabawasan*, *mabilis*, *epektibong*, *paparating*, *ilabas*, *malaman*, *magpaddala*, *magturo*, *malinis*, *dagdagan*, and more were grouped.

Loosening the threshold clustered more related words that were not joined before. Instances include verb pairs such as *magkaroon-nagkakaroon*, *nakakatulong-makakatulong*, *maayos-ayusin*, *nalaman-malaman*, *paligid-daan*, *dagdagan-add*, *mabigyan-provide*, and more. However, the contents in one cluster are general or vague, mixing up the ideas under one large topic. Visible with sub-clusters in nouns, there were lexicalized groups that mixed medicine (*gamot*) with cleanliness (*kalinisan*), barangay and typhoon (*bagyo*), and cause (*sanhi*) and food (*pagkain*).

Unaffected by the absorption, there were highly suggested clusters (e.g., *linisin*, *prepare*, and *giving*) that at the most changed only on the noun sub-clusters. Sample of this change added sewage (*kanal*), garbage (*basura*), and pathway (*daan*) with the original sub-cluster of word variants of surroundings (*kapaligiran* and *paligid*).

Contrary to OAE-T20, OAE-T80 dispersed cluster members in the base OAE. From the most frequent suggestion with actions that indicates essentials to possess, it has been divided into two entries. Although both have the same idea, distances in the vector space did not cut out for the threshold. This effect re-

occurred throughout the whole output, verb and noun clusters alike. A number of clusters was left with only one verb (e.g., magkaroon, pagbibigay, dagdagan, handa, makakatulong, and malinis), which means verbs must be exactly the same to be clustered together.

Due to dispersed verb groups, target fields have lesser appended nouns. As a result, it also lessened the sub-clusters or lexicalized ones. The 80% similarity threshold restricted into joining its members, that is why pairs like harap (front) and paligid (surroundings) were not lexicalized anymore but remained under the same cluster.

OAE-T80's effect however is not all negative. Since the restriction ensures that words are 80% and above close in the vector space, clusters clearly manifested the relationship between them. A good example for this is 'government' being clustered to only 'assembly' and 'council', which depicts a group of people. Another joined 'training' with 'community' and about 'disaster', which are the elements for disaster training.

Clustered per Response Categories

Similar to OAE-T80, ORC experiments divided responses by their designated categories. In this case, a balanced threshold was used for grouping similar ideas together. Analyzing the grouping of actions/verbs and target/nouns, there are certain representative topics that can be taken out of the clusters. Under Information Campaign and Capacity Building, the most frequent ideas consist of verbs that brings out the necessary actions involved (based on have, tell, create, avoid, and give), intended content and delivery medium (based on nais, dapat, malaman, kailangan, and sana), receiving end (based on mabigyan, maging, magkaroon), and decision makers (based on magbigay, magturo, magsagawa, magtayo, and talakayin) of the information.

Nouns involved for the necessary actions are the following: drills, programs, officials, meeting, siren, duty, and tips. In the intended contents and medium for the information, nouns included calamity, method (paraan), seminar, schools (paaralan), and Disaster Risk Reduction Management. For the receiving end of the information, neighbors (kapitbahay) and people (tao) in general must be aware, invited, disciplined (disiplina), and knowledgeable (kaalam). Then for decision makers, they are recommended to conduct programs such as seminars, drills, assemblies (pagpupulong), and orientation.

Under Disaster Relief, the top clusters mostly involve which items the community needs. One sub-cluster combined nouns containing food, aid, supply, relief, and case, then another with food (pagkain) and medicine (gamot). Unfortunately, even with almost the same set of nouns, there were verbs such as giving, provide, magbibigay and pagbibigay (give) that have not been merged together by Word2Vec.

In the same way, Community-wide Logistic Support for Disaster Response provided top suggestions with the same topic. It pointed out clusters specific to equipment and vehicles that can be used to support disaster response. Sub-clusters present groups about emergency, where the following words are included: place, aid, disaster, areas, food, center, people, and technology.

Under Infrastructure Maintenance and Management, most of the highly suggested ideas involve sanitation (cleaning and supplying). A sub-cluster with clear relationship between members are about flooding (pagbaha), that includes catastrophe (sakuna), pathway (daanan), road (kalsada), and flood (baha). In addition, there is a sub-cluster that mentioned the kapaligiran (surroundings) with lalagyan (place), daanan (pathway), and pagtaas (elevate). Moreover, in this category there were instances that lexicalization included orthographic similarities like clustering kapaligiran and paligid together.

In Early Warning System, highly suggested verb clusters indicated ideas that in this category must have. The cluster with highest frequency put up a bunch of actions words where their relationship are not easily distinguishable such as warning, do, make, be, have, coming, drive, is, are, has, eat, visit, update, give, avoid, said, watch, provide. The fact is there are keywords that relates to actions to be implemented for this category.

Focusing on target clusters, one sub-cluster collected ideas that seems to be describing a workshop; members of this include training, technology, community, disaster, warning, sign, typhoon, days, people, house, food, events, and emergency. More notable examples of sub-cluster paired up warning (babala) with typhoon (bagyo), situation (sitwasyon) with detail (detalye), typhoon with flood (baha), and radio with update. Provided with examples, there are cases wherein the orthographic similarities were evident on the vector space such as *maagang* (early) and *maaga*. Within this vector space, there are also cases that opposites were merged such as *mababang* (low) and *mataas* (high).

Under Preparedness for Emergency, a set of ideas on how to act in situations were presented. Word2Vec was able to combine essential actions related to preparedness which are *maayos* (fix), *malaman* (know), *magpadala* (send), *matugunan* (address), and *maiwasan* (avoid). Distinguishable topics in lexical-

ized sub-clusters are about readiness (paghahanda, pangangailangan, kaligtasan, pagkain), news (balita, ulat, programa / disaster, typhoon, week, times), time (oras, araw), and planning (plan, things, management, unit, community, time, proper, case, council, emergency).

Similarly with Early Warning System, in Local Government Accountability category, the cluster with highest frequency indistinguishably combined words such as help, be, are, want, working, make, have, know, having, and feed. Analyzing its target words, a sub-cluster lexicalized with help (people, map, part, government, case) was provided. In line with it are the nouns disasters, barangay, officials, personnel, evacuation, society. Basing on these, the idea for the top cluster signifies recommendation to either help the decision makers or make decision makers focus on helping the people. Moreover, there was a sub-cluster that grouped together problem (problema) with policy (patakaran), method (paraan), and attention (pansin).

In Filipino Values, it would be beneficial to encourage each other to help. With regards to clustering, recognizable positive samples are the following: samahan-tulong, tahanan-bahay, bagyo-baha, and baranggay-barangay. In another view, there are entries that could have been clusters with other entries. Example for these are *magkaisa* and *makiisa*, *tutulong* and *magtutulungan*, *matutong* and *matutunang*, and etc.

Last category is Others. Under this has not been much on both insight and clusters quality. For verbs, examples with clear relationships are humingi-magbigay, dapat-nararapat, and wala-walang; while for sub-clusters are sign-object, emergency-people, typhoon-disaster, warning-device.

Synthesizing ORC experiment using Word2Vec, examples showed its ability to use the vector space into providing semantic relationships between words. It was able to cluster orthographic variants, thematic similarities, usages, and even antonyms. However, given this algorithmic ability, there were high frequency clusters that has members with relationship indistinguishable from each other. In addition, there were some obvious instances that could have been included in either verb or noun clusters.

5.2.4 Experiments Clustered by FastText

Grouping responses using FastText attempted to produce clusters that are semantically related. In OAE experiment, the highly suggested entry has verbs that are about actions to be done in disaster response. In terms of quantity, majority of the verbs are in English and as a whole, this cluster is larger than Word2Vec. Sample English suggested actions include maximize, alert, improve, prepare, enforce, inform, clean, provide, minimize, remind, implement, install, ensure, gather, educate, join, create, communicate, fix, teach, and more. A few Filipino actions include *maiaannounce*, *ma-inform*, *linawin*, and *ialarm*.

Among these verbs, there are instances that are proven to be synonymous with each other. Pairs include enforce-implement, establish-create, educate-teach, contain-hold, support-help, and tell-announce. At the same time, there are also antonyms such as minimize-maximize, let-prevent, spread-gather, add-reduce, and give-take. Other semantically related members describe communication through messaging, texting, and calling.

One special characteristic in FastText that has made its way to be evident in providing orthographically related words is the use of subword information (character n-grams) as representation. Cluster members included verb tenses such as improve, improved, and improving. Furthermore, there were intra-word code-switching connected to each other. Example pairs for this are inform-mainform, provide-magprovide, participate-magparticipate, and alert-ialarm.

Under the same cluster, two of the largest sub-clusters relates to disaster (with awareness, dissemination, consequence, evacuation, etc.) and sirens (with laws, theft, alert, gear, etc.). Smaller sub-clusters still provided clear relationships with each other, with pairs such as responders-response, assembly-house, news-radio, trash-risk, and people-aid.

Analyzing the rest of the clusters, they showed the same kind of relationships in groupings. In fact, on the second largest cluster, most of its verbs were closely related based on tenses. The set includes *magkaroon*, *maging*, *pagkakaroon*, *magtulong*, *maipaabot*, *maiwasan*, *maglagay*, *nagkakaroon*, *magbigay*, *mabawasan*, *maglinis*, *maghanda*, *magbibigay*, *mabigyan*, *magtulong2*, *magpasagawa*, *mabigyang*, *nagbbigay*, *makatulong*, *bigyan*, *bigyang*, *magtanim*, *matugunan*, *magsagawa*, *mabigay*, *mapagbigay-alam*, *magtulongan*, *maibigay*, *ipagbigay*, and more.

Provided with this example, the shortcut words *nagbbigay* and *magtulong2* were able to join the cluster with *magbigay* and *magtulongan* – where Dice's coefficient was also able to but not Word2Vec. On another view, instances that Dice have not covered orthographically but was clustered by FastText are the verb

clusters of *magturo* (teach), fixing, *paalalahanan* (remind), *mabawasan* (reduce), *pagtibayin* (fortify), and more. In the sense, FastText effectively combined the two approaches as to covering orthographically similar words and using the vector space as basis for relationships.

Gathering ideas from other clusters, those with high frequency collected ideas regarding dissemination (e.g., *pagbibigay*, *pagiinform*, *magtawag*, and *pagpapalala*), solidarity (e.g., *magkaisa*, *tumulong*, and *magsama*), preparedness (e.g., lessen, *maagap*, and *maagang*), and maintenance (e.g., *tanggalin*, *palitan*, and *ilagay*). Running through these ideas, one limit with FastText is there were still fragments of word variants scattered across different clusters. Case in point, from the base word *tulong* (help), there were entries that on another cluster contains *tumulong*, and on another is *makakatulong*, which should have been on a single cluster. Reflecting on this, the primary cause is the summed vectors' positioning in the space, which is determined by their usage.

Normalized Evaluation

In OAE-N experiment for FastText, some entries were affected by the shift of words created post-normalization. Shifts in the sense that shortcuts, typographical errors, and separated prefixes were standardized and those that were incorrectly tagged as an insight was replaced with a different set. A few of these changes are the normalized verbs *meron* (have) into *mayroon*, and joined prefixes *magplano* (plan), *magschedule* (schedule), *magreport* (report), *maglaan* (allocate), and *magdagdag* (add), which all are under the same cluster.

In the same way, target clusters were affected, where some were removed than replaced. In one of the clusters, the shortcuts *mg*, *nmin*, *hrpin*, and *dn* were gone from OAE-N report. Another case added and removed some members in lexicalization of targets. Positive samples for this include meetings-conduct adding ‘citizens’ and responders-respondents adding ‘respondihan’. For new merges, *komunikasyon* (communication) was clustered with *impormasyon* (information), *samahan* (organization) with *komunidad* (community) and *pagpupulong* (assembly), and *daluayan* (pathway) with *kanal* (sewer) and *daanan* (road). Removals on the other hand were caused by changes in extracted insights or disbanding in clusters.

It is important to note that even with the change, there are clusters that remained the same in count, but some still changed in lexicalization or its sub-cluster members. In this case, the change in extracted insights are present, specifically on either the verbs which can change entry frequency counts or nouns which can change the target cluster and sub-cluster members.

Similarly with Dice and Word2Vec, OAE-N experiments did not eliminate but reduced instances of responses that were not clustered but should be. Instances like *turuau* (teach), *paghahanda* (ready), *magimbak* (store), and *ibalita* (report) proved this. Regardless, using FastText in general increased the clusters made (decreased the single frequency entries) and decreased the number of entries as compared to the other two approaches by more than half.

Threshold Adjustments

The result of FastText's OAE-T20 showed the gravity of loosening the threshold. Initially from OAE's 114 entries, its value dropped to six entries, reaching about 99% of the responses clustered. Provided with this statistic, almost each of the clusters collected a large amount of actions and target (presuming the same effect on lexicalizations), generalized with topics about disasters.

The outliers for this experiment, however, are the last two clusters. With the rest of the clusters having a hundred or more frequency counts, the lowest only have a single insight, with the proposed action *nangyayaring* (happening) and target *barangay*. Next to it has two responses, with proposed actions *lumawak* (expand) and transport, and targets *kaalaman* (knowledge) and vehicle.

From the threshold of 20%, each pass of the clusters piled loads of responses together, and as one cluster is created, the next would have fewer selections until there is no more. However, FastText still was unable to include these outliers even though there are keywords that could have been potential members.

The reason for this exclusion is the algorithm works by choosing a base word that would be compared to the rest of the verbs. In this case, cluster one's *magkaisa* (unite), cluster two's *wastong* (proper), cluster three's *put*, and cluster four's *pagtawag* (call) as compared to *nangyayaring*, *lumawak* and transport has lower similarity values – visualizing them on opposite “corners” in the space. As a result, they are on separate clusters.

One solution to merge these clusters is to use other members from existing clusters as the base word. However, doing so might result into having one cluster all in all. Since the main idea of this research is to extract and organize the responses to generate a report that could be easy to interpret and implement, the problem with a single huge cluster is it would be harder to understand; it would represent a vague topic about disasters, which is similar to reading all of the responses in the data.

Tightening the clustering condition, OAE-T80 split the large clusters from OAE. Frequency counts and cluster orders shifted, with more specific topics in clusters. Instead of huge topics about logistics or disasters in general, some of the specific topics of action include procurement (e.g., provide), dissemination (e.g., inform), preparedness (e.g., prepare), sanitation (e.g., linisin), and avoidance (e.g., maiwasan). A downside with the spread of clusters or increase of entries however is there are ideas within separate clusters that duplicates the ideas. Nevertheless, these type of clusters makes it easier to interpret the idea behind each.

Similarly with Word2Vec, OAE-T80 created clusters with only a verb, meaning verbs joined has to be identical to be under a cluster. Examples are *magkaroon* (have), *linisin* (clean), *maging* (become), *kailangan* (need), give, *tumulong* (help), and more. Regarding target clusters, it is clear that at 80% mark, there is a high number of sub-clusters that joined words (mostly in pairs) with orthographic similarity. At the same time, FastText still included instances of semantic similarity such as before-after (antonym/preposition) and for-of (preposition).

Clustered per Response Categories

ORC for FastText produced high frequency clusters despite having the responses be segregated into their corresponding categories. Clusters have mixed topics, where some are specific and other are generic. Examined each category exhibited key insights grouped together by the approach.

First category is Information Campaign and Capacity Building. Inside it vaguely represents a multitude of actions that could be involved in disaster strategies. Among the highest cluster, verbs include alert, remind, gather, give, help, prepare, conducting, teach, provide, making, focusing, participate, reduce, ready, and more; while nouns include ideas, flood, place, sirens, seminars, tips, orientation, garbage, disposal, assembly, family, class, project, LGU, and more. As the same with other experiments, FastText was able to cluster this set with distinct relationship with each other. Other ideas from low frequency entries includes education or dissemination, faster actions, assistance, and construction.

Second, Disaster Relief in total have a few responses under it. Majority of its responses are under one cluster that indicated the need for support in the community. Suggestions under this relate to providing supplies like food and medical kits. It is noticeable that beneath the target field, there have been sub-clusters that would be more appropriate to be clustered than its current grouping. One example is the pair ‘aid’ and ‘need’. The noun ‘aid’ could be clustered with ‘response’ or ‘assistance’ but are either non- or members of a different sub-cluster.

Third and in a similar case as Disaster Relief, Community-wide Logistic Support for Disaster Response accumulated a lot of responses for its cluster with the highest number of frequency counts. Within this cluster involves verbs and nouns that indicates the things needed to be provided or prepared. Instances for the nouns grouped emergency with facility and assistance, supplies with shelter, evacuation, equipment and food, and disaster with facilities and for typhoon and storms. Other ideas that belonged to the same idea but were on another cluster mentioned building or buying an area or place, adding budget and community volunteers, and communicate weather predictions.

Fourth is Infrastructure Maintenance and Management which highlighted ideas for sanitation and repair. There were three major clusters, one that pertains to the primary subjects for maintenance or management, another is for those in need of improvement and prevention, and one of which to incorporate in this category.

Regarding the primary subjects, these are to clean and fix waste, garbage, and sewage. In improvement and prevention, target subjects are drainage (that comes with its sub-clusters areas, streets, canals, dumpster, etc.), system, rules, flow of flood, plan, and community. It is also notable how FastText was able to cluster dumpster with garbage, bins, and *garbagetrash*. Regarding which nouns to incorporate for maintenance and management, specific to providing (magkaroon or magbigay) and reducing (mabawasan), here are the following suggested sub-clusters: *basurahan* or garbage (for floods and clogging), *pagtitipon* (gathering), training (or seminars), and *puno* (trees). It is safe to note that *pagtitipon* could be clustered into training, but similarity seemed to be below the threshold. In addition, there could have been other low frequency clusters, especially the single ones, that FastText should merge. Samples for these are lessen, *dagdagan* (add), and *ngbbgay* (nagbibigay or provide).

Fifth is Early Warning System, another category that has a large number of responses. Under this has one cluster with more than a hundred members, while the rest follow with less than 25. The essence of the highly grouped verbs is to provide ideas relating to Early Warning System, which includes: inform, alert, messaging, texting, *namemegaphone* (use of megaphone), update, utilize, provide, and more. In its target/noun list, there is a large sub-cluster that was lexicalized under *theft*. Its members are: plenty, citizens, places, beforehand, incoming, troubles, people, preparedness, sms, alert, constituents, news, person, everyone, theeveryone, damage, needs, need, weather, changes, things, loss, and more. Based on this, lexicalization might not have been appropriate as this was chosen randomly. It would be better if nouns closely related to theft are the only ones included (e.g., troubles, people, damage, loss, things, news, alert, etc.). In another view, there is a sub-cluster that paired ‘media’ and ‘devices’, which would be more appropriate if the nouns ‘sms’, ‘news’, ‘need’, and ‘things’ are under them.

Sixth, the list of highly suggested actions under Preparedness for Emergency are the following: *magtulong* (help), *magkaroon* (have), *maghanda* (ready), *malam-an* (know), *mapagbigay-alam* (inform), and more. A few of the nouns under these are *balita* (news), *gamit* (things), *disiplina* (discipline), *tulong* (help), emergency plan, and more. Although some of the contents have distinguishable connections with each other, there are also some that are indistinguishable to the category. Example is pairing up *atentibo* (attentive) with *kagamitin* (kagamitan or things), *kapit-bahay* (neighbor) with *pagkain* (food), and *bagay* (thing) with *kasanyan* (kasanyan or skill).

Rest of the clusters showed positive samples such as clustering news with week, radio and times, flood with weather, condition, and calamity, and *balita* (news) with *ulat* (report). Negative samples on the other hand could be combined with existing clusters such as *makipagtutulungan* (help), and *ihanda/handa* (ready) with the top cluster, *ibahagi* (share) with *ibalita* (report), and *maiwasan* (avoid) with *makaiwas*.

Next is Local Government Accountability that indicates the responsibility of the government in disaster planning and response. Under it has the following keywords as its top cluster: enforce, hold, gather, facilitate, surrounds, make, raise, and more. Targets under it pertains to trash, map, response, and laws about (as specified in the sub-cluster) disasters, awareness, officials, news, help, and evacuation. Undeniably, these ideas, especially the last one, can give light to the positive and potential use of the government's power to reinforce the nation's disaster prevention and mitigation.

Then, under Filipino values, clusters highlighted solidarity, that is to keep everyone united and participative in helping each other. The cluster with highest frequency indicated ways to help, with majority of its verbs as variants of its Filipino word *tumulong* (tutulong, magtulong², matulungan, magtulungan, makatulong, magtulong, tulongan, hikayatin, makipagtulongan, maitutulong, tutulongan, magtulongan, and makakatulong). Subjects in helping are community, volunteer, *bata* (child), *pagkain* (food), *magkakapamilya* (families), and more. On the cluster about unity, traits of Filipinos or traits that should be are as follows: concern (malasakit), peace, order, and support (suporta). The rest of the clusters, some positive examples grouped *napinsala* (damaged) with *apektado* (affected) and *nasiraan* (damaged), volunteer with duty, and *pakiki-isa* (solidarity) with *komunidad* (community). Negative examples on the other hand paired share-trick and peace-my.

Last category is Others, where its clusters contain mixed ideas about disasters. In terms of clustered words, examples paired dapat-nararapat, walawalang, humingi-manghingi, tulung-tulong, and humingi-maibigay (antonym). Conversely, it failed to pair maayos-pagsasaayos, emergency-responders, barangay-community, barangay-tao, and the likes. Furthermore, there were samples that clusters have indistinguishable relationships such as pockets-disasters, tas-corrupt, deaf-flashfloods, and train-permission.

Synthesizing FastText clustering for ORC, as it combines certain characteristics in Dice's coefficient and Word2Vec, it could be compared to both results. In fact, results in terms of characteristics in positive and negative aspects are similar, but quality of the clustering has changed. Since FastText has more characteristics taken from the two, there were more combinations as to the resulting clusters; specifically, being able to join orthographic and semantic (thematic similarities, usages, and even antonyms) similarity. At the same time, it also suffered from having indistinguishable relationships and missing some potential responses to be merged with others.

5.2.5 Summary of Qualitative Experiments and Results

Information Organization and Clustering consists of experiments on organizational format, lexicalization, normalization, and threshold adjustment, and analysis on the three clustering approaches used. In organizational format, two perspectives can be taken away from the results. One is through Organize All Entries, where all of the insights can be clustered based on their similarity – regardless of which categories they are under. Hence, in representing clusters, they were focused mainly on actions provided. Two is through Organize by Response Categories, where there is an assertion that the insights are already under a particular category, which makes clusters/actions under it more focused into suggesting steps that contributes to that category.

Top ideas collected from organizing all entries are the following:

- Logistics mentioned items in need such as early warning system, medical kits, flashlight, garbage cans, shelter, medicine, and grocery/supply.
- Dissemination suggested information about disasters in general and specifically on typhoons, storms, floods, and consequences that comes with them.
- Sanitation points mostly towards cleaning the surroundings, specifically sewers and areas near households.

- Community and self-preparedness through alarms and designating places such as schools for evacuations were suggested, mainly to be able to avert or avoid ramifications such as floods or clogging, spread of tragedies, and getting trapped.
- Solidarity recommended families and Barangay to continuously help or support each other.

Based on this, Malasakit's predefined categories adequately covered the dataset. No new categories were discovered, as all are under its scope.

Top ideas when organizing by response categories:

- Information Campaign and Capacity Building focused on actions that the community must have such as items or programs, preparations, support, and logistics for information dissemination. It also enumerated intended contents, medium of information, and receiving end of the information. Its key insights are: (1) conducting programs such as seminars, drills, and assemblies, (2) using infographic materials such as signages, posters, and leaflets, and (3) suggested contents for this include reminders or tips on what to do before, during, and after the calamity, while the target audience are family members.
- Disaster Relief pertained to the same ideas which mentioned receiving or providing assistance. Its key insights are receiving goods, food or grocery, and medicine, whilst providing evacuation options.
- Community-wide Logistic Support for Disaster Response provided ideas that involves safety and support. Its key insights are: (1) to have safety gears, sirens, and shelter for the operations, (2) to add more budget, equipment, and volunteers, (3) boats and storage facilities are must-haves, (4) to build or buy an area or place for disaster response, and (5) is to communicate weather predictions to the community.
- Infrastructure Maintenance and Management were about sanitation and repairs. Cleaning the mentioned areas in the minds of the community would ensure prevention of floods and clogs. Its key insights are: (1) clean surroundings such as streets, sewers, rivers, and garbage waste, (2) to avoid throwing trash everywhere, and (3) there must be a proper place or containers for garbage.

- Early Warning System involved information dissemination. Its key insights are: (1) having proper communication, alert, news, updates, and radio, (2) information should be about disaster and announced to the public, specifically people, citizens, and residents, and (3) alerts with regards to calamities, catastrophe, disaster, typhoon, communication, support, and assembly must be provided.
- Preparedness for Emergency showed ideas in a form of a reminder to people. Its key insights are: (1) to alert and be aware or attentive with nature, news, and officials, (2) medium in reminding people include watching weather forecasts or news in television, and (3) preparation is needed on the following: news, evacuation plan, officials, management, unit, and the community.
- Local Government Accountability indicated expectations of the community with the government. Its key insights are: (1) the government should be accountable for subjects such as disasters, evacuations, posters, cases, and the society, (2) it is expected for them to be ready, cooperative, and able to show and send help to the community, and (3) they should be active and focused on helping people.
- Filipino values contained traits that the whole community must have to get through disasters. Its key insights are: (1) encouraging solidarity, that is to keep everyone united and participative in helping each other, (2) help should be observable in preparing families and community, (3) people must be cooperative with duties, plans, and preparation, and (4) traits mentioned tidiness, readiness, equality, kindheartedness, concern, and order.
- Others contained mixed ideas to improve disaster prevention and mitigation. Its key insights are: (1) calling out corruption, which specifically mentioned the act of putting valuables in pockets, (2) endorsing their satisfaction with decision makers, encouraging to continue their work or activities, and (3) minor and major entities in disasters are mentioned, specifically suggesting communication between deaf, citizens, and responders.

In lexicalization, implementation was successful in joining Noun instances, where a representative word is beside the lexicalized words enclosed in parenthesis. In spite the attempt, there were instances of unrelated lexicalizations. In this experiment, the main takeaway is its effectivity highly depends on the performance of clustering approaches. One thing to note of is it has potential in displaying a single word that can represent an entire set of words given a good clustering approach. Selecting and retaining a single word however is another problem to tackle.

In normalizations, it applied standards such as correcting typographical errors and Filipino Colloquialisms. Even though the degree and coverage of the normalizer was sufficient to correct responses, not all were normalized correctly. Since it was built to be dependent on a statistical model, some instances resulted undesirable insertions and replacements. There were also instances that confuses interlingual homographs with colloquialisms and there were some not covered by the normalizer. Observing its effect, frequency counts fluctuated, shifting the order of clusters, but did not affect the top ones. It has been evident that there were shortcuts previously included in the clusters which was then removed, and new or corrected words appeared. Given this, there have been several instances that shortcuts as such were tagged properly. In a similar way, joining Filipino/Tagalog prefixes with their separated root words as part of the normalization task caused a set of words to appear as insights.

In threshold adjustments, it proved that increasing or decreasing the threshold affects the placement of words in their appropriate clusters. It controls the scope and composition of a cluster's contents and excessive amounts of these adjustments could decline the quality of clusters. Thus, threshold value was set into its default and balanced threshold of 50%.

In comparing the approaches, Dice's coefficient was able to group variants of a word, within the constraints of changes in affixes. Characteristic as such enabled it to cover shortcuts and typographical errors not far from the original's form or structure, even without utilizing a normalizer. Provided, clusters have clearer and interpretable relationships. Main difference between Dice and Word2Vec is vectors are positioned based on usage – so words that operate the same way are closer together in the space.

In word embeddings approaches, both were able to cluster based on usages, thematic similarities, and even those with orthographic similarities. Examples for these covered synonyms, antonyms, tense variants, code-switching, shortcut or typos, and other topic/themes. Coverage of orthographic similarity is more evident on FastText as it makes use of sub-word information (character n-grams) as vector representation. With this, verb tenses such as improve and improved were captured; as well as intra-word code-switching such as maimprove. Moreover, it also inherits Dice's capability in clustering shortcuts, one that is not covered by Word2Vec.

Based on clustering results, it is undoubtedly that all approaches were able to fulfill their purposes, capturing and joining conceptually similar ideas together and in some instances formed one, interpretable idea in clusters. It is important to note that based on the clustering process, an insight can only belong to one cluster. Meaning, regardless of formatting, there were no instances of an insight

appearing in multiple clusters. Furthermore, for organize by response categories, an insight under a category could not be placed or clustered in another category. Although, there are instances that separate clusters under different categories have similar ideas or subjects in their insights. An example for this is the need for an assembly which separate insights were mentioned on Information Campaign and Capacity Building and Early Warning System.

Characteristics in positive and negative aspects of the approaches are similar, but quality of clustering has changed. FastText has more characteristics taken from the other two, hence there were more combinations as to the resulting clusters. Comprehensively, these three still have room for improvement, especially in capturing the right balance of relationships between words as there were instances of vague or unrelated relationships.

Negative samples for Dice's coefficient mistakenly joined evacuation-elevation and pagkain-pagkalap, and failed to join magkaisa-makiisa, maglinis-clean, create-make, due to far string similarity distances or the connection between words are through their meanings. For Word2Vec, it failed to join prepare-preparing, giving-pagbibigay, maglinis-clean, announce-inform, and linisin-maglinis. For FastText, it joined a bunch of words indistinguishable from each other such as warning, do, make, be, have, coming, drive, eat, visit, update, give, avoid, said, watch, provide, and failed to join words such as magkaisa-makiisa, tutulong-magtutulungan, and matutong-matutunang.

5.3 Survey on the API and Report

In evaluating the outputs of this study, survey was performed with questions pertaining to their quality and discussions on user feedback was provided.

5.3.1 API Functionalities

Application Programming Interface (API) is a collection of functions which is shown at Appendix B. It has nine parts that represents the tool's modules, namely Data Utilities, Normalization, Language Identification, Filipino Part-of-Speech Tagger, Information Extraction, Information Organization, Information Clustering, Information Ranking, and Report Generation.

Data Utilities module (see Appendix B.1) contains 11 functions. These functions enable the user to process files outside the program. Main functionalities permit the user to read and write on text or excel files.

Normalization module (see Appendix B.2) contains 10 functions, that could be used to utilize or customize the normalizer. There are two normalizers that can be used in this API, namely Nocon et al.'s (2018) normalizer and another that joins prefixes with root words using Oco and Borra's (2011). As default, both are used. There are two ways that normalizations could be done, one is per string or sentence, and another through a list or by batch. The rest of the functions act as means to modify the configuration, that is by setting file paths.

Language Identification module (see Appendix B.3) contains four functions. These can be used to indicate the language a particular string is under. There are two ways it could be used, one is using a string as input (per sentence evaluation), and another using string or object lists (by batch evaluation). Either way, its output provides a tuple of language and confidence value. Additional functionality enables the user to modify the coverage of languages for identification.

The Filipino Part-of-Speech Tagger module (see Appendix B.4) contains six functions, that makes use of FSPOST (Go & Nocon, 2017). Three kinds of functions can be seen in the list, which are about modifying the file path (to call the tagger), tagging options (per string, by object or string list), and formatting options (Part-of-Speech only or Stanford word|tag Format). By default, tagging format displays a word and Part-of-Speech tuple.

Information Extraction module (see Appendix B.5) contains only two functions. These functions enable the user to extract insights in two formats. One is insight phrases which extracts a string starting from a Verb up to a Noun. Another does the same process but formats it with only the Verb and Nouns inside a sub-list or tuple. It is safe to note that on this module, the extractions were made specific to Malasakit responses. Extractions are performed in an object with the following attributes: Response ID, the Response itself, its Response Category, Language identifier, FSPOST tags, container for the extracted insights, and location (a field that can be added by users).

Information Organization module (see Appendix B.6) contains three functions that sets the formatting style of the clusters and report. Main functionalities enable the user to organize their extractions or results all in all or uses a per category style of formatting.

Information Clustering module (see Appendix B.7) contains 11 functions. Three parts can be taken from this list. First is the main function that invokes the clustering algorithm (i.e., Dice's Coefficient, Word2Vec, or FastText). Then, supporting functions that can retrieve insights from the Malasakit object, remove duplicates in clusters, flatten insights in the cluster, and cluster/lexicalize target/noun words. Last is a list of functions that computes for distance or similarity values between two strings using the selected approach.

Information Ranking module (see Appendix B.8) contains only two functions. A function that ranks the clusters by frequency and arranges them in descending order (highest count first, lowest comes last); and a function that ranks by cluster categories, arranging the order of categories (category prioritization has been determined beforehand) and per categories ranks them by frequency in descending order.

Report Generation module (see Appendix B.9) contains seven functions. Its main function generates the report in a Microsoft Word document. The other five are functions that was used as support in generating the report. Examples for this includes adding a timestamp, divider, title, and setting the document margins and page columns. The idea for separating or having these functions is for future applications that intends to create their own format in the report.

As a whole, the functions listed are intended to be useful in future researches or application not necessarily within Malasakit that involves the tasks of data processing, information extraction, language identification, part-of-speech tagging, word clustering, and text ranking.

5.3.2 Report Formatting

The report generated by the tool is a two-column Microsoft word document, with extracted insights/suggestions organized in two ways, organized all entries (OAE) by frequency (see Figure C.2) and organized per response category (ORC) and frequency (see Figure C.1). Each report contains three parts: Introduction, Insight List, and the Malasakit Response List (see Figure C.3).

In the Introduction, a title can be seen above the document's production timestamp (in Month-Day-Year and HH:MM:SS format). Below them is a short description which informs the readers that the suggestions that has been provided was automatically extracted and organized.

In the Insight List, Entries or also considered as clusters are numbered. Under that entry heading, fields such as Sentence ID numbers (of cluster members), Fre-

quency Counts (number of responses under the cluster), Proposed Action (verbs extracted and joined together), and Target (nouns under those verbs clustered) were included.

Moreover, lexicalization, a process in placing a word that can represent a cluster, has been applied in Target fields, resulting into sub-clusters enclosed in parentheses. The only difference between OAE and ORC in terms of format is ORC's category heading enclosed with two horizontal column-length lines. The length of this part varies, depending on the experiment (approach and configurations used).

In the Malasakit Response List, the full responses in the data are provided. It is formatted in a table containing two columns, one for Sentence ID or row number in the data, and the next with the responses (sentences of suggestions). The length for this part is 33 pages.

5.3.3 Survey Procedures

In conducting the survey, several steps were followed for its completion. Prior to answering the questionnaires, an informed consent form (see Appendix E) was provided indicating the purpose, procedures, duration, participation and confidentiality notes, risks, benefits, contact information, and consent signatures.

Under the purpose of the survey, it has been indicated details about the research (what is it about or what is it for) and how the respondents relate to it or why are they chosen to partake in it. For procedures, a step-by-step on how the survey is going to be performed was iterated. In duration, an expected time for completion was provided. For participation, it has been noted that the survey is voluntary and that respondents can withdraw at any time. In confidentiality, assurances and steps that will be done to protect the respondents was described. For risks and benefits, it showed how respondents will be responsible for the effects and if there are negative ones, they will be addressed immediately. In contact information, details about Research Ethics Committee in De La Salle University was provided. Finally, are the concluding remarks and proof of participation in the survey.

After signing the consent form, materials for the survey such as supporting documents and API were provided. The documents include the copy of the consent form for both parties, survey questionnaires and API functions list (see Appendix B). There are two questionnaires (see Appendix F), one intended for assessing the API and another for the report.

The API questionnaire consists of two parts: a quantitative and qualitative part, modified from the USE Questionnaire (Lund, 2001) and NormAPI (Nocon, Cuevas, Magat, Suministrado, & Cheng, 2014, as cited in Regalado et al., 2015) version of the USE Questionnaire. In the quantitative part, every question is measured by numbers from one to five (5-point Likert scale), where it indicates in the same order strongly disagree, disagree, neutral, agree, and strongly agree. It has been divided under four themes, namely usefulness (e.g., it is effective partnered with other software tools), ease of use (e.g., it is user friendly), ease of learning (e.g., I easily remember how to use it), and satisfaction (e.g., I would recommend it to be used in the future). In qualitative, there are three items that needed opinions about. The statement is to provide the functions used (or functions that will be most likely useful to me), comments and suggestions about the API, and negative aspects (if there are any).

Uniformly, the Report questionnaire has two parts and similar assessment measurement. It has been created from the TAM Model (Davis, Bagozzi, & Warshaw, 1989) that measures a tool's perceived usefulness and ease of use. It has been indicated in the survey the primary targets for its quantitative part are focused in terms of design (e.g., the report's design is acceptable), quality (e.g., the information fields are appropriate and enough to make a decision or action), format (e.g., the information is easy to interpret), efficiency (e.g., using the report in my job would enable me to accomplish tasks more quickly), and potential (e.g., the report would enhance my effectiveness on the job). For its qualitative assessment, it has been stated to indicate the organizational preference (which approach and format), comments and suggestions about the report, and its negative aspects.

5.3.4 Survey Results

Initially, the survey was intended to be answered by decision makers. However, at this time writing, Coronavirus disease (COVID-19) enveloped the world, making particular organizations be preoccupied with the disaster or closed due to quarantine. Instead, the survey was conducted on five (5) Malasakit team members as its respondents. They were chosen to assess the content of the generated report and the developed software tool, due to their knowledge and exposure to the topic and technologies used, as well as their accessibility to the tool and availability. Particularly, this survey focuses on the report's aesthetics and information quality, and the tool's usability. Their feedback determined the impact of the research, specifically its overall usability, and were able to provide ideas for improving the report and tool (see Appendix G for their answers).

API Assessment

In assessing the API, quantitative scores were provided with criteria under Usefulness (see Table 5.11), Ease of Use (see Table 5.12), Ease of Learning (see Table 5.13), and Satisfaction (see Table 5.14). The format in tallying the result was based on Nocon et al. (2014), which accommodates the scores regardless of the number of participants.

In each row displays the description or questions under a criterion. It uses a 5-point Likert scale with numbers one to five representing strongly disagree to strongly agree. Under the scores indicate how many respondents chose that value (blank fields count as zero), rather than putting scores per respondent. At the last column, average scores per description was provided.

As a whole, the API was rated with 4.07 out of five score, representing agreeable to strongly agreeable elements. It excelled in terms of usefulness and user satisfaction, indicating its importance and potential, and fulfilling its current expectations.

Table 5.11: API Results on Usefulness

Description	1	2	3	4	5	Avg.
It is effective partnered with other software tools			1	1	3	4.4
It can help raise the productivity rate when used in conjunction with other software tools				2	3	4.6
It is useful for my tasks			2	1	2	4.0
It makes it easier to accomplish my tasks			2		3	4.2
It helps save time				2	3	4.6
It does everything I would expect it to do			2	1	2	4.0

Particularly, its overall rating in usefulness criterion produced a value of 4.3 (see Table 5.11). It has been described to excel in providing increased productivity rate when partnered with other tools and reduces time in accomplishing tasks. Moreover, applying it to other tools was deemed effective. Lowest ratings for this criterion got the average of 4.0, a point lower to the highest score, which described how useful the tool was to the participants' tasks and if it runs how they expect it to be.

Table 5.12: API Results on Ease of Use

Description	1	2	3	4	5	Avg.
It is easy to use			1	2	2	4.2
It is simple to use			2	1	2	4
It is user friendly	1	1	3			3.4
It requires the fewest steps possible to accomplish what I want to do with it			2	2	1	3.8
It is flexible			1	4		3.8
Using it is effortless	1	2	2			3.2
I can use it without written instructions	1	2	2			3.2
I don't notice any inconsistencies as I use it			2	2	1	3.8
Both occasional and regular users would like it			1	3	1	4
I can recover from mistakes quickly and easily			2	2	1	3.8
I can use it successfully every time			1	2	2	4.2

In Ease of Use, its overall rating garnered 3.76 – the lowest among all criterion (see Table 5.12). With this in mind, unfavored ones described the API as difficult to use, especially without proper documentation. Furthermore, there were ratings as low as 2 points which pertains to user-friendliness. Despite this, there were participants that still found the API easy and simple to use. In addition, it also runs successfully when they use it. Even having in mind that with these characteristics, occasional and regular users would like its functionalities. Other distinguishable characteristic pertains to the modularity of the API's functions, where it got four out of five participants agreeing that it is flexible.

Table 5.13: API Results on Ease of Learning

Description	1	2	3	4	5	Avg.
I learned to use it quickly	1		4			3.6
I easily remember how to use it			1	1	3	4.4
Learning how to use it is easy		1		4		3.6
I quickly became skillful with using it			2	3		3.6

For Ease of Learning, it got an overall rating of 3.8 (see Table 5.13). Supporting this, three descriptions under this criterion got the same average score of 3.6. They are regarding the learning curve involved in using the tool; where in terms of time invested and difficulty, there were four out of five participants that agreed to it to be fast and easy, while one found it hard. Correspondingly, applying it in their programs, most have been skillful in doing so. As a matter of fact, the highest description under this criterion is how easy a user can be accustomed to the API.

Table 5.14: API Results on Satisfaction

Description	1	2	3	4	5	Avg.
I am satisfied with it			1		4	4.6
I would recommend it to be used in the future				1	4	4.8
It works the way I want it to work			1	4		3.8

Last criterion is on Satisfaction. It got 4.4 overall rating, the highest among other criteria (see Table 5.14). It describes how satisfied the users are, if it works as intended, and if it would be recommended to others. Majority of the participants were strongly satisfied and positive with recommending it. Additionally, most agreed that the API worked the way they want or expected it to be. There are, however, an account that the participant is neutral with the expectations and being satisfied.

On another part of the survey form, a qualitative section was provided. It has three fields which asked for the participants' functions that were used in the evaluation or those likely useful to them, comments and suggestions about the tool, and its negative aspects, if there are any. Under this are answers that shed light upon ideas on improvements for the API.

Listing the functions used, every module was utilized, spread across the five participants. All of them were able to try out the organization or clustering modules. Specifically mentioned under this module are the functions for organizing response categories and sub-lists. Apart from that, the major tasks (i.e., Part-of-Speech Tagging, Information Extraction, and Information Ranking) in the tool was also tested. Specific to Part-of-Speech Tagging, formatting options was applied. For Information Extraction, participants tried out extracting word sets. In ranking module, ranking by response categories was tested. Additionally on these major tasks, supporting tasks such as Data Utilities (e.g., read excel file, refresh excel cells, and write report), Language Identification (e.g., identify language of a string) and Normalization (e.g., normalize a list, normalize a string, and translate Filipino colloquialism) was utilized.

Taking the respondents' comments and suggestions, there are a lot of inputs to consider. One of the highlights is their appreciation with the tool. Quoting from one of the respondents, "...this can be very useful especially that this is tested on a domain in the Filipino language. Users can now have a toolkit that can be used for preprocessing and information extraction for their Filipino dataset." Adding from another comment, it has been stated that the functions created can be "... used externally for other data processing tasks."

In addition, there were positive comments about the API being properly documented and readable. Even included a comment where atomic functions were very detailed. However, not all were satisfied with this documentation. Describing the current documentation, outside the API, respondents were provided with a list of functions and its description. While inside the API, each function consists of block comments that describes what it does, its parameters, and return values. Not to mention that during execution of the functions, there are progress updates to provide awareness on the events happening inside the program. Despite this, based on the comments provided, these are not enough.

Suggestions to improve documentation contain three main ideas. First is to make a README file that provides a brief overview on the API and state tasks that indicates its intended use. Second is to include a requirements list – these are Python modules or packages that is needed to be installed in order to properly use the API. Last, is to provide premade or sample scripts that utilizes the functions, specifying its input, the set of codes to run it, and how the output will be shown. In fact, one of the negative aspects stated its lack of requirements and instructions on how to use it. Rounding up the ideas about documentation, development of these documents would boost the low criteria ease of use and learning. One note is to make use of API documentation tools to accomplish this.

Following up the suggestions, respondents listed options in the future to extend this API/tool that addresses improvements in ease of use. One is to commercialize it by making it into a RESTful API, using HTTP requests; for reference, REST (Representational State Transfer) is an architecture for web APIs. Another is to package it in a way that utilizes pip install. Last is to introduce a Graphical User Interface (GUI).

For issues and negative aspects of the API, aside from documentations, a participant noticed tracing prompts on the runtime updates, which indicated “No insights (words) extracted at Response #: <response_ID>”. In the participant’s perspective, it appeared as errors, but in this case indicated that the Part-of-Speech tagger and Information Extraction’s pattern matching were not able to capture such ideas. One way to mend this is to remove tracing prompts but leave the module updates intact. Another issue, in one of the participant’s account, language identification was unable to run as it was not recognized in the program. This proves the importance of having the README and requirements file, to prevent instances such as the one provided.

Issues addressed in the API consists of creating supporting documents such as Technical Manual, User’s Manual, and README file. To ensure proper running of modules, requirements.txt were also created. These files were packaged together with the API and uploaded on Github Repository and Google Drive.

Report Assessment

Table 5.15: Report Survey Results

Description	1	2	3	4	5	Avg.
The report's design is acceptable			1	3	1	4
The different elements of the report (e.g., title, insights list, and Malasakit response list) is necessary				3	2	4.4
The information fields are appropriate and enough to make a decision or action				4	1	4.2
The information is clear and readable				4	1	4.2
The information is easy to interpret	1	1	3			3.4
The information is useful in my job			1	1	3	4.4
The information is displayed in an organized manner			1	1	3	4.4
Using the report in my job would enable me to accomplish tasks more quickly				3	2	4.4
The report would enhance my effectiveness on the job			1	1	3	4.4
I can make decision/s based on the information provided				3	2	4.4
I am satisfied with the report				2	3	4.6

In assessing the generated report, 11 descriptions were provided for its quantitative evaluation (see Table 5.15). The format of presenting the results is similar to the quantitative evaluation of the API. Overall, the report has been rated 4.25, a value that resides with agreeable to strongly agreeable elements.

Based on the results, the highest average score garnered 4.6, where all participants agreeing to being satisfied with the report. It then was followed with six descriptions tied for the value of 4.4, in which answers mixed counts on either agree or strongly agree. Descriptions involved the appropriateness of the elements in the report, its impact to the participant's job and tasks, presentation or organization style of the information, and usefulness of the information.

The lowest score on the other hand, got 3.4 in average. It states that interpreting the information was found to be easy for most of the participants, but there were some that were neutral and even disagree with the level of difficulty. Despite having this, participants agreed that contents are clear, readable, and sufficient make decisions based on the information. With this, it goes to show the need to

provide a background on what to expect in the report and instructions on how to do interpretations given the information – to conveniently push users to easily learn making use of the report.

Exploring ideas to improve the report, three fields were provided to take these inputs namely, to provide their organizational and approach preference, comments and suggestions, and negative aspects of the report. In organizational preference, most preferred organized by response categories (ORC), where each response is organized per categories first then by frequency.

Although, one of the participants noticed an important element in clustering the entries entirely or using organized all entries (OAE); quoting, “having a frequency only ranking for the information also has its merits since it will give the users an idea on which category is most important followed by the rest.” In this perspective, the participant was able to point out the importance of OAE, which is to show the urgent topic or category, actions, and target subjects that is in need of prioritization in disaster planning.

On another perspective, there was one preference that indicated a reverse order, grouping all entries by frequency first then order by categories. In this case, using the response categories may be inappropriate as one cluster could belong to multiple response categories. The task that would be more appropriate for this is applying a topic modeling task; where given a set of clusters, they will be labeled depending on a topic they are generally under. Nonetheless, this is an idea for future extensions.

Regarding the participants’ preference in the clustering approach. The setup prior to assessment provided six sets of report with a combination of the organizational format (i.e., OAE and ORC) and clustering approaches (i.e., Dice’s Coefficient, Word2Vec, and FastText). The approaches were labeled A, B, and C, respectively, to remove bias on the innovative approaches.

Based on the results, four out of five participants preferred how the information were clustered through Dice’s Coefficient. Due to the fact that it displays a clearer relationship on the actions and targets through its orthographic similarities, as compared to semantic similarities which have instances of vaguely related and diverse clusters. Complementing with ORC, the layout looked more presentable and easier to interpret. Quoting from one of the answers, “...even though the proposed action is limited, insights are better gained using this approach by being specific on what the proposed action is”. One participant, however, have a different opinion and pointed out FastText’s output to be easier and clearer.

Collecting their comments and suggestions, the reports were judged to be “aesthetically formal and proper to look at.” The information is understandable and actionable. However, some initially found it hard to understand, which took time before fully understanding what to do with the information. Having said, the report was not as intuitive as expected. Factors for this are the length of pages and contents in the reports.

With regards to the pages, participants found it overwhelming as it contains pages that stretch around 30 or more pages. Suggestions to reduce this include the removal of entries with frequency counts of one. As stated by one of the participants, it is not necessary as those could be redundant ideas from top or highly frequent clusters. For the same reason, entries could also be limited into the top 5, 10 or 25 entries. Another idea is to add a layer of generalization to the clusters. It could be through lexicalization of actions (and omitting members) or placing topic labels to aid categories whilst summarizing contents of the cluster.

Alternatively, a different way of displaying the information can be supplied, not as replacement but as aid. The idea of showing keywords, by only looking at the action and its target subject, was not easily interpreted by some. There might be a need for showing context with less effort than redirecting from sentence numbers to Malasakit Response list. In this situation, it could be an infographic or visualization that would give an overview in the contents of the report – particularly urgent steps to prevent and mitigate disasters. With this in mind, it would seem that the current report will be used when the user is ready for the specifics. Example suggestion to accomplish this is through “... simple generated graphs such as word clouds or any frequency-based visualization”, applying User Interface Design or User Experience design methodologies.

With regards to the contents, few of the respondents indicated improvements in clustering. Even though the clusters were said to be understandable, they found the proposed actions broad and disconnected with the target nouns. As a result, there were verbs that lack additional, relatable information. Take for example the word ‘mag’, that since it is a prefix for Filipino verbs, it covers a wide range of words that it could be under. One solution for this is to implement the normalizer in the report or filter the clustered words to reduce and narrow down the ideas. Granted, these actions will improve the value of clustered contents, not just for disaster planning’s benefit, but also for developers that would implement the tool.

A Hypertext Markup Language (HTML) version (Figure 5.1) was supplied to address the respondents’ comments about the report. It is automatically generated by reading an excel file with cluster values. It uses Bootstrap for its design and can support the two organizational formats such as organize all entries and organize by response categories.

Clustered Insights Report

The information below were extracted and organized automatically. Each cluster contain ideas collected from local communities about preventing and mitigating disaster in the Philippines. *Number of Entries* indicates the number of times an idea was mentioned, *Proposed action* indicates the suggested actions to take, and *Target* indicates which subjects should the actions be applied to.

Pro tip: Click on Target words to display the original post.

Cluster List

INFORMATION CAMPAIGN AND CAPACITY BUILDING

Cluster 1

Number of Entries: 20

Proposed Action: dapat

Targets: barangay, seminar, oras, prepared, kalamidad, sakuna, posters, paalala, drill, like, for, example, about, paraan, encourage, before, during, and, after, of, the, calamity, my, weekly, mg, roong

Figure 5.1: HTML Report: Front Page

On its front page, a brief description was provided to give an idea on what to expect in the report. Similar to the Word Document version, it shows the list of clusters (Figure 5.2) with the number of entries, proposed action, and target fields, and Malasakit Response List (Figure 5.3). Main intention is to make the report readable by adjusting the contents included, while maintaining its simplicity.

Cluster List

INFORMATION CAMPAIGN AND CAPACITY BUILDING

Cluster 1

Number of Entries: 20

Proposed Action: dapat

Targets: barangay, seminar, oras, prepared, kalamidad, sakuna, posters, paalala, drill, like, for, example, about, paraan, encourage, before, during, and, after, of, the, calamity, my, weekly, mg, roong

Sample Cluster Members

Response 26: dapat ang barangay ay sanayin ang mga naninirahan sa kanilang lugar tungkol sa drim na may sapat na pangangailangan sa pagsapit ng bagyo baha

Response 208: dapat magkaroon ng palagiang seminar upang maging handa ang mga tao sa nasaabing lugar o pinamumunuan barangay

Response 209: magkaroon ng seminar para malaman ng mga tao kung ano ang dapat gawin sa oras ng sakuna

Response 339: sana po magkaroon ng mga awareness programs po sa schools and sa community dapat po maging prepared din po bawat family at irequired sila na magkaroon po ng mga emergency kits

Response 507: sa pagpapaalala saamen ang mga dapat gawin upang makapaghanda sa kalamidad

Figure 5.2: HTML Report: Clusters

Additional features that comes with this report, as indicated on it, includes target words assigned with hyperlinks, that upon clicking them redirects users to the original post. A *scroll-to-top* functionality was also provided for convenience. Moreover, for every cluster, at most five sample cluster members are shown.

Malasakit Response List

Response #1: magkaisa dapat ang mga tao
Response #2: mag karon ng pagkakaisa upang sa mga darating na mga sakuna ay malalagpasan
Response #3: magkaroong ng komunikasyon kung saan magkikita sa panahon ng kalamidad
Response #4: paglinis ng kanal wastong pagtatapon ng basura at kailangan mag ikot ikot ang mga tanod upang bantayan mga gamit ng tao
Response #5: malawakang information drive
Response #6: bago dumating ang bagyo magkaroong ng early warning system para mas maging handa ang mga tao
Response #7: magkaroong ng early warning upang maging handa ang mga tao sa darating na bagyo
Response #8: higit na pagtibayin ang early warning system device magkaroong ng maintenance quarterly para masigurong mayos ito bago dumating ang isang kalamidad
Response #9: lalo pang lumawat at lumago ang pagmamalasakit sa aming ka-barangay
Response #10: pagsunod sa sinasabi sa kung ano ang dapat gawin paghandaan ang lahat ng bibitbitin sa tuwing may sakuna
Response #11: pagkakaroon ng early warning device
Response #12: pagbibigay ng humanitarian assistance goods sa panahon ng kalamidad
Response #13: nais ko po sana makaroon da po norma ibano paraan uoano lalo nano macino handa ano amino moa kanitbahav do



Figure 5.3: HTML Report: Malasakit Response List

Aside from the two, generating the report enables users to specify limit on the number of clusters displayed. It has been included to reduce the shown clusters than overwhelm users with information. With this feature, top (e.g., 5, 10, or 25) clusters are shown, lessening cluster instances with frequency of one.

Given this implementation, there are other improvements that could be done for better user experience. The most straight-forward way is to extend the functionalities and design of this report. It could be done by making hyperlinks and original posts more accessible through pop-ups or collapsing text components. Although, it is highly recommended to present the information through graphs or word networks (Figure 5.4) – as they could link information with respective fields and posts whilst providing better design, interactivity, features, and usability.

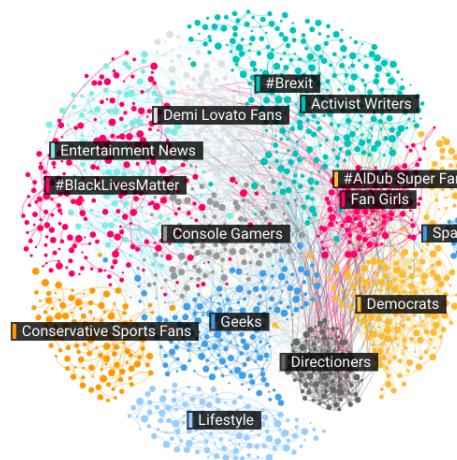


Figure 5.4: Word Network or Graph²

²Source: <https://www.business2community.com/big-data/data-shows-millennials-use-2016-word-year-01736594>.

5.4 Test on News Dataset

In terms of data, there was a test on another dataset in another domain. The assumption is that since the solution processes texts, it should be able to handle inputs regardless of the domain. This test is a proof of concept that it can handle any text given the current implementation. The domain of choice was News.

The News dataset was taken from a Philippine language resource collection (Oco, Syliongka, Allman, & Roxas, 2016). Sources used for this experiment are from *Pang-Masa* and *Pilipino Star Ngayon* 2015 which are sister newspapers under the latter. Both are in Filipino and has collected 576 and 1,013 articles, respectively. Moreover, the two are considered as the leading tabloid newspaper in the Philippines which makes it ideal as subject for this experiment.

2,000 sentences were randomly taken from each of the categories under each newspaper. Pang-Masa has labels with Movies, *Punto Mo* (Your Point), Sports, and Metro, while Pilipino Star Ngayon has *Bansa* (Nation), Metro, *Opinyon* (Opinion), *Palaro* (Sports), *Probinsya* (Province), and Showbiz. As part of the experiment, these categories were merged and labeled into four main categories based on their topic or theme, namely Entertainment (for Movies and Showbiz), Opinion (for Your Point and Opinion), Sports (for Palaro and Sports), and World (for Metro, Nation, and Province). Having each category with 200 sentences, the highest number of contents is World category, with 800.

In this news dataset, sentences were cleaned and tokenized, then processed through the main tasks: Information Extraction (making use of Language Identification and Part-of-Speech Tagger) and Information Organization/Clustering. Organization were implemented to process by category with category prioritization set by alphabetical order. Similar to previous experiments, in terms of clustering, the dataset undergone Dice's Coefficient, Word2Vec, and FastText clustering.

5.4.1 General Statistics

There were three runs on this experiment, one for each clustering approach. As this is a qualitative analysis, the only statistics provided indicates the runtime and number of pages. As for the extracted insight phrases and word sets, and entry/cluster counts, it is provided on their respective sections.

Generally, processing the 2,000 sentences dataset took more than an hour to finish running all the modules. Among these tasks, the longest processing time is the Part-of-Speech Tagger with around 50 minutes to an hour to complete. Two other tasks also contribute to lengthening the time, that is the Information Clustering and Report Generation.

Specifically, processing through Dice's Coefficient completed its execution time with one hour, three minutes and one second. With the number of sentences and entries produced, Report Generation took 8 minutes and 58 seconds to complete. Other than the two, the rest of the tasks were under 11 seconds. It produced 114 pages worth of clusters/insights and adding 77 pages for a list of the 2,000 sentences, the total pages for this report is 191.

Another, using Word2Vec, its total execution time was one hour, 10 minutes, and 48 seconds. The report was generated in 10 minutes and six seconds, while its clustering task took four minutes and 34 seconds. It produced 108 pages worth of clusters/insights, which in total is 185 pages. A smaller number of pages, given that the extractions are consistent, means that there were more sentences clustered as compared to Dice's.

The third approach used FastText. Its execution time lasted for one hour, 31 minutes, and 40 seconds. Report generation took less than 10 minutes (exactly nine minutes and 14 seconds), while its clustering task took almost 30 minutes (exactly 28 minutes and 14 seconds). Among the approaches, FastText took the longest time in completing the whole task. It produced 52 pages worth of clusters/insights, which in total has 129 pages. By far, FastText took the longest but in turn provided lower number of pages. That having said, more ideas were clustered in this approach.

5.4.2 Information Extraction

Expectation for this task is to produce examples that indicates an action and a target that is the subject of that action. Based on the results, the information extracted was consistent, regardless of the clustering approach. There are two types of formatting, namely insight phrases and word sets. For insight phrases, it produced 3,490 entries from 2,000 sentences. Whereas, there were 3,503 entries for word sets. Phrases proves to be usable by providing readability for analyzing extracted ideas, while word sets are for clusters. Together, they point into the same ideas that can be taken from the news snippet.

Table 5.16: Extracted Information in News Dataset

Category	Examples
Entertainment	Humihingi na siya ng pambili ng pagkain
	Magkaroon ng healthy lifestyle
	i-promote ang teleserye
	Nag-tweet si Vice
	natuwa ang fans
Opinion	NAGULAT ang mga kidnaper
	masagot ang nasabing mga tanong
	hilingin ni Pope Francis
	Lumalabas ang sintomas
	nakuha ang plaka
Sports	sumabak si Pacquiao
	Tumanggap diumano ng 3 million
	sinimulan na ang fitness test
	mapapanood ng live
	idineklarang may injury
World	binili ng Ecowaste Coalition
	inaprubahan ang bail petitions
	Ituturo sa lahat ang tamang paghuhugas
	nakaresponde ang mga bumbero
	patay sa ligaw na bala

The algorithm for extraction makes use of Part-of-Speech patterns. It looks for a verb tag and records it until it reaches a noun group. Among the insights, this extraction algorithm exhibited decent results. It shows that as long as the Part-of-Speech tagger is reliable, it would be able to process texts that can point out the actions found on a sentence and record that phrase until its subject, which is a noun is found. Examples of these phrases taken from each category are shown at Table 5.16. In addition, as the tool covers English, there are positive instances of extracting English phrases. Few of these are “controlled the game”, “received proper medical attention”, “plays the title role”, “worked hard for the movie”, and “uniting our people”.

However, not all ranges of verbs to nouns provide a legitimate verb to noun entry. In some cases, the Part-of-Speech tagger failed to label the words correctly, producing incorrect entries such as “asul na van” (blue van), “liblib na lugar” (secluded place), “bandang alas-7” (around 7), “33-anyos na suspek” (33 year old suspect), “nationwide sa January” (nationwide on January), and “nakaraang week” (last/past week). These examples do not point to an action instead describes a noun. Moreover, there are some that does not exhibit enough idea,

meaning there are lacking elements to it (e.g., am Maldives, sa semis, ani Chulani, is what St Paul, taus puso, and more). Furthermore, out of 2,000 sentences, only 1,689 sentences have insights extracted, that is 311 sentences without insights.

On the condition that extracting information in the news domain has been successful to capture actions and targets, the main question is its potential beneficial use in the news/media setting. A few of the ideas is insight phrases can provide a summary or gist of the contents in an article. By taking parts of the article, specifically the actions or events in it, readers would have an idea on what already happened in the article. At the same time, word sets can be used as keywords or tags, displaying the main occurrences (e.g., pinatay, nadakip, nakaresponde, mapatalsik, nangunguna, etc.) and involved subjects (e.g., which celebrity, victim, government official, sports team, area/country, etc.) of the article. Regardless, this experiment showed the potential in being applied to other domains.

5.4.3 Information Clustering

Three approaches were used for this experiment, namely Dice's Coefficient for string similarity and Word2Vec and FastText for semantic similarity. Discussions about the quality and contents of the clusters are provided.

Using Dice's Coefficient

In this clustering approach, it produced a total of 1,663 entries spread amongst four categories (shown at Table 5.17). The highest number of sentences clustered together is 51, exhibiting action words or ideas used in this category is closely related. In terms of cluster entry counts (CEC), over 25% of the entries have at least two news sentences or frequency in it. Highest CEC percentage is under World with 41.41%. On the other hand, cluster insight counts (CIC) percentages were above 55%, which means at least half of the sentences were clustered into an entry. Highest CIC percentage is 77.52% under World category.

Table 5.17: Report Composition of News through Dice's Coefficient Clustering

Category	Total Entries	CEC	CIC	Highest Freq.	Total Freq.
Entertainment	352	26.14	58.73	27	623
Opinion	356	27.81	55.69	24	580
Sports	332	41.27	71.15	21	676
World	623	41.41	77.52	51	1624

Each category produced clusters that aimed to connect two or more words together based on their orthographic similarity. In Entertainment category, its top cluster joined the exact verbs *may* (to possess or have). Its targets pointed to articles with topics that indicated possessing: *plano* (plan), lovescenes, crush, *problema* (problem), birthday, *laban* (fight/game), and more. Although the clusters were orthographically joined, consolidating the ideas can be represented by the verbs it is under. Beside this, lexicalizing combined nouns under a cluster, which produced a notable pair *ugali-galit* (habit-anger) coincidentally showing semantic similarity.

Other clusters or ideas in Entertainment showed actions that a target has become (naging) something, does not have (walang) something, is with or together with (nakasama or makakasama) someone, huge (malaking) physical or emotional aspect, been seen (makikita), is attractive (maganda), and more. Other target or nouns provided names of celebrities (e.g., Vice, Sarah Geronimo, and Gary Valenciano), politicians/positions (e.g., chairman, senator, and presidente), characteristics (e.g., happy, healthy, and kalungkutan), and other related keywords in Entertainment (e.g., aktor, direktor, ratings, fans, and teleserye).

Identically for Opinion, the most frequent cluster contain the verb *may*. In this case, target contains assorted ideas due to its category where articles are based on author's perspective or judgement on a matter. Given that, the nouns found for this cluster indicates what the articles have, such as there is a cobra, rally, hidden camera, internet, *terorista* (terrorist), *reklamo* (complaint), and more.

Other highly frequent clusters have the following base verbs (verbs that was compared): *nawala* (lost), *kailangan* (need), *maging* (become), *dapat* (should), *mayroon* (there is), *bagong* (new), *sinabi* (told), *nangyari* (happened), and more. Nouns on the other hand have distinct orthographic composition, to the point that there was not a single cluster with a lexicalized or sub-clustered target.

For Sports, the most frequent cluster contain a pair of verbs *matapos* (finish or after) and *matatapik* (tap). Under this cluster pointed out targets that talked about a specific event (e.g., competition, laro, round, quarterfinals, and more) or athlete/teams (e.g., Mayweather, Tropang Texters, D' Angelo Russel, Pirates, and more).

Other clusters with high frequency have base verbs of *may*, *sunod* (following), *nasabing* (said), *nakaraan* (previous), *kailangan* (need), *naging* (become), *dating* (before), and more. Positively on one of the clusters grouped prefixes with *pinaka-* (e.g., pinakamalaki, pinakamabigat, pinakamasamang, pinakamarining, pinakamahabang, etc.); although they are not semantically related, it is one with a large, yet distinct set of words with similar strings.

Generally, in target and lexicalization, the approach was not able to produce a lot of sub-clusters, due to nouns in this category involved names (e.g., Cleveland, Pacquiao, Romero, NBA, etc.) and things (e.g., medalya, tiket, record, season, etc.) related to sports. In some that mentioned the same nouns, unfortunately they are on different clusters. Despite this, there are still some examples of sub-clusters which are manlalaro-manlalarong (player/s), player-players, mar-mark (names of players), kampeon-kampeong (champions), and alas-9-alas-2 (time).

In World news, the cluster with highest number of frequency count is the same as most, the verb *may*. As contents concern updates with the entire world, its targets have mixed nouns pertaining to different topics. A few of these indicated that there are *armas* (weapons), MERS-CoV, water, inmates, tanks, cargo, shabu, *pondo* (funds), fiesta, and more. There are also samples of lexicalized entries, which paired apartment-compartment, marking-markings, and dbb-11-db-12.

World category indeed contain a huge number of sentences. The top 10 clusters contain at least 20 counts with the following base verbs: *wala* (nothing or gone), *maging* (become), *nasabing* (said), *matapos* (after), *sinabi* (said), *kinilala* (known), *patay* (dead), *nabatid* (realized), and *dating* (previous). It is noticeable that in this selected dataset, there are a number of accounts that pertain to deaths. Extending the top into 20, aside from *patay*, other verbs related to that crime have the words *nasawi* (casualty), *sinaksak* (stabbed), *nakaratay* (bedridden), and *natagpuan* (found).

For target, since there are a lot of sentences under the category, this is also the category with the most sub-clusters. Instances of these paired biktima-biktimang (victim), imbestigasyon-pag-iimbestiga (investigation), psupt-supt (superintendent), kasong-kaso (case), national-international, infomercial-commercial, and even have samples of semantic relevance (e.g., permit-permiso, pulisya-pulis, suspek-suspect, hospital-ospital, and insidente-akdidente).

Furthermore, there is a particular set of instances that joined targets that relates to time, specifically those with a prefix of *ala-/alas-* followed by a number. It has spread through different clusters, and to visualize it is in this example, alas-900 was paired with alas-915, alas-4, alas-6, and alas-10.

Taking from these results, Dice's coefficient was able to cluster verbs mainly but was ineffective to targets. It is due to a lot of the clusters containing distinct set of nouns that made string distances farther from each other. However, given the chance, it could cluster nouns that has been generally used such as victim, suspect, case, and the likes.

Using Word2Vec

Results of clustering using Word2Vec can be seen at Table 5.18. It produced 1,522 entries in total, about a hundred less than Dice's. The highest frequency under this experiment is 93, under World News. Aside from the highest frequency, World category also gained the highest percentage value in CIC. CEC on the other hand, more clusters were made on the Sports category. Overall, CEC and CIC values are above 15% and 58%, respectively.

Table 5.18: Report Composition of News through Word2Vec Clustering

Category	Total Entries	CEC %	CIC %	Highest Freq.	Total Freq.
Entertainment	306	15.36	58.43	49	623
Opinion	291	16.84	58.28	44	580
Sports	303	22.44	65.24	46	676
World	622	20.42	69.52	93	1624

Analyzing its clusters, it is expected that sentences are joined by their semantic relationship. In the Entertainment category, the highest cluster from the base verb *malaman* (know) produced different ways to communicate or give information. Example of its members include *makikilala* (meet), *makita* (see), *narinig* (hear), *mapapansin* (notice), *sinabi* (told), *hinahanap* (looking for), *importante* (important), *magkasama* (together), and more. It is also important to note that under Entertainment, these are actions mostly done by the subject in the topics.

Forward with targets, there are nouns that were lexicalized that has relatable members. A few of these are about franchise (with actor, character, teleserye, ratings, fans, kumpetisyon, etc.), celebrity names (with Nora, Vilma, Sharon, Herbert, Enrique, Gary, Martin, etc.), places (with Quezon, city, and mall members), and other contextual similarities like *kuwento-buhay* (story-life), *aktres-pelikula* (actress-film), and *kaibigan-kapatid* (friend-sibling). However, even with sub-clusters as such, there are members that were not merged along with them.

Checking on the other highly frequent clusters, they showed verb clusters that showed action in events such as *kinanta* (sang), *dumalo* (visited) and *ipinakilala* (introduced), and features or characteristics such as *maganda* (beautiful) with *matanda* (old), *malaking* (huge) and *mahina* (weak). Other distinguishable sub-clusters combined Daniel-Padilla, Manny-senador, San-Juan, ina-ama, Araneta-coliseum, Muntinlupa-city, Facebook-account, Cebu-Pacific, music-video, acting-career, aktor-aktres, pari-obispong, final-episode, over-acting, emosyonal-problema, and problema-pagsubok-sitwasyon.

In the Opinion category, the highly frequent cluster exhibited verbs that pertains to statements indicating belief. The verbs described here are *sana* (with hope / should be), *kailangan* (need), *sabi* (said), *inisip* (thought), *gusto* (want), *nais* (desire/want), *pwede* (can), and more. Under this cluster, noun sub-clusters contained pairs such as supreme-commission, sampaguita-bakuran, pasahero-sakay, pamahalaan-desisyon, and more. Other members that could be joined include taumbayan-tao, pamahalaan-otoridad, traffic-enforcers, banko-tax, and more.

The rest of the clusters contain a wide distribution of cluster members depending on their base verb. Some examples from the highly frequent ones are the following: *maiwasan* (with huminto, magpadala, mabawasan, mabilis, kumilos, inihanda, ipagpatuloy, etc.), *nawala* (with dumating, nangyari, Nakita, bumalik, nananatili, iniwan, etc.), *nagulat* (with nakatakas, nahuli, pinadala, binaril, nasangkot, nagsalita, malungkot, etc.), and more. While, target sub-clusters' notable examples are pairs such as Aquino-Roxas, Andres-Bonifacio, opisina-trabaho, department-government, camera-picture, pulis-terorista, kaso-criminal, gamot-sintomas, balita-impormasyon, and more.

In Sports, the highest verb cluster is comprised of mostly auxiliary verbs. These are words such as 'am', 'are', 'be', 'do', 'had', 'has', 'have', 'is', 'were', 'was', and the likes. Inside this cluster, target words indicated major sub-clusters relating to the base noun coach (with players, team, championship, league, plan, win, etc.) and time (with tournament, year, and game). Other members include referees, intensity, teams, experience, Lebron, owner, Macau, and more. Provided, it is observable that there are words that could be placed on some of the lexicalized nouns such as teams under coach and time under experience.

Aside from the highly clustered example, other verb clusters were grouped that indicated actions that thematically is related to Sports. Instances for this are the pairs tinapos-tinalo, matapos-sinimulan, napatalsik-nabigo, nakatakda-inaasahan, pinamumunuan-namuno, nakaraang-naunang, maabot-makapasok, and nakuhan-nakamit. Turning to nouns, distinguishable lexicalizations pointed to subjects that are related to each other like Sabado-Linggo, Setyembre-Enero, Malaysia-Singapore, San-Miguel, Manny-Pacquiao, manlalaro-miyembro, laban-labanan, conference-cup, finals-kampeonato, and more.

In the context of World news, the cluster with highest frequency contained keywords of those appropriate for headlines. Take for example using verbs to start or provide a gist of the article: *kinasuhan* (sued), *nasawi* (death), *inaresto* (arrested), *pumatay* (killed), *nagpalabas* (issued), *nilagdaan* (signed), *nasangkot* (involved), *nadiskubre* (discovered), and many more. The topics under this mixed noun about crime (e.g., suspek, pulis, rebelde, and biktima), politics (e.g., president, secretary, and Moreno), and other issues and announcements for the public.

Analyzing the rest, there were distinguishable clusters that provided closely related verbs. Instances involved those about movement (e.g., lumisan, tumakas, iniwan, dinala, umakyat, umalis, humiwalay, dumating, etc.), suggestions (e.g., mangyari, magsagawa, magkaroon, magbibigay, maiwasan, makamit, simulant, siyasatin, etc.), events (e.g., may, mayroong, etc.), and more.

The lexicalization of the targets also expressed the capability of Word2Vec to be able to make use of the vector space in measuring similarities in between nouns. Even on distinct orthographic cases, it still provided relationships such as *mag-asawa* (married couple) with *ina* (mother), *anak* (child), and *lalaki* (male), as members of the family; *eroplano* (airplane) with *sakay* (aboard), *kotse* (car), and *pasahero* (passenger), as subjects about transportation; *sakuna* (catastrophe) with *sunog* (fire), *trapiko* (traffic), *krimen* (crime), and *aksidente* (accident), as mishaps; and more.

These clusters have frequency counts with two-digit values until the 28th entry. In spite having large members count, there still are some entries that could have been included in those clusters. Take the case of *maaresto* (arrested), *naaresto* and *arestado* which could be under a crime related cluster or simplistically clustered with each other, because they are under the same root word. In addition, there are lexicalized instances from Dice's Coefficient that were not found to be close vector distances in Word2Vec. One of the examples is the *ala-/alas-* prefixes, where Word2Vec was not able to lexicalize.

Comparing Word2Vec with Dice's coefficient, it has been proven that both were able to cluster verbs and nouns. However, Word2Vec produced more members for both fields. Word2Vec merged words based on its usage and there are an ample number of examples that exhibited semantic similarity. Extending this, it also covers some clusters with orthographic similarity. One flaw in this approach is there are some words that were not captured in the cluster, as its vector distance is far from each other even with obvious semantic similarities – for example are proper names.

Using FastText

Applying FastText, this approach produced 547 total entries as shown at Table 5.19, the lowest among the three. A smaller number of entries, in this case, mean that there have been more sentences clustered together. As a matter of fact, the highest frequency among all categories is 234. Supporting this, CEC and CIC values are above 60% and 91%, respectively – an amount generally higher than the other approaches. Specifically for CEC, the highest value garnered was

70.52%; whereas, in CIC, 96.86% of the sentences were clustered. Both of these values are under the World category. It signifies that under this approach, it is appropriate to combine Metro, National, and Province news together to produce a larger quantity in the category that substantiates similar properties in reporting news under this consolidated category; thus, being able to cluster more sentences.

Table 5.19: Report Composition of News through Word2Vec Clustering

Category	Total Entries	CEC %	CIC %	Highest Freq.	Total Freq.
Entertainment	120	63.33	92.94	47	623
Opinion	130	60.00	91.03	32	580
Sports	124	62.10	93.05	72	676
World	173	70.52	96.86	234	1624

Observing cluster contents in Entertainment, its highly frequent verb cluster collected words that indicates events happening within the category. It specified that targets are *ine-endorse*, *nakumbinse* (convinced), *natsismis* (gossip), *naka-schedule*, *nag-cancel*, *nag-extend*, and more. Basing on a few of these verbs, most pertains to status of the shows or film. In terms of its targets, words collected are movie, commercial, episode, shooting, press, *produksyon* (production), *programa* (program), *eksena* (scene), and more.

In lexicalization, there were a few sub-clusters that showed clear relationships such as life-people, movie-shooting, entertainment-music, final-season, and more. Even with these, there are members in the cluster that was not lexicalized such as eksena-scene, pamilya-anak, skin-collagen, episode-season, and more. Furthermore, there were inexplicit relationships in the sub-clusters such as collagen-bet, fans-go, leaf-day, na-taong, and credit-wansapanataym.

The following clusters within Entertainment, provided status (e.g., naging, nannatiling, naging, nagkaron, etc.), characteristics (e.g., mabait, matangkad, mahirap, kalmado, masama, etc.), and incidents (e.g., nag-viral, siniraan, lumaylay, nagkukuwento, etc.). There were also instances that mixed the clustered verbs with orthographic similarity such as *makikita* (see), *nakikita*, *pinapakita*, and *nagkita*. Adding semantic relationships, under this cluster included the words *ipinapanuod* (watch) and *mapapansin* (notice).

Other positive instances for nouns paired the following: pelikula-aktor, ratings-happy, emosyonal-damdamin, GMA-bulaga, century-year, buwang-buwan, Smart-Araneta, Muntinlupa-city, reserved-seat, film-festival, Facebook-account, fresh-fruits, direktor-aktres and more. Negative instances on the other hand, in the

sense that the relationship is not obvious are the following: crush-cubao, march-scratches, Instagram-tirik, pagkakasakit-pagtaas, issue-welcome, and more.

Under Opinion, top cluster touched ideas that explicitly showed subjects to possess (i.e., may and mayroong). Target words under this are camera, *testigo* (witness), rally, *reklamo* (complaint), *pagkain* (food), *himig* (song), and more. Distinguishable targets showed relationship between internet-email, lungga-hidden, and lalaki-sanggol.

Following this, more verbs clustered contained mixed and diverse set of action words. sinabi-nagsalita, kailangan-mahalaga, nawala-nagkaroon, and ginagawanganayayari. There was even a cluster that contained a large number of verbs that starting with the word *nagbibiruan* (joke), and clustered together *nagtatanong* (ask), *nagtaksi* (took a taxi), *nagbigay* (give), *pagbibitbit* (carry), *nagpaalam* (said goodbye), *naghudad* (undress), and more.

For nouns, positive instances consist of the following sample pairs: opisina-gobyerno, manpower-highways, taumbayan-pamahalaan, sakay-sasakyan, MRT-LRT, kapulisan-terorista, airport-seaport, kasong-kriminal, highway-patrol, toxic-hazardous, and more. On the other hand, negative examples are luha-paradahan, pagpapawis-prosecutors, pananalita-mamasapano, bumuwelocorridor, and many more.

Under Sports category, its cluster with the highest frequency produced groups with indistinguishable relationship. From the base forged, some of the cluster members are prepared, *ipinoste* (posted), ranking, evaluating, deliver, breaking, uniting, fostering, mens, Benilde, and other auxiliary verbs such as ‘do’, ‘were’, ‘is’, ‘had’, and ‘been’. Mixing verbs with nouns may be caused by inaccuracy in the Part-of-Speech tagger module, however adding auxiliary verbs made this cluster hazier than it already is. It is also surprising how the words ‘forged’ and ‘mens’ or ‘Benilde’ are close in distance in the vector space to be clustered together.

Although there is a variety on its verb clustering, contents in target combined ideas resulted into having large lexicalized sub-clusters, wherein from the word ‘challenge’ it was able to come up with 45 members. Sample members for this sub-cluster are time, message, experiences, referees, barriers, intensity, tournament, mistake, adjustment, tiredness, years, and more.

Subsequent verb clusters also provided similar diversity in its clusters but were more specific as compared to the top cluster. There were clusters that focused on the progression (e.g., matapos, makaraan, kasunod, etc.), responses (e.g., sinabi, nangako, inihayag, sisiguraduhin, etc.), and status (e.g., umusad, makabangon, inilunsad, nabigo, etc.) of the sport events.

Positive instances of target sub-clusters showed the following relationships: laro-game, points-round, lider-chairman, national-pambansang, players-koponan, kabiguan-pagkatalo, siko-braso, Olympics-medalya, Malaysia-Singapore, MVP-famous, Mavericks-Heat, Setyembre-Enero, point-guard, and many more. On the other hand, negative examples paired offensive-no, plan-tickets, duda-swerte, fans-contract, fight-president, and more.

The category with the highest frequency amongst all is World category. As compared to the other approaches FastText was able to cluster 234 ideas. Under this cluster contain actions relating to World events; sample of these are *nasuspinde* (suspended), *abandonadong* (abandoned), *patay* (dead), *isinailalim* (undergone), *sinunog* (burned), *madugong* (bloody), *kinumpiska* (confiscated), *pagkakabangga* (crash), *kuwestiyonable* (questionable), *nag-organisa* (organized), *inoobserbahan* (observed), *nasaksihan* (witnessed), and many more.

Provided with that set of verbs, nouns present on those topics involved the following: *bomba* (bomb), *biktim* (victim), *pagamutan* (hospital), *bangkay* (corpse), *suspek* (suspect), *lindol* (earthquake), compound, city, and many more. By counting the number of sub-clusters, under this cluster alone have 39. Largest sub-cluster have 62 members, merging commissioner, requirement, surveillance, cam, murder, bail, proceedings, interrogation, arson, SUV, telephone, rollback, and other nouns. Notable lexicalized pairs joined bahay-silid, suspek-kasong, anak-ama, kaalaman-impormasyon, operation-operatiba, ulo-katawan, hospital-medical, batay-resulta, oras-minuto and more.

Succeeding the top cluster, highly frequent clusters have verb groups that concerns with bumaba ‘drop’ (with nahulog, nakaratay, bumulagta, nagpabagsak, tumaas, etc.), iniulat ‘report’ (with kinasuhan, inaresto, napatay, kinumpirma, inalam, etc.), iniwān ‘left’ (with napansin, naalarma, nakilala, pinasok, binalikan, etc.), and matapos ‘after’ (with nasabing, sumunod, nakalipas, pagkatapos, etc.) base verbs. Observing the rest of the clusters’ lexicalized nouns, some examples that displayed clear relationships are provided; recognized pairs are the following: milyon-million, river-ilog, Filipino-Pilipino, bansa-Pilipinas, suspek-biktim, awtopsiya-inspeksyon, kamay-paa, pangulo-pinuno, gunman-shot, pagulan-tagtuyot, pera-alahas, and more.

Additionally, comparing to Word2Vec, FastText was able to group *ala-/alas-* prefixes, which evidently showed the effectiveness of using n-gram vectors. Even though n-gram vectors prove to be effective on word variants, there are clusters wherein words are not exactly related to each other. Take the case of residente-residue, where it indeed showed orthographic similarity, but its meaning did not match up.

Examining the approach entirely, the most compelling quality of FastText is its ability to be able to capture both orthographic and semantic similarities. Having said, positive instances of Dice's Coefficient and Word2Vec approaches was covered or present. Although this is the case, their negative qualities are also passed down, where it was unable to combine closely related terminologies and instances that have uncertain relationship with each other were still clustered. Regardless, in this experiment, it has been shown the extent of relationships that FastText was able to cover.

Moreover, with this coverage, there were large quantity of sentences that were clustered, that positively designated more sentences in clusters, but negatively was harder to interpret for its vague relationships. Other than that, it is important to remember its capability to be able to group words through a combination of vector space and n-grams which produced an approach that complements previous approaches.

Chapter 6

Conclusion and Recommendations

In this paper, the development of an automated insight extraction and organization was discussed. It collected ideas or actionable points from textual data, specifically Malasakit community responses. It is then presented in a list report, ordered by highly suggested ideas, that serves as medium that can connect local communities with respective decision makers in disaster planning and response.

Results showed that the use of Part-of-Speech Information Extraction achieved satisfactory marks in collecting these insights. It was measured through standard metrics such as Precision (P), Recall (R), Accuracy (A), and F-Measure (F). In extracting insight phrases, scores were recorded with $P = 70.09$, $R = 80.87$, $A = 72.57$, and $F = 75.09$. Whereas on word sets, scores achieved $P = 62.14$, $R = 74.06$, $A = 51.03$, and $F = 67.58$.

There are two options available in organizing the extracted insights, where one can organize all entries or organize by response categories. Under these formatting, three clustering approaches were applied, namely Dice's Coefficient, Word2Vec, and FastText. Undoubtedly, all approaches were able to join similar ideas together and some instances formed one, interpretable idea. At the same time, the three approaches still have room for improvement, especially in capturing the right balance of relationships between words. Even though each have their own advantages and disadvantages, in one of the assessments, organizing by response categories and using Dice's Coefficient as clustering approach was the preferred combination in presenting the results.

Experimenting on the modules, lexicalization, normalization, and threshold adjustments were implemented. Results showed the effectivity of placing a word that represents a cluster, and that there could be other suitable representatives in the group. In normalization, changes were evident in words and through shifts in Part-of-Speech tags. On threshold adjustment of clustering approaches, primary learning is it controls the scope and composition of a cluster's contents. Furthermore, another experiment changed the domain of the data. The whole process was tested on a News dataset, where it was able to extract and organize ideas under topics of Entertainment, Opinion, Sports, and World-related articles.

In assessing the entirety of the tool and its generated report, a survey was conducted. The tool was rated 4.07 out of five and the report got 4.25 out of five. The tool excelled in usefulness and satisfaction, while the report on appropriateness, impact, presentation, usefulness, and satisfaction. Areas for improvement with regards to the tool is on its ease of use and ease of learning. Whereas for the report, content readability and interpretability need work. Nonetheless, participants appreciated the benefit and potential of having these two outputs and provided feedback on possible improvements.

Specifying the contributions of this research, in terms of using the reports, the Excel file contain extensive details about the responses such as the response itself and its category, and output of the architectures modules such as normalized responses, Part-of-Speech tags, extracted insight phrases and word sets, and clustered unranked and ranked insights. It could serve as a data source for future disaster-related researches and its simplified Word document or HTML report could be used to assist decision makers in disaster planning and response.

In terms of the API, as it is open-source and online distributed, it could be used to develop and supplement other Data Processing, Text Analysis, or Natural Language Processing tasks. Since it is also open-sourced, the API can be extended and developed further for optimizations and addition of features. In the sense, this serves as the start of collecting Filipino Natural Language Processing tools. Furthermore, this document, as well as an abridged version and supporting documents involved in this study are available in public to maximize its potential of being extended.

Future directions of this study involve improving the heuristics of the extractor, exploring more novel clustering approaches, adding Graphical User Interface in the API, including other report formats such as infographics or visualizations, commercializing the tool, integrating the tool on other software applications, and studying the effects of applying the report in disaster planning and response.

Appendix A

Filipino Part-of-Speech (POS) Tagset

MGNIN TAGSET

Total of 230 Tags (69 Basic and 161 Compound)
Updated as of March 1, 2017

Part of Speech	Tag Name	Example
Noun (Pangngalan)	NN (4)	
Common Noun (count noun)	NNC	papel, tao, pag- + verb, kapalit
Proper Noun	NNP	Pilipino, Lasaliano
Proper Noun Abbreviation	NNPA	Dra., Bb., G., Kgg., etc.
Common Noun Abbreviation	NNCA	in, km, m, cm, measurements, et al, etc.
Pronoun (Panghalip)	PR (10)	
as Subject (Palagyo)/Personal Pronouns Singular	PRS	ako, ikaw, ka, siya, ko, mo, niya, kita, nya
Personal Pronouns (Plural)	PRP	kami, tayo, kayo, sila, nila, naming, natin, ninyo
Possessive Subject (Paari)	PRSP	akin, iyo, kanya, amin, atin, inyo, kanila
Pointing to an Object		
Demonstrative/(Paturrol/Pamatlig)	PRO	ito, iyan, iyon, iri/e, niyan, niyon/noon, nito, naroon, nariyan, yaon
Question/Interrogative (Panamong)/Singular	PRQ	sino, saan, alin, ilan, gaano, kanino, magkano
Question/Interrogative Plural	PRQP	sinu-sino, saan-saan, alin-alin, ilan-ilan, gaa-gaano, kani-kanino, etc.
Location (Panlunan)	PRL	dito, doon, diyan, riyan, roon, rito, nandito, etc.
Comparison (Panulad)	PRC	ganyan, ganito
Found (Pahimaton)	PRF	ayto, heto, hayan, ayon, yun, hayun
Indefinite	PRI	kuwan, iba, kapwa, isa, lahat, marami, kaunti, sinuman, alinman, anuman, kalahatan, kabuuan
Determiner (Pantukoy)	DT (4)	
for Common Noun	DTC	ang
for Common Noun Plural	DTCP	(ang) mga, (ng) mga
for Proper Noun	DTP	si, ni, kay
for Proper Noun Plural	DTPP	sina, nina, kina
Lexical Marker	LM (1)	ay
Conjunctions (Pang-ugnay)	CC (6)	
	CCT	o, saka, ni-, maging, pero, subalit, ngunit, bagkus, kundi, imbes, kahit, halip, maliban, sa, sa pamamagitan ng, bilang, bagamat, datapwat, samantala, habang, para
	CCR	kaya (tuloy), kaya (nga ba), kaya (ngayon), kasi, dahil sa, dahil kasi, kung dangan kasi, papaano kasi, sapagkat, kasi, dahilan sa, palibhasa
	CCB	at saka, at gayon din, at...rin, kasama, upang, ng, nang gayundin, palibhasa, sa sandaling, basta't
	CCA	at, pati
Ligatures (Pang-angkop)	CCP	na, -ng, -g
Preposition (Pang-ukol)	CCU	laban sa, dagdag pa
Verb (Pandiwa)	VB (14)	
Neutral/Infinitive	VBW	mag-, ma-, mang-, sana, sabi, ka- + verb, mapag- + verb, makipag- + verb, maging
Auxiliary, Modal/Pseudo-verbs	VBS	kailangan, pwede, dapat, maari, gusto, ayaw, ibig, nais

Existential	VBH	mayroon, meron, may
Non-existential	VBN	wala, ala
Time Past (Perfective)	VBTS	nahulog, kumain, pinaalis, nag-, naging
Time Present (Imperfective)	VBTR	nahuhulog, kumakain, pinapaalis, nagiging
Time Future (Contemplative)	VBTF	mahuhulog, kakain, papaalisin, magiging
Recent past	VBTP	kahuhulog, kakakain, kapapaalis
Actor Focus	VBAF	-um-, mag-, ma-, mang-
Object/Goal Focus	VBOF	-in-, -an, i-
Benefactive Focus	VBOB	i-, ipag-
Locative Focus	VBOL	-an, -in, pag...an
Instrumental Focus	VBOI	ipang-
Referential/Measurement Focus	VBRF	pinag-
Adjective (Pang-uri)	JJ (6)	
Describing (Panlarawan)	JJD	maganda, mabait, buo, masyado, bawat
Used for Comparison (same level) (Pahambing Magkatulad)	JJC	sing-, kasing, kapwa, pareho, magsing, magkasing, gangga, ga, tulad ng, gaya ng, kaysa sa
Comparison Comparative (more) (Palamang)	JJCC	mas, medyo, higit, lalo, lalong
Comparison Superlative (most) (Pasukdol)	JJCS	pinaka-, ubod, sakdal, ulo, labis, hari
Comparison Negation (not quite) (Di-Magkatulad)	JJCN	di-gasinong, di-gaano
Describing Number (Pamilang)	JJN	tatlong, labinlima
Adverb (Pang-Abay)	RB (15)	
Describing "How" (Pamaraan)	RBD	mabilis na tumakbo, masayang umuwi, pa + verb, sabay, naka- + verb
Number (Pangaano/Panukat)	RBN	nang limang libra, + apat na guhit
Conditional (Kondisyunal)	RBK	kung, sakali, pagka, kapag, pag
Causative (Panahi)	RPB	dahil sa, dahil dito, kaya
Benefactive (Benepaktibo)	RBB	para sa, para kay
Referential (Pangkaukulang)	RBR	tungkol sa, ukol, hinggil, patungkol, ayon sa, ukol sa, hinggil sa, alinsunod sa, sabi ni, wika ni, tanong ni
Question (Pananong)	RBQ	bakit, paano, baga, kaya, gaano
Agree (Panang-ayon)	RBT	talaga, oo, tunay, mangyari, opo, oho, siyang pala, sadya, maaaring, totoo
Disagree (Pananggi)	RBF	hindi nga, hindin-hindi, walang, huwag, ewan, aywan, ayaw, malay, wag, ayoko
Frequency (Pamanahon)	RBW	tuwing, muli, ngayon, laging, pagkatapos, noon, mamaya, parati, bihira, bago, uli, sandali, minsan, samantala, habang, kapag, buhat, mula ng, umpsisa, hanggang, kahapon, kanina, bukas, araw-araw, galing
Possibility (Pang-agam)	RBM	baka, tila, marahil, yata, siguro, wari, malamang, maaaring
Place (Panlunan)	RBL	kina Thelma, nasa, sa + bahay, amin, ilalim, likod, itaas, harap, mula sa, kinaroroongan, tungo sa
Enclitics (Paningit)	RBI	na, pa, rin, din, man, muna, kaya, naman, sana, yata, ba, nga, daw, raw, kasi, lang, lamang, pala, tuloy
Interjections (Sambitila)	RBj	hoy, aba, ay, aray, naku, ha
Social Formula (Formularyong)	RBS	Tao po!, Magandang umaga! Mano po., Salamat po.,

Panlipunan)		Pasensya na po., Sori po.
Cardinal Number (Bilang)	CD (1)	
Digit, Rank, Count	CDB	I, una, tatlo, II
Topicless (Walang Paksa)	TS (1)	Umuulan., Alas dos na., May tao., Ang tapang mo pala.
Foreign Words	FW (1)	English, Spanish, Latin
Punctuation (Pananda)	PM (6)	
Period	PMP	"."
Exclamation Point	PME	"!"
Question Mark	PMQ	"?"
Comma	PMC	","
Semi-colon	PMSC	";"
Symbols	PMS	"@, /, +, *, (,), ‘, ~, &, %, \$, #, =, -, :"
Compound Tags	<tag1> <tag2> ... <tagN>	
CCB_CCP	JJC_VBTR_CCP	PRI_CCT
CCR_CCA	JJC_VBTR	PRI_LM
CCR_CCB	JJC_VBTR_VBOF	PRL_CCP
CCR_CCP	JJC_VBTR_VBRF	PRL_LM
CCR_LM	JJC_VBTS	PRO_CCB
CCT_CCA	JJC_VBW	PRO_CCP
CCT_CCP	JJC_CCB	PRO_LM
CCT_LM	JJC_CCP	PRP_CCB
CCU_DTP	JJC_JID	PRP_CCP
CDB_CCA	JJC_PRL	PRP_LM
CDB_CCP	JJD_CCA	PRQ_CCP
CDB_LM	JJD_CCB	PRQ_LM
CDB_NNC	JJD_CCP	PRSP_CCP
CDB_NNC_CCP	JJD_CCT	PRS_CCB
JJCC_CCP	JJD_NNC	PRS_CCP
JJCC_JID	JJD_NNP	PRS_LM
JJCN_CCP	JJN_CCA	RBD_CCB
JJCN_LM	JJN_CCB	RBD_CCP
JJCS_CCB	JJN_CCP	RBD_LM
JJCS_CCP	JJN_NNC	RBF_CCP
JJCS_JJC	JJN_NNC_CCP	RBF_JID
JJCS_JJC_CCP	JNC_CCA	RBF_JID_CCP
JJCS_JJD	JNC_CCB	RBF_LM
JJCS_JJD_CCB	JNC_CCP	RBF_RBW
JJCS_JJD_CCP	JNC_LM	RBF_VBTR
JJCS_JJD_NNC	JNC_PMC	RBF_VBW_CCP
JJCS_JJN	NNP_CCA	RBI_CCA
JJCS_JJN_CCP	NNP_CCP	RBI_CCP
JJCS_RBF	PRC_CCB	RBI_LM
JJCS_VBAF	PRC_CCP	RBJ_CCP
JJCS_VBAF_CCP	PRI_CCB	RBK_LM
JJCS_VBN_CCP	PRI_CCP	RBL_CCP
JJCS_VBOF	PRI_CCP_NNP	RBL_CCP_NNP
		VBRF_CCP

Tagset used at Nocon, N. and Borra, A.'s "*SMTPOST: Using Statistical Machine Translation Approach in Filipino Part-of-Speech Tagging*" (2016) from De La Salle University, Manila, Philippines.
(<https://www.aclweb.org/anthology/Y/Y16/Y16-3010.pdf>)

Appendix B

Application Programming Interface (API) Functions

Table B.1: Data Utilities Module

Function Name	Description
refresh_excel	Clears out values in excel, excluding values under a protected_cell variable.
read_candidate_excel	Reads the values in the candidate's excel file and stores the value in a list.
read_gold_standard_excel	Reads the values in the gold standard's excel file and stores the value in a list.
read_clusters_excel	Reads the values in an excel file and stores the ranked cluster values in a list.
read_excel	Reads the values in an excel file and stores the first two columns (response and tag) into the MalasakitResponse object.
write_excel	Writes the system output in an excel file.
text_to_list	Transforms a given text file into a list of list [s1 [w1, w2, ..., wN], ..., sN [w1, w2, ..., wN]].
text_to_list_without_stopwords	Transforms a given text file into a list of lists removing Tagalog/English stopwords in the process.
write_text_file	Writes the strings in a list to a text file.
read_text_file	Reads the strings in a file and transfer them into a list.
get_stopwords_from_file	Transforms stop words in a given file to a list.

Table B.2: Normalization Module

Function Name	Description
normalize_object	Normalizes the MalasakitResponse object and updates the object after.
normalize_list	Normalizes a given list of strings and returns the normalized list.
normalize_string	Normalizes a given string and returns the normalized string.
join_prefix_word	Joins the prefixes that are separated with a word by whitespace. More prefixes can be added in the tl_prefixes.txt file.
translate_filipino_colloquialism	Runs the Filipino Colloquialism Translator or Normalizer through command prompt. File path parameters can be changed by the user.
set_moses_file_path	Sets Moses' executable file path with the one provided by the user.
set_model_file_path	Sets Moses' model configuration file path with the one provided by the user.
set_input_file_path	Sets Moses' input text file path with the one provided by the user.
set_output_file_path	Sets Moses' output text file path with the one provided by the user.
set_prefix_file_path	Sets a user-provided prefix list text file.

Table B.3: Language Identification Module

Function Name	Description
set_language	Changes the scope of languages.
identify_language_string	Identifies language (ISO 639-1 code) of a given string. Returns a (language, confidence) tuple.
identify_language_object_list	Identifies language between Tagalog or English of MalasakitResponse object's responses and updates the language field in the object.
identify_language_string_list	Identifies language of a sentence list. Returns a list of languages respective to the sentences.

Table B.4: Filipino Part-of-Speech Tagger Module

Function Name	Description
set_java_path	Sets the java path to make Stanford POS Tagger work.
tag_string	Tags a sentence/string. Returns a (word, pos) tuple.
tag_object_list	tagging a list of MalasakitResponse object's sentence. This updates the MalasakitResponse object.
tag_string_list	Tags a sentence list. Returns a list of (word, pos) tuple.
format_pos	Formats a tuple into a POS-only string.
format_stanford	Formats a tuple into Stanford word tag string.

Table B.5: Information Extraction Module

Function Name	Description
extract_insights_phrases	Extracts phrase insights (action word to target/s). The MalasakitResponse object is updated after.
extract_insights_words	Extracts word insights or word sets (action word and target/s). The MalasakitResponse object is updated after.

Table B.6: Information Organization Module

Function Name	Description
organize_sublist	Organizes a given sublist (a single category or the current category).
organize_by_response_categories	Organizes the information based on (or per) response categories.
organize_all_entries	Organizes the entire information (per entry).

Table B.7: Information Clustering Module

Function Name	Description
string_similarity_fasttext	Computes FastText's vector similarity (high value is better) between two strings.
string_distance_fasttext	Computes FastText's vector distance (low value is better) between two strings.
string_similarity_word2vec	Computes Word2Vec's vector similarity (high value is better) between two strings.
string_distance_word2vec	Computes Word2Vec's vector distance (low value is better) between two strings.

Table B.7 continued from previous page

Function Name	Description
string_similarity_dice	Computes Dice's Coefficient similarity (high value is better) between two strings.
string_distance_dice	Computes Dice's Coefficient distance (low value is better) between two strings.
collect_all_insights_from_object	Retrieves all insights in MalasakitResponse object and stores them in one list.
merge_cluster_insights	Merges the insights in one cluster into a single line.
remove_duplicate	Removes duplicate strings in a list (cluster).
cluster_words	Clusters target/noun words. Given a list it will join similar words using the 'word_1 (word_2, ..., word_N)' notation.
cluster_information	Clusters text using either Srensen-Dice Coefficient (String Clustering), Word2Vec, or FastText Word Embeddings (Semantic Clustering) and returns a list of clusters.

Table B.8: Information Ranking Module

Function Name	Description
rank_clusters_by_frequency	Ranks the clusters based on their frequency counts (descending order: highest count first).
rank_by_response_categories	Ranks the clusters (rearranges the groups) based on their response category prioritization order. Categories follow the Malasakit Codebook 4.7.

Table B.9: Report Generation Module

Function Name	Description
add_timestamp	Adds a timestamp in the document. Follows Month-Date-Year Hours-Minutes-Seconds format (e.g., Jan-01-2020 23:59:59).
add_divider	Adds a divider in the document that is made from a 1x1 table object.
add_title	Adds the title in the document and formats it for display.
set_document_margin	Sets the document's margin (in inches).
set_number_of_page_columns	Sets the number of columns in a section through xpath.

Table B.9 continued from previous page

Function Name	Description
write_report	Generates the report in word document (.docx) format. Default filename: Report.docx
generate_web	Generates the report in web/HTML format. Target words are linked to their respective Malasakit responses.

Appendix C

Report Screenshots

FILIPINO TEXT ANALYSIS TOOL REPORT

Mar-24-2020 20:24:54

The information below were extracted and organized automatically.

INFORMATION CAMPAIGN AND CAPACITY BUILDING

Entry 1

ID/s:

26|208|209|339|507|576|580|649|762|766|
811|813|836|852|854|885

Frequency: 20

Proposed action: dapat

Target: barangay, seminar, oras, prepared, kalamidad, sakuna, posters, paalala, drill, like, for, example, about, paraan, encourage, before, during, and, after, of, the, calamity, my, weekly, mg, roong

Entry 2

ID/s:

13|46|208|236|243|307|311|598|603|618|6
31|651|758|760|766|854|903|925

Frequency: 19

Proposed action: maging, palaging, laging

Target: kapitbahay, tao, disaster, mamamayan, aware, kabbarangay, kabaro, kalamidad, daring, beforeafter, and, during, pra, bagyo, trahedy, gawim, paalala

Entry 3

ID/s:

33|68|209|289|307|310|311|339|696|716|7
62|787|800|847|885|926|930

Frequency: 19

Proposed action: magkaroon, magkaron, magkakaroon, karoon

Target: seminars (seminar, seminarsdrill), about, kaalaman (kaalam), disaster, drill, sos, program, regarding, and, iinvite, emergency, kits, training, progma, pra, disciplina, mamamayan, weekly, meeting, organisasyon, orientation, barangay

Entry 4

ID/s:

80|121|132|149|156|169|215|222|234|340|
585|654|794|831|929

Frequency: 15

Proposed action: be

Target: advantage, possibilities, calamities, dont, times (time), drills, typhoon, disaster, preparedness (prepare), programs, officials, instructions, orientation, effects, duty

Entry 5

ID/s:

130|215|245|249|259|266|340|390|794|838

Frequency: 12

Proposed action: have

Target: assembly, disaster, drill, place, representative, check, emergency, needs, training (taraining), barangay, officials, seminar, meeting

Entry 6

ID/s:

169|222|303|331|343|358|385|614|627|654
|661

Frequency: 11

Proposed action: help

Target: prepare, families, disasters (disaster), outcomes, community, idea, seminars, people, signage, calamity

Entry 7

ID/s:

101|104|169|287|333|343|347|396|583|644

Frequency: 10

Proposed action: preparing, prepare

Target: typhoon, disaster, seminar, times

Entry 8

Extracting and Organizing Disaster-related Philippine Community Responses for Aiding Nationwide Risk Reduction Planning and Response (N. Nocon, 2020)

Figure C.1: Word Report: First Page (By Categories)

FILIPINO TEXT ANALYSIS TOOL REPORT

Mar-04-2020 19:40:07

The information below were extracted and organized automatically.

Entry 1

ID/s:

26|28|31|39|58|69|73|90|148|151|167|203|
204|208|209|211|306|339|341|342|366|436|
|440|448|455|495|506|507|510|529|554|57|
6|580|612|629|643|649|673|732|743|757|7|
62|766|767|777|791|793|811|812|813|821|
836|850|851|852|853|854|885|888|890|891|
|892|894|896|899|900|901|902|922|923|

Frequency: 80

Proposed action: dapat, sapat

Target: barangay (barangay, baranggay), active, pakiki-isa, komunidad, sakuna, eg, komunikasyon, kita, kalinisan, mamamayan, disaster, balita, evacuation, center, aware, pamamagitan, seminar, oras, anunsyo, prepared, updated (update), kanal, pangongolekta, lugar, alert, tao, araw, kalamidad, alirto, radio, abiso, saknila, posters, paalala, official, ugnayan, membro, i, lage, panahon, drill, like, for, example, about, paraan, truck, encourage, pagbabahahindi, dn, before, during, and, after, of, the, calamity, my, weekly, paligid, mg, roong, mgkaisa, ulat, darating, pra, programa (program), pagkain, pangangailangan

Entry 2

ID/s:

6|7|13|16|19|41|46|122|128|208|236|243|3|
07|311|341|365|419|439|443|447|457|471|
502|513|557|598|603|611|616|618|621|626|
|631|651|660|690|714|721|728|731|747|74|
8|753|758|760|766|783|785|786|806|815|8|
17|820|823|824|830|833|854|889|903|925|

Frequency: 63

Proposed action: maging, palaging, messaging, laging, pagiging, magiging, nagiging

Target: tao, kapitbahay, bagay, bagyobaha (bagyo), darating (daratinga), disaster,

mamamayan, aware, oras, sakuna, paglaki, pra, kalagayan, alertoat, barangay (barangay), tagapag, handa, pamaraan, kalamidad (kalamid), agrisibo, kabaro, atentibo, balita, cla, brgys, pkiking, beforeafter, and, during, kalikasan, kagamatin, opisyal, sanhi, kasanyan, xa, trahedy, media, etc, gawim, paalala, prepared, dahilan

Entry 3

ID/s:

80|112|121|123|125|132|137|139|140|145|
146|149|153|154|156|165|169|175|192|195|
|198|199|215|222|233|234|235|248|285|29|
2|308|309|318|340|371|392|575|585|619|6|
54|755|794|795|816|831|859|861|863|868|
909|928|929|932|934|

Frequency: 56

Proposed action: be

Target: advantage, training, possibilities, news, plenty, calamities, community, beforehand, incoming, disaster (disasters), officials, time (times), dont, publicsoundnotifsystem, happenings, typhoon, aware, garbage, evacuation, plan, drills, things, preparedness (prepare), programs, constituents, flood, devices, instructions, government, technology, need, orientation, effects, management, response, unit, barangay, events, posters, case, duty, info, society

Entry 4

ID/s:

3|6|7|8|11|14|20|27|33|35|40|44|50|66|68|
209|289|307|310|311|321|339|341|368|467|
|478|488|550|555|696|716|752|762|779|78|
7|800|847|885|926|930|

Frequency: 42

Extracting and Organizing Disaster-related Philippine Community Responses for Aiding Nationwide Risk Reduction Planning and Response (N. Nocon, 2020)

Figure C.2: Word Report: First Page (All Entries)

MALASAKIT RESPONSES REFERENCE LIST

I Response	D
1 magkaisa dapat ang mga tao	7 bawat isa sa kanilang tungkulin
2 mag karoon ng pagkakaiisa upang sa mga darating na mga sakuna ay malalagpasan	1 bawat pamilya ay dapat nagkakaisa
3 magkaroon ng komunikasyon kung saan magkikita sa panahon ng calamidad	8
4 paglilinis ng kanal wastong pagtatapon ng basura at kailangan mag ikot ikot ang mga tanod upang bantayan mga gamit ng tao	1 maging handa sa ano mang dumating na bagyoberba sa komunidad sa ating bansa
5 malawakang information drive	2 magkaroon ng komunikasyon at ikutin ng
6 bago dumating ang bagyo magkaroon ng early warning system para mas maging handa ang mga tao	0 mga council members para makita ang sitwasyon ng barangay sa tuwing may calamidad
7 magkaroon ng early warning upang maging handa ang mga tao sa darating na bagyo	2 tumulong sa mga karatig bayah at
8 higit na pagtibayin ang early warning system device magkaroon ng maintenance quarterly para masigurong maayos ito bago dumating ang isang calamidad	1 magbigay ng naitabing pagkain sa mga napinsala ng bagyo
9 lalo pang lumawat at lumago ang pagmamalasakit sa aming ka-barangay	2 maayos at malinaw na pagpaplano bilang paghahanda sa calamidad
1 pagsunod sa sinasabi sa kung ano ang	2 maagang paghahanda bago dumating ang
0 dapat gawin paghandaan ang lahat ng bibitbitin sa tuwing may sakuna	3 sakuna
1 pagkakaroon ng early warning device	2 maging alerto at maging handa
1	4
1 pagbibigay ng humanitarian assistance	2 magdasal
2 goods sa panahon ng calamidad	5
1 nais ko po sana magkaroon pa po ngmga	2 dapat ang barangay ay sanayin ang mga
3 ibat ibang paraan upang lalo pang maging handa ang aming mga kapitbahay po	6 naninirahan sa kanilang lugar tungkol sa drrm na may sapat na pangangailangan sa pagsapit ng bagyoberba
1 siguruhing magkaroon ng mga basurahan	2 magkaroon ng maayos na komunikasyon
4 sa buong barangay dahil ito ang pangunahing sanhi ng calamidad	7 gaya ng paggamit ng public address system para mabilis ang pagpaparating ng inormasyon
1 seminar tungkol sa bagyo at lindol para	2 mabuti ang aming barangay at
5 mapaghandaan	8 nakatutulong sa amin pag may problema dapat mas maging active pa ang ibang opisyal ng barangay kung maari 100 performance na maipaabot ang serbisyo sa mga tao
1 maging alerto lagi sa mga di inaasahang	2 bahay-bahay na pag-inform tungkol sa
6 bagay magtulong tulong opisyal man o hindi ng brgy	9 sakuna
1 magkaisa ang lahat maging totoo ang	3 mabilis na paginform ng chairman upang

Extracting and Organizing Disaster-related Philippine Community Responses for Aiding Nationwide Risk Reduction Planning and Response (N. Nocon, 2020)

Figure C.3: Word Report: Malasakit Responses List

A	B	C	D
1 magkaisa dapat ang mga tao	Filipino values	tl magkaisa dapat ang mga tao	
2 magkaroong ng pagkakaisa upang sa mga Filipinio values		tl magkaroong ng pagkakaisa upang sa mga darating na mga sakuna ay malalagpasan	
3 magkaroong ng komunikasyon kung saan mearly warning system		tl magkaroong ng komunikasyon kung saan magkikita sa panahon ng kalamidad	
4 paglilinis ng kanal wastong pagtatapon ng infrastructure maintenance and management		tl paglilinis ng kanal wastong pagtatapon ng basura at kailangan magikot ikot ang mga tanod upang bantayan	
5 malawakang information drive	information campaign and capacity building	tl malawakang information drive	
6 bago dumating ang bagyo magkaroong ng early warning system		tl bago dumating ang bagyo magkaroong ng early warning system para mas maging handa ang mga tao	
7 magkaroong ng early warning upang maginj ng early warning system		tl magkaroong ng early warning upang maging handa ang mga tao sa darating na bagyo	
8 higit na pagtibayin ang early warning systeearly warning system		tl higit na pagtibayin ang early warning system device magkaroong ng maintenance quarterly para masigurong	
9 lalo pang lumawat at lumago ang pagmam Filipino values		tl lalo pang lumawat at lumago ang pagmamalasakit sa aming kabarangay	
10 pagsunod sa sinasabi sa kung ano ang dapat preparedness for emergency		tl pagsunod sa sinasabi sa kung ano ang dapat gawin paghandaan ang lahat ng bibitbitin sa tuwing may sakuna	
11 pagkakaroong ng early warning device	early warning system	tl pagkakaroong ng early warning device	
12 pagbibigay ng humanitarian assistance goods disaster relief		tl pagbibigay ng humanitarian assistance goods sa panahon ng kalamidad	
13 nais ko po sana magkaroong pa ng mga it information campaign and capacity building		tl nais ko po sana magkaroong pa ng mga ibat ibang paraan upang lalo pang maging handa ang aming mga kisiguruhing magkaroong ng mga basurahan sa buong barangay dahil ito ang panganahing sanhi ng kalamidad	
14 siguruhing magkaroong ng mga basurahan sinfrastructure maintenance and management		tl seminar tungkol sa bagyo at lindol para magsapaghandaan	
15 seminar tungkol sa bagyo at lindol para magsapaghandaan		tl siguruhing magkaroong ng mga basurahan sa buong barangay dahil ito ang panganahing sanhi ng kalamidad	
16 maging alerto lagi sa mga inaasahang ba preparedness for emergency		tl maging alerto lagi sa mga hini inaasahang bagay tulog opisyal man o hindi ng brgy	
17 magkaisa ang lahat maging totoo ang bawFilipino values		tl magkaisa ang lahat maging totoo ang bawat isa sa kanilang tungkulin	
18 bawat pamilya ay dapat nagkakaisa	Filipino values	tl bawat pamilya ay dapat nagkakaisa	
19 maging handa sa ano mang dumating ng bpreparedness for emergency		tl maging handa sa ano mangdumating na bagyobaha sa komunidad sa ating bansa	
20 magkaroong ng komunikasyon at iktutin ng rearly warning system		tl magkaroong ng komunikasyon at iktutin ng mga council members para makita ang sitwasyon ng barangay sa tuwing may sakuna	
21 tumulong sa mga karatig bahay at magbigay	Filipino values	tl tumulong sa mga karatig bahay at magbigay ng naitabing pagkain sa mga napinsala ng bagyo	
22 maayos at malinaw na pagpapiano bilang ipreparedness for emergency		tl maayos at malinaw na pagpapiano bilang paghandaan sa kalamidad	
23 maagang paghandaan bago dumating ang ipreparedness for emergency		tl maagang paghandaan bago dumating ang sakuna	
24 maging alerto at maging handa	ipreparedness for emergency	tl maging alerto at maging handa	
25 magdasal	Filipino values	en magdasal	

Figure C.4: Excel Report: Response Information

E
1 magkaisa VBW dapat VB5 ang DTC mga DTCP tao NNC
2 magkaroong VBAF ng CCB pagkakaisa NCC upang CCB sa CCT mga DTCP darating NNC na CCP mga DTCP sakuna NCC ay LM malalagpasan VBTF
3 magkaroong VBAF ng CCB komunikasyon NCC kung RBQ saan RBQ magkikita VBTF sa CCT panahon NCC ng CCB kalamidad NNC
4 paglilinis NCC ng CCB kanal NCC wastong JJID_CCP pagtatapon NNC ng CCB basura NNC at CCA kailangan VBS magikot VBFW ikot NNC ang DTC mga DTCP tanod NNC upang CCB ba
5 malawakang JJID_CCP information FW drive FW
6 bago RBW dumating VBAF ang DTC bagyo NCC magkaroong VBW ng CCB early FW warning FW system FW para CCT mas JJCC maging VBW handa JJD ang DTC mga DTCP tao NNC
7 magkaroong VBAF ng CCB early FW warning FW upang CCB maging VBW handa JJD ang DTC mga DTCP tao NNC sa CCT darating NNC na CCP bagyo NNC
8 higit JJCC na CCP pagtibayin VBOF ang DTC early FW warning FW system FW device FW magkaroong VBAF ng CCB maintenance FW quarterly FW para CCT masigurong VBW_CCP m
9 lalo JJCC pang RBL_CCP lumawat VBAF at CCA lumago VBAF ang DTC pagmamalasakit NNC sa CCT aming PRSP_CCP kabarangay NNC
10 pagsunod NNC sa CCT sinasabi VBTR sa CCT kung RBK and RBQ ang DTC dapat VBS gawin VBOP paghandaan VBOF ang DTC lahat PRI ng CCB bibitbitin VBF sa CCT tuwing RBW ma
11 pagkakaroong VBW ng CCB early FW warning FW device FW
12 pagbibigay VBW ng CCB humanitarian VBT assistance FW goods FW sa CCT panahon NNC ng CCB kalamidad NNC
13 nais VBS ko PRS po RBS sana VBS magkaroong VBW pa RBL po RBS ngmga NNC ibat JJD ibang PRI_CCP paraan NNC upang CCB lalo JJCC pang RBL_CCP maging VBW handa JJD ang I
14 siguruhing RBD_CCP magkaroong VBW ng CCB mga DTCP basurahan NNC sa CCT buong PRI_CCP barangay NNC dahil CCR ito PRO ang DTC pangunahing JJID_CCP sanhi NNC ng CCB I
15 seminar JJID tungkol RBR sa CCT bagyo NNC at CCA lindol NNC para CCT mapaghandaan VBW
16 maging VBW alerto JJD lagi RBW sa CCT mga DTCP hindri RBF inaasahang VBT_CCP bagay NNC magtulog VBW_CCP tulog NNC opisyal JJD man RBI o CCT hindri RBF ng CCB brgy
17 magkaisa VBW ang DTL lahat PRI maging VBW totoo RB7 ang DTC bawat PRI isa PRI sa CCT kanilang PRSP_CCP tungkulin NNC
18 bawat PRI pamilya NNC ay LM dapat VB5 nagkakaisa VBTR
19 maging VBW handa JJD sa CCT ano PRQ mangdumating VBTF na CCP bagyobaha NNC sa CCT komunidad NNC sa CCT ating PRSP_CCP bansa NNC
20 magkaroong VBAF ng CCB komunikasyon NNC at CCA iktutin VBOF ng CCB mga DTCP council FW members FW para CCT makita VBW ang DTC sitwasyon NNC ng CCB barangay NNC
21 tumulong VBAF sa CCT mga DTCP karatig JJD bahay NNC at CCA magbigay VBAF ng CCB naitabing VBT_CCP pagkain NNC sa CCT mga DTCP napinsala VBT ng CCB bagyo NNC
22 maayos JJD at CCA malinaw JJD na CCP pagpapiano VBW bilang CCT paghandaan NNC sa CCT kalamidad NNC
23 maagang JJID_CCP paghandaan NNC bago RBW dumating VBAF ang DTC sakuna NNC
24 maging VBW alerto JJD at CCA maging VBW handa JJD
25 magdasal NN

Figure C.5: Excel Report: Part-of-Speech (Stanford Format)

F	G	H	I
1 magkaisa dapat ang mga tao			
2 magkaroong ng pagkakaisa			
3 magkaroong ng komunikasyon	magkikita sa panahon		
4 5 wastong pagtatapon	kailangan magikot ikot	bantayan mga gamit	
5 malawakang information drive			
6 6 dumating ang bagyo	magkaroong ng early warning system	maging handa ang mga tao	
7 7 magkaroong ng early warning	maging handa ang mga tao		
8 8 pagtibayin ang early warning system device	magkaroong ng maintenance quarterly	masigurong maayos ito bago dumating ang isang kalamidad	
9 9 lumawat sa lumago ang pagmamalasakit			
10 10 sinasabi sa kung ano ang dapat gawin paghandaan ang lahat ng bibitbitin sa tuwing may sakuna			
11 11 pagkakaroong ng early warning device			
12 12 pagbibigay ng humanitarian assistance goods			
13 13 nais ko po sana magkaroong pa ng mga	ibat ibang paraan	maging handa ang aming mga kapitbahay	
14 14 magkaroong ng mga basurahan	pangunahing sanhi		
15 15 seminar tungkol sa bagyo at lindol			
16 16 maging alerto lagi sa mga hini inaasahang bagay	magtulog tulog	opisyal man o hindi ng brgy	
17 17 magkaisa ang lahat maging totoo ang bawat isa sa kanilang tungkulin			
18 18			
19 19 maging handa sa ano mangdumating na bagyobaha			
20 20 magkaroong ng komunikasyon at iktutin ng mga council members	makita ang sitwasyon	may kalamidad	
21 21 tumulong sa mga karatig bahay at magbigay ng naitabing pagkain	napinsala ng bagyo		
22 22 maayos at malinaw na pagpapiano bilang paghanda	dumating ang sakuna		
23 23 maagang paghanda			

Figure C.6: Excel Report: Insight Phrases

	A	B	C	D	E	F
1	1	magkaisa	tao			
2	2	magkaroon	pagkakaiisa			
3	3	magkaroon	komunikasyon			
4	3	magkikita	panahon			
5	4	wastong	pagtatapon			
6	4	kailangan	ikot			
7	4	bantayan	gamit			
8	5	malawakang	information	drive		
9	6	dumating	bagyo			
10	6	magkaroon	early	warning	system	
11	6	maging	tao			
12	7	magkaroon	early	warning		
13	7	maging	tao			
14	8	pagtibayin	early	warning	system	device
15	8	magkaroon	maintenance	quarterly		
16	8	masigurong	kalamidad			
17	9	lumawat	pagmamalasakit			
18	10	sinasabi	sakuna			
19	11	pagkakaroon	early	warning	device	
20	12	pagbibigay	assistance	goods		
21	13	nais	ngmga			
22	13	ibat	paraan			
23	13	maging	kapitbahay			
24	14	magkaroon	basurahan			
25	14	pangunahing	sanhi			
26	15	cominar	baus	lindol		

Figure C.7: Excel Report: Insight Word Sets

A	B	C	D	E	F	G	H
1 COMMUNITY-WIDE LOGISTIC SUPPORT FOR DISASTER RESPONSE							
2 Cluster 1							
3 932	1 be		info				
4 FILIPINO VALUES							
5 Cluster 1							
6 1 17 63 427 433 508 832	7 magkaisa, nagkakaisa		tao, tungkulin, samang, samahan, ren, kalamidad				
7 Cluster 2							
8 2 44	2 magkaroon, nagkakaroon		pagkakaiisa, lugar				
9 Cluster 3							
10 9	1 lumawat		pagmamalasakit				
11 Cluster 4							
12 21 49 423 525 540	6 tumulong		bahay, pagkain, bagyo, magcommunity, volunteer, bata, tulong, barangay				
13 Cluster 5							
14 21	1 napinsala		bagyo				
15 Cluster 6							
16 31 39 757 888	4 dapat		pakikiisa, komunidad, panahon, mgkaisa				
17 Cluster 7							
18 31 426 520 715 832	5 may		sakunang, bagyo, bayanihanpara, mwasan, kalamidadwag (kalamidad)				
19 Cluster 8							
20 31	1 dumating		kalamidad				
21 Cluster 9							
22 41 443 621 721	4 maging		darating, kalagayan, agrisibo, kalamidad				
23 Cluster 10							
24 53	1 makiisa		komunidad				
25 Cluster 11							
26 55 161 133 142 152 1521 1611 604 1722 12 tumulong, mangtulong? makatulong, minaghahanda, barangay/barangay, kalamidad, komunidad, tulong, ectaca							

Figure C.8: Excel Report: Cluster List

A	B	C	D	E	F	G	H	I	J
1 INFORMATION CAMPAIGN AND CAPACITY BUILDING									
2 Cluster 1									
3 13 46 208 236 243 307 311 598 603 19 maging, palaging, laging			kapitbahay, tao, disaster, mamawayan, aware, kabarangay, kabaro, kalamidad, darating, pagbe						
4 Cluster 2									
5 26 208 209 339 507 576 649 762 766 18 dapat			barangay, seminar, oras, prepared, kalamidad, sakuna, posters, drill, like, for, example, about, p						
6 Cluster 3									
7 33 68 209 307 310 311 339 696 716 16 magkaroon, magkakaroon			seminars (seminar, seminarsdrill), about, kaalaman (kaalam), disaster, drill, sos, program, regar						
8 Cluster 4									
9 80 121 132 149 156 169 215 222 234 15 be			advantage, possibilities, calamities, barangay, times (time), drills, typhoon, disaster, preparedne						
10 Cluster 5									
11 130 215 245 249 259 266 340 390 79 12 have			assembly, disaster, drill, place, representative, check, emergency, needs, training (taraining), ba						
12 Cluster 6									
13 207 238 269 289 332 343 344 382 39 12 conducting, magconduct, conduct, pagseminars (seminar), drills, community, assembly, lot, organisasyon, regarding									
14 Cluster 7									
15 101 104 169 287 333 343 347 396 58 10 preparing, prepare			typhoon, disaster, seminar, times						
16 Cluster 8									
17 169 222 303 331 343 358 614 627 65 10 help			prepare, families, disasters (disaster), outcomes, community, idea, people, signage, calamity						
18 Cluster 9									
19 26 362 764 813 852 883 903 926 9 may			pangangailangan, seminar, meetings, kalamidad, conduct, tao, kinalamnan, sakuna, darating						
20 Cluster 10									
21 101 104 179 218 337 347 350 386 40 9 inform, informs, informing			consequence, people, neighbor, subordinate, community, kind, whoe, sitio						
22 Cluster 11									
23 152 200 216 303 594 627 798 876 92 9 giving, living			disaster, drill, knowledge, seminars, leaflets, information, community						
24 Cluster 12									
25 344 382 390 398 485 700 842 846 86 9 regarding			prevention, disaster (disasters), awareness, preparation (preparations), incoming, calamities, pr						
26 cluster 13									

Figure C.9: Excel Report: Ranked Cluster List

Appendix D

Generated Reports

These are merely samples from three Word document reports.

D.1 Sørensen-Dice Coefficient, Organized by Response Categories

FILIPINO TEXT ANALYSIS TOOL REPORT

Mar-24-2020 20:24:54

The information below were extracted and organized automatically.

INFORMATION CAMPAIGN AND CAPACITY BUILDING

Entry 1

ID/s:

26|208|209|339|507|576|580|649|762|766|
811|813|836|852|854|885

Frequency: 20**Proposed action:** dapat

Target: barangay, seminar, oras, prepared, kalamidad, sakuna, posters, paalala, drill, like, for, example, about, paraan, encourage, before, during, and, after, of, the, calamity, my, weekly, mg, roong

Entry 2

ID/s:

13|46|208|236|243|307|311|598|603|618|6|
31|651|758|760|766|854|903|925

Frequency: 19**Proposed action:** maging, palaging, laging

Target: kapitbahay, tao, disaster, mamamayan, aware, kabbarangay, kabaro, kalamidad, darating, beforeafter, and, during, pra, bagyo, trahedy, gawim, paalala

Entry 3

ID/s:

33|68|209|289|307|310|311|339|696|716|7|
62|787|800|847|885|926|930

Frequency: 19**Proposed action:** magkaroon, magkaron, magkakaroon, karoon

Target: seminars (seminar, seminarsdrill), about, kaalaman (kaalam), disaster, drill, sos, program, regarding, and, iinvite, emergency, kits, training, progma, pra, disciplina, mamamayan, weekly, meeting, organisasyon, orientation, barangay

Entry 4

ID/s:

80|121|132|149|156|169|215|222|234|340|
585|654|794|831|929

Frequency: 15**Proposed action:** be

Target: advantage, possibilities, calamities, dont, times (time), drills, typhoon, disaster, preparedness (prepare), programs, officials, instructions, orientation, effects, duty

Entry 5

ID/s:

130|215|245|249|259|266|340|390|794|838

Frequency: 12**Proposed action:** have

Target: assembly, disaster, drill, place, representative, check, emergency, needs, training (taraining), barangay, officials, seminar, meeting

Entry 6

ID/s:

169|222|303|331|343|358|385|614|627|654|
661

Frequency: 11**Proposed action:** help

Target: prepare, families, disasters (disaster), outcomes, community, idea, seminars, people, signage, calamity

Entry 7

ID/s:

101|104|169|287|333|343|347|396|583|644

Frequency: 10**Proposed action:** preparing, prepare

Target: typhoon, disaster, seminar, times

Entry 8

ID/s: 207 238 269 332 343 344 382 396 595 677 Frequency: 10 Proposed action: conducting, conduct Target: seminars, drills, community, assembly, lot	ID/s: 243 536 579 598 599 618 631 736 Frequency: 8 Proposed action: magpasagawa, magsagawa, magsasagawa Target: disaster, drill (drills), pagpupulong, orientation, seminars (seminar), grupo, mg, programa
Entry 9	Entry 15
ID/s: 101 104 179 218 337 347 350 386 405 Frequency: 9 Proposed action: inform, informs, informing Target: consequence, people, neighbor, subordinate, community, kind, whoe, sitio	ID/s: 26 362 764 813 883 926 Frequency: 7 Proposed action: may Target: pangangailangan, seminar, meetings, kalamidad, conduct, kinalaman, darating
Entry 10	Entry 16
ID/s: 152 200 216 303 594 627 798 876 927 Frequency: 9 Proposed action: giving, living Target: disaster, drill, knowledge, seminars, leaflets, information, community	ID/s: 214 215 269 283 382 677 Frequency: 7 Proposed action: having Target: meeting, drills, class, seminar, knowledge, house
Entry 11	Entry 17
ID/s: 344 382 390 398 485 700 842 846 860 Frequency: 9 Proposed action: regarding Target: prevention, disaster (disasters), awareness, preparation (preparations), incoming, calamities, preparedness	ID/s: 46 54 469 649 736 836 Frequency: 6 Proposed action: magbigay, bibigay, mabigay Target: seminar, kaalaman, bahay, information
Entry 12	Entry 18
ID/s: 120 385 614 707 869 895 Frequency: 8 Proposed action: do Target: disasters (disaster), calamities, drills, seminar, information, drive, emergencies, training	ID/s: 209 469 735 744 813 903 Frequency: 6 Proposed action: malaman, malalaman, nalaman, nalamang Target: tao, sakuna, kabbarangay, seminar
Entry 13	Entry 19
ID/s: 160 181 389 400 403 406 485 893 Frequency: 8 Proposed action: give Target: seminar (seminars), seminars, information, tips, knowledge, ideas	ID/s: 236 289 462 553 604 Frequency: 6 Proposed action: mag Target: participate, conduct, impormasyon, schedule, update, seminars
Entry 14	Entry 20
	ID/s: 576 598 717 854 903 Frequency: 6

Proposed action: makakatulong	Proposed action: tamang
Target: barangay, lubus	Target: kaukulan, oras, desaster, impormansyon
Entry 21	Entry 28
ID/s: 491 628 735 797 897	ID/s: 15 592 736
Frequency: 5	Frequency: 3
Proposed action: kailangan, kilangan, kakailanganin	Proposed action: seminar
Target: kaalaman, disiplina, pgdating, conduct, drmm, local, government, magkroon, residente	Target: bagyo, lindol, disaster, information, drive
Entry 22	Entry 29
ID/s: 130 149 591 687	ID/s: 26 339 771
Frequency: 4	Frequency: 3
Proposed action: is	Proposed action: sanayin, sana
Target: barangay, disaster, lgu, calamity	Target: lugar, awareness, programs, weekly
Entry 23	Entry 30
ID/s: 149 157 297 389	ID/s: 84 88 846
Frequency: 4	Frequency: 3
Proposed action: think	Proposed action: organize
Target: barangay, meeting, seminars, officials	Target: events, disaster, drills, talk
Entry 24	Entry 31
ID/s: 169 331 585 630	ID/s: 249 717 797
Frequency: 4	Frequency: 3
Proposed action: know, knowing	Proposed action: handle, handa
Target: panicking, case, majority, places	Target: emergency, situations, sakuna, paaralan, evacuation, center
Entry 25	Entry 32
ID/s: 287 303 610 919	ID/s: 766 800
Frequency: 4	Frequency: 3
Proposed action: provide, provides, providing	Proposed action: magandang, maganda
Target: needs, seminars, awareness, evacuation, center	Target: conduct, tulong, seminar
Entry 26	Entry 33
ID/s: 337 630 677 801	ID/s: 13 811
Frequency: 4	Frequency: 2
Proposed action: are	Proposed action: nais, naisip
Target: disaster, things, typhoon	Target: ngmga, barangay
Entry 27	Entry 34
ID/s: 717 766 897 931	ID/s: 13 930
Frequency: 4	Frequency: 2
Proposed action: ibat	Proposed action: ibat
Target: paraan, sona	Target: paraan, sona

Entry 35	Target: seminar, barangay
ID/s: 54 602	Entry 43
Frequency: 2	
Proposed action: maglinis, linisin	ID/s: 249 254
Target: kanal, harap, daan	Frequency: 2
Entry 36	Proposed action: teach
ID/s: 120 287	Target: citizen, disaster
Frequency: 2	Entry 44
Proposed action: remind	
Target: citizens, community, members	ID/s: 305 582
Entry 37	Frequency: 2
ID/s: 126 931	Proposed action: make
Frequency: 2	Target: drill, information, drive
Proposed action: respond, rumesponde	Entry 45
Target: calamities, sakuna	
Entry 38	ID/s: 310 760
ID/s: 126 627	Frequency: 2
Frequency: 2	Proposed action: lumawak, gumawa
Proposed action: staying, stay	Target: kaalaman, programa
Target: touch, times	Entry 46
Entry 39	
ID/s: 130 157	ID/s: 347 583
Frequency: 2	Frequency: 2
Proposed action: gather	Proposed action: create
Target: citizens, share, ideas	Target: project, disaster, risk, orrientation
Entry 40	Entry 47
ID/s: 130 178	
Frequency: 2	ID/s: 362 578
Proposed action: educate	Frequency: 2
Target: disaster, people	Proposed action: ginagawa, ginagamit
Entry 41	Target: samin, babala
ID/s: 132 595	Entry 48
Frequency: 2	
Proposed action: require, requiring	ID/s: 469
Target: family, members (member)	Frequency: 2
Entry 42	Proposed action: dating, dadating
ID/s: 132 842	Target: kalamidad, drills, paraan
Frequency: 2	Entry 49
Proposed action: join	
	ID/s: 565 602
	Frequency: 2
	Proposed action: makaiwas, mkaiwas
	Target: skuna, kalamidad, storms
	Entry 50
	ID/s: 588 804
	Frequency: 2

Proposed action: participating, participate	ID/s: 26
Target: garbage, disposal, disaster, awareness, activities	Frequency: 1
Entry 51	Proposed action: pagsapit
	Target: bagyobaha
	Entry 59
ID/s: 588 933	ID/s: 33
Frequency: 2	Frequency: 1
Proposed action: adds, add	Proposed action: mapaglalaban
Target: disaster, drainage, systems	Target: sakuna
Entry 52	Entry 60
ID/s: 603 630	ID/s: 42
Frequency: 2	Frequency: 1
Proposed action: drill	Proposed action: pagsabihan
Target: dapat, facebook	Target: kaalaman
Entry 53	Entry 61
ID/s: 610 860	ID/s: 42
Frequency: 2	Frequency: 1
Proposed action: held	Proposed action: nasasakupan
Target: program, session	Target: kalamidad
Entry 54	Entry 62
ID/s: 628 797	ID/s: 80
Frequency: 2	Frequency: 1
Proposed action: papaalam, nagpapaalala	Proposed action: door
Target: tao, oras	Target: disaster, awareness
Entry 55	Entry 63
ID/s: 630 681	ID/s: 80
Frequency: 2	Frequency: 1
Proposed action: using, focusing	Proposed action: maximize
Target: news, idea	Target: information, dissemination
Entry 56	Entry 64
ID/s: 801 831	ID/s: 81
Frequency: 2	Frequency: 1
Proposed action: hit	Proposed action: put
Target: cases, area	Target: sirens
Entry 57	Entry 65
ID/s: 5	ID/s: 81
Frequency: 1	Frequency: 1
Proposed action: malawakang	Proposed action: alert
Target: information, drive	Target: citizens
Entry 58	

Entry 66	Target: people
ID/s: 84	Entry 74
Frequency: 1	
Proposed action: concerning	
Target: disaster, preparedness	
Entry 67	
ID/s: 121	Entry 75
Frequency: 1	
Proposed action: implement	
Target: disaster, drills	
Entry 68	
ID/s: 121	Entry 76
Frequency: 1	
Proposed action: cause	
Target: damage	
Entry 69	
ID/s: 129	Entry 77
Frequency: 1	
Proposed action: wasnt	
Target: drill	
Entry 70	
ID/s: 129	Entry 78
Frequency: 1	
Proposed action: been	
Target: help	
Entry 71	
ID/s: 130	Entry 79
Frequency: 1	
Proposed action: thrice	
Target: month	
Entry 72	
ID/s: 169	Entry 80
Frequency: 1	
Proposed action: initiating	
Target: seminars	
Entry 73	
ID/s: 169	Entry 81
Frequency: 1	
Proposed action: comes	
	ID/s: 218
	Frequency: 1
	Proposed action: go
	Target: calamities
	ID/s: 256
	Frequency: 1

Proposed action: tell	Frequency: 1
Target: information	Proposed action: turuan
Entry 82	Entry 90
ID/s: 259	ID/s: 336
Frequency: 1	Frequency: 1
Proposed action: afformentioned	Proposed action: maihnada
Target: problems	Target: tao
Entry 83	Entry 91
ID/s: 259	ID/s: 343
Frequency: 1	Frequency: 1
Proposed action: encounter	Proposed action: establish
Target: disaster	Target: program
Entry 84	Entry 92
ID/s: 281	ID/s: 344
Frequency: 1	Frequency: 1
Proposed action: magturo	Proposed action: disseminating
Target: first, aid	Target: information
Entry 85	Entry 93
ID/s: 289	ID/s: 364
Frequency: 1	Frequency: 1
Proposed action: gusto	Proposed action: planning
Target: volunteer	Target: incase
Entry 86	Entry 94
ID/s: 296	ID/s: 381
Frequency: 1	Frequency: 1
Proposed action: guide	Proposed action: performing
Target: community	Target: drill
Entry 87	Entry 95
ID/s: 296	ID/s: 383
Frequency: 1	Frequency: 1
Proposed action: doing	Proposed action: t
Target: meeting	Target: incoming, disaster
Entry 88	Entry 96
ID/s: 296	ID/s: 389
Frequency: 1	Frequency: 1
Proposed action: talks	Proposed action: pay
Target: safety	Target: visit, house
Entry 89	Entry 97
ID/s: 307	

ID/s: 389	Entry 105
Frequency: 1	
Proposed action: contain	
Target: safety, tips	
Entry 98	
ID/s: 394	Entry 106
Frequency: 1	
Proposed action: discussing	
Target: safety, protocols	
Entry 99	
ID/s: 394	Entry 107
Frequency: 1	
Proposed action: simulating	
Target: disasters	
Entry 100	
ID/s: 400	Entry 108
Frequency: 1	
Proposed action: needed	
Target: evacuation	
Entry 101	
ID/s: 403	Entry 109
Frequency: 1	
Proposed action: survive	
Target: disaster	
Entry 102	
ID/s: 462	Entry 110
Frequency: 1	
Proposed action: natutunan	
Target: disaster, survey	
Entry 103	
ID/s: 469	Entry 111
Frequency: 1	
Proposed action: maiging	
Target: barangay	
Entry 104	
ID/s: 565	Entry 112
Frequency: 1	
Proposed action: magsgwa	
Target: kaalaman	

Target: series	Proposed action: talakayin
Entry 113	Entry 121
ID/s: 645	ID/s: 677
Frequency: 1	Frequency: 1
Proposed action: ibahagi	Proposed action: making
Target: paghahanda	Target: household
Entry 114	Entry 122
ID/s: 647	ID/s: 717
Frequency: 1	Frequency: 1
Proposed action: please	Proposed action: maaari
Target: information, dissemination	Target: pamamagitan
Entry 115	Entry 123
ID/s: 651	ID/s: 751
Frequency: 1	Frequency: 1
Proposed action: magpasimuno	Proposed action: magkron
Target: drill	Target: pagpupulong
Entry 116	Entry 124
ID/s: 658	ID/s: 797
Frequency: 1	Frequency: 1
Proposed action: hold	Proposed action: nakikita
Target: seminars	Target: disciplina
Entry 117	Entry 125
ID/s: 658	ID/s: 797
Frequency: 1	Frequency: 1
Proposed action: happen	Proposed action: ipinatutupad
Target: beforeduring	Target: tagapamuno
Entry 118	Entry 126
ID/s: 667	ID/s: 797
Frequency: 1	Frequency: 1
Proposed action: leaflets	Proposed action: naglalagay
Target: safety, tips	Target: karatula
Entry 119	Entry 127
ID/s: 676	ID/s: 797
Frequency: 1	Frequency: 1
Proposed action: magtayo	Proposed action: nagkakabit
Target: programs	Target: tarpaulin
Entry 120	Entry 128
ID/s: 676	ID/s: 800
Frequency: 1	

Frequency: 1	ID/s: 842
Proposed action: nangyayari	Frequency: 1
Target: s	Proposed action: cleaning
Entry 129	Target: canals
ID/s: 800	ID/s: 842
Frequency: 1	Frequency: 1
Proposed action: maganada	Proposed action: attending
Target: programa	Target: seminars
Entry 130	Entry 137
ID/s: 800	ID/s: 842
Frequency: 1	Frequency: 1
Proposed action: maiwasan	Proposed action: hinahanda
Target: trahedya	Target: tao
Entry 131	Entry 138
ID/s: 801	ID/s: 847
Frequency: 1	Frequency: 1
Proposed action: being	Proposed action: gagawin
Target: disaster	Target: kalamidad
Entry 132	Entry 139
ID/s: 801	ID/s: 847
Frequency: 1	Frequency: 1
Proposed action: starts	Proposed action: reduce
Target: others	Target: damages
Entry 133	Entry 140
ID/s: 811	ID/s: 869
Frequency: 1	Frequency: 1
Proposed action: pumunta	Proposed action: maayos
Target: briefing	Target: pagpapatupad
Entry 134	Entry 141
ID/s: 811	ID/s: 883
Frequency: 1	Frequency: 1
Proposed action: makaligtas	Proposed action: mabigyan
Target: taas	Target: ideya
Entry 135	Entry 142
ID/s: 811	ID/s: 885
Frequency: 1	Frequency: 1
Proposed action: magparticcate	Proposed action: paparating
Target: klase	Target: kalamidad
Entry 136	Entry 143

<p>Entry 144</p> <p>ID/s: 886 Frequency: 1 Proposed action: w Target: brgy, level</p> <p>Entry 145</p> <p>ID/s: 903 Frequency: 1 Proposed action: magpatawag Target: drill, pra</p> <p>Entry 146</p> <p>ID/s: 903 Frequency: 1 Proposed action: ready Target: lahatmgipon</p> <p>Entry 147</p> <p>ID/s: 903 Frequency: 1 Proposed action: importanteng Target: bagaypagkain</p> <p>Entry 148</p> <p>ID/s: 910 Frequency: 1 Proposed action: nagsmemeeting Target: araw</p> <p>Entry 149</p> <p>ID/s: 910 Frequency: 1 Proposed action: pumasok Target: bagyo</p> <p>Entry 150</p> <p>ID/s: 910 Frequency: 1 Proposed action: mabilis Target: aksyon</p> <p>Entry 151</p> <p>ID/s: 926 Frequency: 1 Proposed action: sakunamaari</p>	<p>Target: araw</p> <p>Entry 152</p> <p>ID/s: 929 Frequency: 1 Proposed action: assist Target: citizen</p> <hr/> <p style="text-align: center;">DISASTER RELIEF</p> <hr/> <p>Entry 1</p> <p>ID/s: 12 460 559 566 776 Frequency: 5 Proposed action: pagbibigay, magbibigay, bibigay Target: assistance, goods, pagkain, medisina, evacuation, tulong, pangangailangan (panganga), grocery</p> <p>Entry 2</p> <p>ID/s: 110 261 795 878 Frequency: 4 Proposed action: provide Target: food, aid, kits, supply, goods, relief, assistance</p> <p>Entry 3</p> <p>ID/s: 205 522 776 Frequency: 3 Proposed action: kailangan, ilangan Target: relief, goods, resque, pagkain, pamilya</p> <p>Entry 4</p> <p>ID/s: 421 828 Frequency: 2 Proposed action: giving Target: relief, goods</p> <p>Entry 5</p> <p>ID/s: 110 Frequency: 1 Proposed action: are Target: need</p>
---	--

Entry 6	Target: pagkaen, gamot
ID/s: 205	Entry 14
Frequency: 1	ID/s: 440
Proposed action: handa	Frequency: 1
Target: sakuna	Proposed action: dapat
Entry 7	Target: lugar
ID/s: 299	Entry 15
Frequency: 1	ID/s: 440
Proposed action: cannot	Frequency: 1
Target: afford	Proposed action: mailikasunahin
Entry 8	Target: pagbibigay
ID/s: 379	Entry 16
Frequency: 1	ID/s: 446
Proposed action: support	Frequency: 1
Target: goods	Proposed action: papadala
Entry 9	Target: kunting
ID/s: 418	Entry 17
Frequency: 1	ID/s: 497
Proposed action: matulungan	Frequency: 1
Target: bangka, pagkain	Proposed action: maagap
Entry 10	Target: tulong
ID/s: 420	Entry 18
Frequency: 1	ID/s: 526
Proposed action: gawing	Frequency: 1
Target: pagtulong	Proposed action: ibakwisyon
Entry 11	Target: pagbbigay
ID/s: 420	Entry 19
Frequency: 1	ID/s: 548
Proposed action: tinamaan	Frequency: 1
Target: kalamidad	Proposed action: sana
Entry 12	Target: brgy
ID/s: 420	Entry 20
Frequency: 1	ID/s: 548
Proposed action: may	Frequency: 1
Target: sakit	Proposed action: inaasahang
Entry 13	Target: sakuna
ID/s: 437	Entry 21
Frequency: 1	ID/s: 795
Proposed action: mag	Frequency: 1

Proposed action: i	Entry 4
Target: government	ID/s: 134 879
	Frequency: 3
	Proposed action: need, needed
	Target: supplies, assistance, disaster
	Entry 5
	ID/s: 154 392 932
	Frequency: 3
	Proposed action: be
	Target: publicsoundnotifsystem, technology, info
	Entry 6
	ID/s: 167 629 777
	Frequency: 3
	Proposed action: dapat
	Target: evacuation, center, barangay, truck
	Entry 7
	ID/s: 242 247 659
	Frequency: 3
	Proposed action: build
	Target: place, evacuation, area
	Entry 8
	ID/s: 244 556
	Frequency: 3
	Proposed action: may
	Target: tao, bangka, baha
	Entry 9
	ID/s: 488 590 810
	Frequency: 3
	Proposed action: gamit
	Target: food, for, volunteers, darating, bumbero
	Entry 10
	ID/s: 40 488
	Frequency: 2
	Proposed action: magkaroon
	Target: medical, kit, flashlight, budget
	Entry 11
	ID/s: 95 911

COMMUNITY-WIDE LOGISTIC SUPPORT FOR DISASTER RESPONSE

Entry 1	
ID/s: 106 114 142 147 171 212 276 397	
Frequency: 8	
Proposed action: have	
Target: emergency, supplies, sirens, equipment, funds, evacuation, place, aid, kit	
Entry 2	
ID/s: 109 221 302 329 409	
Frequency: 5	
Proposed action: providing, provide	
Target: shelterarea, emergency, kits, evacuation, center, safety, gears	
Entry 3	
ID/s: 71 417	
Frequency: 3	
Proposed action: karagdagang, dagdagan	
Target: kagamitan, community, volunteers, budget	

Frequency: 2	ID/s: 114
Proposed action: prepare	Frequency: 1
Target: place, roadway	Proposed action: armed
Entry 12	
ID/s: 142 675	Target: forces
Frequency: 2	ID/s: 135
Proposed action: having	Frequency: 1
Target: boats, secure, storage, facility	Proposed action: create
Entry 13	
ID/s: 154 392	Target: evacuation, area
Frequency: 2	ID/s: 142
Proposed action: is	Frequency: 1
Target: facilities, source	Proposed action: back
Entry 14	
ID/s: 171 270	Target: funds
Frequency: 2	ID/s: 142
Proposed action: make	Frequency: 1
Target: barangay, gym, evacuation, center	Proposed action: know
Entry 15	
ID/s: 242 247	Target: advance
Frequency: 2	ID/s: 154
Proposed action: are	Frequency: 1
Target: condition	Proposed action: cover
Entry 16	
ID/s: 51	Target: areas
Frequency: 1	ID/s: 170
Proposed action: dumating	Frequency: 1
Target: dilubyobagyo	Proposed action: disseminate
Entry 17	
ID/s: 51	Target: evacuation, plan
Frequency: 1	ID/s: 171
Proposed action: nakahanda	Frequency: 1
Target: gamit	Proposed action: etc
Entry 18	
ID/s: 51	Target: evacuees
Frequency: 1	ID/s: 171
Proposed action: mataas	Frequency: 1
Target: lugar	Proposed action: contains
Entry 19	
	Target: food, electricity
Entry 20	
ID/s: 135	
Frequency: 1	
Proposed action: create	
Target: evacuation, area	
Entry 21	
ID/s: 142	
Frequency: 1	
Proposed action: back	
Target: funds	
Entry 22	
ID/s: 142	
Frequency: 1	
Proposed action: know	
Target: advance	
Entry 23	
ID/s: 154	
Frequency: 1	
Proposed action: cover	
Target: areas	
Entry 24	
ID/s: 170	
Frequency: 1	
Proposed action: disseminate	
Target: evacuation, plan	
Entry 25	
ID/s: 171	
Frequency: 1	
Proposed action: etc	
Target: evacuees	
Entry 26	
ID/s: 212	
Frequency: 1	
Proposed action: contains	
Target: food, electricity	

Entry 27	Target: instructions
ID/s: 242	Entry 35
Frequency: 1	
Proposed action: serve	ID/s: 378
Target: place	Frequency: 1
	Proposed action: giving
	Target: equipments, incase
Entry 28	Entry 36
ID/s: 242	
Frequency: 1	ID/s: 392
Proposed action: suffered	Frequency: 1
Target: typhoons	Proposed action: benefit
	Target: baranggay
Entry 29	Entry 37
ID/s: 244	
Frequency: 1	ID/s: 392
Proposed action: magpatayo	Frequency: 1
Target: evacuation, center	Proposed action: communicating
	Target: weather, predictions
Entry 30	Entry 38
ID/s: 244	
Frequency: 1	ID/s: 417
Proposed action: tamaan	Frequency: 1
Target: sakuna	Proposed action: makabii
	Target: gamit, volunteer
Entry 31	Entry 39
ID/s: 302	
Frequency: 1	ID/s: 434
Proposed action: ex	Frequency: 1
Target: boat	Proposed action: makasakay
	Target: bangka, tao
Entry 32	Entry 40
ID/s: 302	
Frequency: 1	ID/s: 484
Proposed action: help	Frequency: 1
Target: people	Proposed action: magtapon
	Target: basura
Entry 33	Entry 41
ID/s: 330	
Frequency: 1	ID/s: 534
Proposed action: designed	Frequency: 1
Target: calamity	Proposed action: nakakatulong
	Target: barangay
Entry 34	Entry 42
ID/s: 378	
Frequency: 1	ID/s: 534
Proposed action: buy	Frequency: 1

Proposed action: sana
Target: gamit

Entry 43

ID/s: 653
Frequency: 1
Proposed action: maglagay

Target: evacuation, center

Entry 44

ID/s: 656

Frequency: 1

Proposed action: ready

Target: evacuation, center, int

Entry 45

ID/s: 777

Frequency: 1

Proposed action: mastrandead

Target: pasahero

Entry 46

ID/s: 780

Frequency: 1

Proposed action: kailangan

Target: bangka

Entry 47

ID/s: 810

Frequency: 1

Proposed action: mag-deploy

Target: siren

Entry 48

ID/s: 911

Frequency: 1

Proposed action: transport

Target: vehicle

Entry 49

ID/s: 918

Frequency: 1

Proposed action: inform

Target: storm, signal

INFRASTRUCTURE MAINTENANCE AND MANAGEMENT

Entry 1

ID/s: 52|57|60|74|78|166|279|288|449|466|480|493|543|648|725|743|746|761|790|819|820|822|823|829

Frequency: 25

Proposed action: maglinis, linisin, linisan, panglinis, malinis

Target: kanal, kapaligiran (sakapaligiran, paligid), pausok, siguraduhing, pagtatapon, especially, kalamidad, ilogwag, basura, daanan, imburnal

Entry 2

ID/s: 38|45|368|500|504|512|743|779|820

Frequency: 10

Proposed action: maiwasan, iwasan

Target: pagbaha (pagbabaha, pagbara, baha), pagbabara, paglaganap, kalamidad, pagtatapon

Entry 3

ID/s: 62|476|504|537|822

Frequency: 5

Proposed action: tamang

Target: lugar, tapunan, kapaligiran, lalagyan

Entry 4

ID/s: 69|366|743|812|853

Frequency: 5

Proposed action: dapat

Target: kalinisan, kanal, pangongolekta, lage, pagbabahahindi, dn, paligid

Entry 5

ID/s: 107|361|584|808|834

Frequency: 5

Proposed action: clean, cleaning

Target: streets, canals, surroundings

Entry 6

ID/s: 14|368|478|779

Frequency: 4

Proposed action: magkaroon	ID/s: 127 617
Target: basurahan, sariling, pagtitipon, training, barangay	Frequency: 3
Entry 7	Proposed action: implement
ID/s: 47 231 442 779	Target: cleaning, programs, segregation
Frequency: 4	
Proposed action: ayusin	
Target: drainage, kanal, daanan, lgu	
Entry 8	
ID/s: 62 449 560 822	
Frequency: 4	
Proposed action: itapon, magtapon	
Target: basura, maayus	
Entry 9	
ID/s: 4 38 69	
Frequency: 3	
Proposed action: kailangan, kaylangan	
Target: ikot, waste, dsiposal, kanal	
Entry 10	
ID/s: 43 465 541	
Frequency: 3	
Proposed action: maglagay, ilagay	
Target: garbage, can, kanal, basura	
Entry 11	
ID/s: 62 537 560	
Frequency: 3	
Proposed action: magbara	
Target: kanal, ilog	
Entry 12	
ID/s: 94 474 601	
Frequency: 3	
Proposed action: improve, maimprove	
Target: drainage, system, areas	
Entry 13	
ID/s: 107 282 635	
Frequency: 3	
Proposed action: lessen	
Target: probability, garbage, chances	
Entry 14	
ID/s: 127 617 617	
Frequency: 3	
Proposed action: ensure, ensuring	
Target: flow, surrounding	
Entry 15	
ID/s: 478 779 807	
Frequency: 3	
Proposed action: regular	
Target: basura, pagtatapon, paglilinis, garbagewaste, collections	
Entry 16	
ID/s: 817 820 823	
Frequency: 3	
Proposed action: nagiging, maging	
Target: dahilan, sakuna, sanhi	
Entry 17	
ID/s: 4 482	
Frequency: 2	
Proposed action: wastong	
Target: pagtatapon (pagtapon)	
Entry 18	
ID/s: 4 482	
Frequency: 2	
Proposed action: bantayan	
Target: gamit, kabataang	
Entry 19	
ID/s: 57 905	
Frequency: 2	
Proposed action: walang	
Target: baha	
Entry 20	
ID/s: 98 601	
Frequency: 2	
Proposed action: have	
Target: rules, elevation	
Entry 21	
ID/s: 127 843	
Frequency: 2	
Proposed action: ensure, ensuring	
Target: flow, surrounding	

Entry 22	Proposed action: avoid Target: floods (flooding)
ID/s: 198 308	
Frequency: 2	
Proposed action: be	
Target: garbage, plan	
Entry 23	Entry 30
ID/s: 282 796	
Frequency: 2	
Proposed action: put	
Target: garbage, dumpster, bins	
Entry 24	Entry 31
ID/s: 288 589	
Frequency: 2	
Proposed action: nagtulung, pagtulungan	
Target: kapaligiran, lugar	
Entry 25	Entry 32
ID/s: 308 843	
Frequency: 2	
Proposed action: is	
Target: factor, disposal	
Entry 26	Entry 33
ID/s: 444 825	
Frequency: 2	
Proposed action: maayos	
Target: daanan, kalsada, tapunan	
Entry 27	Entry 34
ID/s: 504 589	
Frequency: 2	
Proposed action: itapun	
Target: basura (basurahan)	
Entry 28	Entry 35
ID/s: 574 635	
Frequency: 2	
Proposed action: perform	
Target: community, cleanups, clean-up, session	
Entry 29	Entry 36
ID/s: 574 796	
Frequency: 2	
	Entry 37
	ID/s: 52

Frequency: 1	ID/s: 308
Proposed action: tanggalin	Frequency: 1
Target: basura, sakuna	Proposed action: mentioned
Entry 38	Target: survey
ID/s: 69	Entry 46
Frequency: 1	ID/s: 372
Proposed action: ilabas	Frequency: 1
Target: basura	Proposed action: pick
Entry 39	Target: trashes
ID/s: 69	Entry 47
Frequency: 1	ID/s: 454
Proposed action: dumating	Frequency: 1
Target: truck	Proposed action: ngbbgay
Entry 40	Target: dhalan
ID/s: 117	Entry 48
Frequency: 1	ID/s: 458
Proposed action: fixing	Frequency: 1
Target: drainage, system	Proposed action: palitan
Entry 41	Target: bata
ID/s: 127	Entry 49
Frequency: 1	ID/s: 458
Proposed action: install	Frequency: 1
Target: drainage, systems	Proposed action: nagtatapon
Entry 42	Target: ngbasura, kanal
ID/s: 163	Entry 50
Frequency: 1	ID/s: 458
Proposed action: secure	Frequency: 1
Target: drainage	Proposed action: daming
Entry 43	Target: tindahan
ID/s: 198	Entry 51
Frequency: 1	ID/s: 458
Proposed action: fix	Frequency: 1
Target: drainage, system	Proposed action: gitna
Entry 44	Target: kalsada
ID/s: 279	Entry 52
Frequency: 1	ID/s: 465
Proposed action: tumulong	Frequency: 1
Target: kagamitan	Proposed action: maipon
Entry 45	Target: tubig

Entry 53	Target: pgbaha
ID/s: 474	Entry 61
Frequency: 1	
Proposed action: mainform	
Target: tao	
Entry 54	
ID/s: 480	
Frequency: 1	
Proposed action: magtanim	
Target: puno, basura	
Entry 55	Entry 62
ID/s: 482	
Frequency: 1	
Proposed action: dagdagan	
Target: bantay	
Entry 56	Entry 63
ID/s: 482	
Frequency: 1	
Proposed action: pasaway	
Target: basura	
Entry 57	
ID/s: 482	
Frequency: 1	
Proposed action: magbigay	
Target: lubid	
Entry 58	Entry 64
ID/s: 482	
Frequency: 1	
Proposed action: gagawing	
Target: hagdan	
Entry 59	Entry 65
ID/s: 482	
Frequency: 1	
Proposed action: makasagip	
Target: bata	
Entry 60	Entry 66
ID/s: 493	
Frequency: 1	
Proposed action: makaiwas	
	Entry 67
	ID/s: 601
	Frequency: 1
	Proposed action: eg
	Target: rivers
	Entry 68
	ID/s: 617
	Frequency: 1
	Proposed action: cause
	Target: flood
	ID/s: 617
	Frequency: 1

Proposed action: thrice	Frequency: 1
Target: week	Proposed action: gumawa
	Target: programa
Entry 69	Entry 77
ID/s: 635	ID/s: 812
Frequency: 1	Frequency: 1
Proposed action: flooding	Proposed action: nakakatulong
Target: rains	Target: pagtaas
Entry 70	Entry 78
ID/s: 688	ID/s: 812
Frequency: 1	Frequency: 1
Proposed action: keep	Proposed action: khit
Target: cleaning	Target: lakas
Entry 71	Entry 79
ID/s: 779	ID/s: 812
Frequency: 1	Frequency: 1
Proposed action: pumping	Proposed action: malaki
Target: station	Target: kanal, basura
Entry 72	Entry 80
ID/s: 779	ID/s: 817
Frequency: 1	Frequency: 1
Proposed action: pinondohan	Proposed action: tanging
Target: taong, bayan	Target: kalsada
Entry 73	Entry 81
ID/s: 790	ID/s: 817
Frequency: 1	Frequency: 1
Proposed action: bumara	Proposed action: namen
Target: basura, baha	Target: epeko
Entry 74	Entry 82
ID/s: 799	ID/s: 817
Frequency: 1	Frequency: 1
Proposed action: magsawa	Proposed action: napupunta
Target: seminars	Target: lugar
Entry 75	Entry 83
ID/s: 807	ID/s: 841
Frequency: 1	Frequency: 1
Proposed action: delay	Proposed action: help
Target: cause	Target: garbagetrash
Entry 76	Entry 84
ID/s: 809	

ID/s: 843
Frequency: 1
Proposed action: prevent
Target: blockage

EARLY WARNING SYSTEM

Entry 1

ID/s:
58|90|204|306|436|448|495|529|554|821|891|894
Frequency: 16
Proposed action: dapat
Target: sakuna, eg, komunikasyon, kita, mamamayan, pamamagitan, anunsyo, barangay (barangay), alert, balita, radio, abiso, saknila, update, my, baranggay, darating

Entry 2

ID/s:
112|125|137|139|140|146|175|235|292|309|575|619|859|909
Frequency: 15
Proposed action: be
Target: training, plenty, community, beforehand, incoming, disaster, time, typhoon, constituents, devices, need, news, happenings, events

Entry 3

ID/s:
30|90|128|189|223|309|313|354|547|619|782|837|844|874
Frequency: 14

Proposed action: pagiinform, ma-inform, inform, iinform
Target: chairman, dilubyo, citizens, typhoon, preparedness, disaster, residents, announce, public, tao, community, na, earlier

Entry 4

ID/s:
20|263|450|481|712|774|814|833|857|914
Frequency: 13
Proposed action: may

Target: kalamidad (calamidad), sakunang (sakuna), paalam, bagyo (bagyono), assembly, iparating, baha, anu, disaster, komunikasyon, suporta, pra

Entry 5

ID/s: 3|6|7|8|11|20|27|50|66|467|550|555
Frequency: 12
Proposed action: magkaroon, pagkakaroon
Target: komunikasyon, early, warning, system, maintenance, quarterly, council, members, tao, speaker, notice, device

Entry 6

ID/s: 58|436|438|535|558|685|788
Frequency: 7
Proposed action: mag
Target: coordinate, signal, warning, detalye, report, announce, bagyo

Entry 7

ID/s: 77|335|413|749|774|789|894
Frequency: 7
Proposed action: magbibigay, nagbbigay, magbigay, ipagbigay, mgbibigay
Target: babala, impormasyon, warning, barangay, signal

Entry 8

ID/s: 6|7|128|419|557|833
Frequency: 6
Proposed action: maging, messaging, magiging
Target: tao, paglaki, kalamidad, media, etc, aware

Entry 9

ID/s: 67|263|697|727|765|837
Frequency: 6
Proposed action: paparating, darating, iparating
Target: sakuna, kalamidad, na, bigay, lugar, s

Entry 10

ID/s: 112|140|176|224|300|313
Frequency: 6
Proposed action: have

Target: technology, warning, sign, disaster, sms, alert, house, megaphones	Proposed action: is
	Target: disaster, theeveryone, community
Entry 11	Entry 18
ID/s: 137 313 619 664 921	ID/s: 6 65 554
Frequency: 6	Frequency: 3
Proposed action: prepare, prepared	Proposed action: dumating
Target: community, disaster, things, alarm, weeks, way	Target: bagyo, kalamidad
Entry 12	Entry 19
ID/s: 278 697 727 772 856 914	ID/s: 102 105 112
Frequency: 6	Frequency: 3
Proposed action: maagang, maaga, maagap	Proposed action: do
Target: announcement, if, ever, impormasyon, kalamidad, anunsyo, palang	Target: announcements, discipline
Entry 13	Entry 20
ID/s: 111 291 300 373 575	ID/s: 103 112 133
Frequency: 5	Frequency: 3
Proposed action: make	Proposed action: improving, improved, providing
Target: announcement (anouncements), people, everyone, warning	Target: imformation, dissemination, communication, safety, places
Entry 14	Entry 21
ID/s: 272 273 573 671 686	ID/s: 264 519 619
Frequency: 5	Frequency: 3
Proposed action: alert, alerting	Proposed action: updated, update
Target: people, citizens, everyone, residents	Target: palagi, weather, changes, people
Entry 15	Entry 22
ID/s: 66 139 176 237	ID/s: 272 708
Frequency: 4	Frequency: 3
Proposed action: maiaannounce, announcing	Proposed action: giving
Target: balita, time, people, news	Target: announcements, newsinformation, beforehand, updates
Entry 16	Entry 23
ID/s: 137 250 573	ID/s: 788 821 844
Frequency: 4	Frequency: 3
Proposed action: coming, incoming, upcoming	Proposed action: nasasakupan
Target: disaster (disasters), floods	Target: announce, kalamidad, tulong
Entry 17	Entry 24
ID/s: 224 290 300 909	ID/s: 58 547
Frequency: 4	Frequency: 2
	Proposed action: tamang
	Target: panahon, tulong

Entry 25	Target: barangay, community, person
ID/s: 58	Entry 33
Frequency: 2	
Proposed action: paglikas	ID/s: 291 309
Target: pamilya, gamit	Frequency: 2
Entry 26	Proposed action: are
ID/s: 75 448	Target: disaster, calamities
Frequency: 2	Entry 34
Proposed action: pagtawag, magtawag	
Target: anak, baranggay	ID/s: 300 304
Entry 27	Frequency: 2
ID/s: 75 814	Proposed action: suggest
Frequency: 2	Target: barangay, brgy
Proposed action: kailangan	Entry 35
Target: tahanan, better, way	
Entry 28	ID/s: 334 814
ID/s: 77 473	Frequency: 2
Frequency: 2	Proposed action: spread, ma-spread
Proposed action: mabigyan, bigyang	Target: news, barangay
Target: impormasyon, babala	Entry 36
Entry 29	
ID/s: 97 388	ID/s: 413 419
Frequency: 2	Frequency: 2
Proposed action: regarding	Proposed action: wang
Target: disaster (disasters)	Target: wang
Entry 30	Entry 37
ID/s: 102 105	
Frequency: 2	ID/s: 413 419
Proposed action: roam	Frequency: 2
Target: barangay	Proposed action: ipagpatuluy
Entry 31	Target: barangay, pagbigay, babala
ID/s: 164 319	Entry 38
Frequency: 2	
Proposed action: galingan, dealing	ID/s: 431 445
Target: gid, disaster	Frequency: 2
Entry 32	Proposed action: ialarm, magalarm
ID/s: 189 277	Target: barangay, sirena
Frequency: 2	Entry 39
Proposed action: texting	
	ID/s: 445 727
	Frequency: 2
	Proposed action: malaki, malakas
	Target: tubig, n, bagyo
	Entry 40
	ID/s: 547 724
	Frequency: 2

Proposed action: lumikas	Frequency: 1
Target: sakuna, bgyan, talang	Proposed action: mapalaganap
Entry 41	Entry 49
ID/s: 558 844	Target: informasyon
Frequency: 2	
Proposed action: makatulong, makakatulong	
Target: bahaa, barangay	
Entry 42	Entry 50
ID/s: 857 917	Target: koordinasyon
Frequency: 2	
Proposed action: magprovide, provide	
Target: early, warning, emergency, exits	
Entry 43	Entry 51
ID/s: 3	Target: barangay
Frequency: 1	
Proposed action: magkikita	
Target: panahon	
Entry 44	Entry 52
ID/s: 8	Target: pasabi
Frequency: 1	
Proposed action: pagtibayin	
Target: early, warning, system, device	
Entry 45	Entry 53
ID/s: 8	Target: signal
Frequency: 1	
Proposed action: masigurong	
Target: kalamidad	
Entry 46	Entry 54
ID/s: 20	Target: theft
Frequency: 1	
Proposed action: makita	
Target: sitwasyon	
Entry 47	Entry 55
ID/s: 29	Target: troubles
Frequency: 1	
Proposed action: bahay-bahay	
Target: pag-inform	
Entry 48	Entry 56
ID/s: 30	Target: people

ID/s: 176	Entry 64
Frequency: 1	
Proposed action: built	
Target: speakers	
	Entry 57
ID/s: 176	
Frequency: 1	
Proposed action: drive	
Target: barangay, days	
	Entry 58
ID/s: 257	
Frequency: 1	
Proposed action: nagkulang	
Target: pagbibigay, babala	
	Entry 59
ID/s: 263	
Frequency: 1	
Proposed action: inaasahan	
Target: anunsyo	
	Entry 60
ID/s: 286	
Frequency: 1	
Proposed action: maitutulong	
Target: barangay	
	Entry 61
ID/s: 286	
Frequency: 1	
Proposed action: ipamahagi	
Target: impormasyon	
	Entry 62
ID/s: 286	
Frequency: 1	
Proposed action: mangyaring	
Target: sakuna	
	Entry 63
ID/s: 300	
Frequency: 1	
Proposed action: has	
Target: food	
	Entry 64
ID/s: 300	
Frequency: 1	
Proposed action: eat	
Target: damage	
	Entry 65
ID/s: 300	
Frequency: 1	
Proposed action: caused	
Target: typhoon	
	Entry 66
ID/s: 304	
Frequency: 1	
Proposed action: visit	
Target: house	
	Entry 67
ID/s: 306	
Frequency: 1	
Proposed action: gagawin	
Target: disaster	
	Entry 68
ID/s: 313	
Frequency: 1	
Proposed action: accommodate	
Target: barangay	
	Entry 69
ID/s: 313	
Frequency: 1	
Proposed action: allow	
Target: residents	
	Entry 70
ID/s: 319	
Frequency: 1	
Proposed action: participating	
Target: needs	
	Entry 71
ID/s: 413	
Frequency: 1	
Proposed action: kpag	

Target: tubig	Proposed action: mailikas
Entry 72	Entry 80
ID/s: 431	ID/s: 535
Frequency: 1	Frequency: 1
Proposed action: namemegaphone	Proposed action: sabihin
Target: ilog	Target: bagyong
Entry 73	Entry 81
ID/s: 431	ID/s: 550
Frequency: 1	Frequency: 1
Proposed action: ingat	Proposed action: malaman
Target: level	Target: pangangailangan
Entry 74	Entry 82
ID/s: 448	ID/s: 554
Frequency: 1	Frequency: 1
Proposed action: handa	Proposed action: mapaghandaan
Target: talad	Target: comunidad
Entry 75	Entry 83
ID/s: 481	ID/s: 557
Frequency: 1	Frequency: 1
Proposed action: kapit	Proposed action: makipag
Target: bahay	Target: ugnayan
Entry 76	Entry 84
ID/s: 490	ID/s: 571
Frequency: 1	Frequency: 1
Proposed action: paqaabiso	Proposed action: facebook
Target: paqdating	Target: twitter
Entry 77	Entry 85
ID/s: 490	ID/s: 573
Frequency: 1	Frequency: 1
Proposed action: lagi	Proposed action: give
Target: atpagbigay	Target: tips
Entry 78	Entry 86
ID/s: 529	ID/s: 619
Frequency: 1	Frequency: 1
Proposed action: makausap	Proposed action: evacuate
Target: tulong	Target: area
Entry 79	Entry 87
ID/s: 529	ID/s: 683
Frequency: 1	

Frequency: 1	ID/s: 712
Proposed action: knowing	Frequency: 1
Target: disaster	Proposed action: matigas
Entry 88	Target: ulo
ID/s: 684	ID/s: 713
Frequency: 1	Frequency: 1
Proposed action: avoid	Proposed action: makinig
Target: loss	Target: advisory
Entry 89	Entry 96
ID/s: 685	ID/s: 724
Frequency: 1	Frequency: 1
Proposed action: aga	Proposed action: nag
Target: respondihan	Target: anouse, merng
Entry 90	Entry 97
ID/s: 698	ID/s: 724
Frequency: 1	Frequency: 1
Proposed action: walang	Proposed action: umiikot
Target: sawang	Target: tao
Entry 91	Entry 98
ID/s: 698	ID/s: 724
Frequency: 1	Frequency: 1
Proposed action: pagpapa-alala	Proposed action: mababang
Target: sakuna	Target: lugar
Entry 92	Entry 99
ID/s: 708	ID/s: 724
Frequency: 1	Frequency: 1
Proposed action: said	Proposed action: masabihan
Target: disaster	Target: tao, n
Entry 93	Entry 100
ID/s: 710	ID/s: 724
Frequency: 1	Frequency: 1
Proposed action: utilize	Proposed action: tumira
Target: media	Target: evacuoation, center
Entry 94	Entry 101
ID/s: 710	ID/s: 724
Frequency: 1	Frequency: 1
Proposed action: disseminate	Proposed action: ipabatid
Target: awareness	Target: balita
Entry 95	Entry 102

<p>Entry 103</p> <p>ID/s: 788 Frequency: 1 Proposed action: pupunta Target: baha</p>	<p>ID/s: 16 19 122 341 365 439 447 502 616 626 6 60 690 714 728 731 747 748 753 783 785 786 806 815 824 830 889 Frequency: 26 Proposed action: maging, palaging, pagiging, magiging Target: bagay, bagyobaha, oras, sakuna, pra, alertoat, handa, atentibo, balita, daratinga (darating), cla, brgy, pkiking, kalamidad (kalamid), kalikasan, kagamatin, opisyal, kasanyan, xa, prepared</p>
<p>Entry 104</p> <p>ID/s: 789 Frequency: 1 Proposed action: papalapit Target: bagyo, baranggay</p>	<p>Entry 105</p> <p>ID/s: 789 Frequency: 1 Proposed action: mataas Target: lugar</p>
<p>Entry 106</p> <p>ID/s: 833 Frequency: 1 Proposed action: epetibo Target: kng</p>	<p>Entry 107</p> <p>ID/s: 859 Frequency: 1 Proposed action: watch Target: news</p>
<p>Entry 108</p> <p>ID/s: 887 Frequency: 1 Proposed action: mabilis Target: mensahe</p>	<p>Entry 109</p> <p>ID/s: 914 Frequency: 1 Proposed action: naapektuhan Target: bagyo</p>
PREPAREDNESS FOR EMERGENCY	
Entry 1	
<p>Extracting and Organizing Disaster-related Philippine Community Responses for Aiding Nationwide Risk Reduction Planning and Response (N. Nocon, 2020)</p>	

Entry 2

ID/s:
73|148|151|203|211|341|342|506|510|612|
673|767|793|850|851|890|892|896|899|900|
901|922
Frequency: 24
Proposed action: dapat, sapat
Target: barangay (baranggay), disaster,
balita, aware, oras, updated, araw, alirto,
official, membro, i, ulat, pra, programa
(program), sakuna, pagkain, pangangailangan

Entry 3

ID/s:
123|145|165|199|233|248|755|816|863|868
Frequency: 10
Proposed action: be
Target: news, officials, happenings,
evacuation, plan, things, flood, disaster,
management, unit, community

Entry 4

ID/s: 61|122|365|492|506|510|613|748|753
Frequency: 9
Proposed action: may
Target: sakuna (sakunang), bagyo,
kapaligiran, sakona, pomoang, barangay
(kabaranggay), nkalagay, darating

Entry 5

ID/s: 36|61|642|722|726|742|899|908
Frequency: 8
Proposed action: ihanda, maghanda, handa

Target: pweding, gamit, sakuna, anu, barangay, bagay, mangyari, karapat, pagkain	Frequency: 3
Entry 6	Proposed action: is
ID/s: 37 56 528 533 562 803 818 824	Target: week, solution, times
Frequency: 8	Entry 13
Proposed action: kaylangan, kailangan, nangangailangang	ID/s: 203 722 818
Target: aware, mamamayan, radyo, ata, always, alert, disciplina, mangyare, interaksyon	Frequency: 3
Entry 7	Proposed action: nangyayari, mangyare
ID/s: 315 452 463 613 624 729 850	Target: paligid, darateng, bagyo
Frequency: 7	Entry 14
Proposed action: mag	ID/s: 248 755 840
Target: food, supplies, kagamitang, monitor, samahan, plano, laan, gamot, imbak	Frequency: 3
Entry 8	Proposed action: help, helps
ID/s: 23 56 533 818	Target: time, calamity, community
Frequency: 5	Entry 15
Proposed action: dumating	ID/s: 72 122
Target: sakuna, kalamidad, obserbatibo	Frequency: 2
Entry 9	Proposed action: malaman
ID/s: 312 690 729 890 912	Target: panahon, kalagayan
Frequency: 5	Entry 16
Proposed action: alert, alerto	ID/s: 73 769
Target: calamity, bagay, kalamidada, tv, darating	Frequency: 2
Entry 10	Proposed action: wala
ID/s: 35 321 341 752	Target: kalamidad
Frequency: 4	Entry 17
Proposed action: magkaroon	ID/s: 248 864
Target: participation, sariling, disciplina, uusap, emergency, plan	Frequency: 2
Entry 11	Proposed action: comes
ID/s: 159 228 284	Target: time, funds
Frequency: 3	Entry 18
Proposed action: watch, watching	ID/s: 312 672
Target: weather, forecast, news, television	Frequency: 2
Entry 12	Proposed action: prepare
ID/s: 159 324 738	Target: disaster, needs
	Entry 19
	ID/s: 622 626
	Frequency: 2
	Proposed action: makinig
	Target: anunyo, balita
	Entry 20

ID/s: 623 915	Target: brgy
Frequency: 2	Entry 28
Proposed action: canned	
Target: food, batteries, goods	
Entry 21	
ID/s: 626 752	ID/s: 22
Frequency: 2	Frequency: 1
Proposed action: mapagbigay-alam, ipagbigay	Proposed action: maayos
Target: kapit-bahay, mamamayan	Target: paghahanda
Entry 22	Entry 29
ID/s: 748 908	ID/s: 23
Frequency: 2	Frequency: 1
Proposed action: mabilis	Proposed action: maagang
Target: bayah, sakuna	Target: paghahanda
Entry 23	Entry 30
ID/s: 806 881	ID/s: 36
Frequency: 2	Frequency: 1
Proposed action: mapaghandaan, paghahanda	Proposed action: magamit
Target: tulong, kalamidad	Target: comunicasion
Entry 24	Entry 31
ID/s: 896 924	ID/s: 72
Frequency: 2	Frequency: 1
Proposed action: nagaganap, laganap	Proposed action: nakatuon
Target: barangay, kurapsyon	Target: weather, forecast
Entry 25	Entry 32
ID/s: 10	ID/s: 99
Frequency: 1	Frequency: 1
Proposed action: sinasabi	Proposed action: doing
Target: sakuna	Target: flood, prone, areas
Entry 26	Entry 33
ID/s: 16	ID/s: 158
Frequency: 1	Frequency: 1
Proposed action: magtulong	Proposed action: check
Target: tulong	Target: typhoon
Entry 27	Entry 34
ID/s: 16	ID/s: 159
Frequency: 1	Frequency: 1
Proposed action: opisyal	Proposed action: know
	Target: weather
	Entry 35
	ID/s: 165
	Frequency: 1

Proposed action: listen	Frequency: 1
Target: radio	Proposed action: pwede
Entry 36	Entry 44
ID/s: 165	ID/s: 380
Frequency: 1	Frequency: 1
Proposed action: regarding	Proposed action: being
Target: weather, condition	Target: disaster, preparedness
Entry 37	Entry 45
ID/s: 211	ID/s: 447
Frequency: 1	Frequency: 1
Proposed action: magpadala	Proposed action: makipagtutulungan
Target: tao	Target: barangay
Entry 38	Entry 46
ID/s: 211	ID/s: 452
Frequency: 1	Frequency: 1
Proposed action: tignan	Proposed action: binabaha
Target: bagyo	Target: tao
Entry 39	Entry 47
ID/s: 280	ID/s: 492
Frequency: 1	Frequency: 1
Proposed action: bumuo	Proposed action: kumplitohin
Target: emergency, plansan	Target: gamit
Entry 40	Entry 48
ID/s: 280	ID/s: 506
Frequency: 1	Frequency: 1
Proposed action: pupunta	Proposed action: matugunan
Target: oras	Target: pangangailangan, kaligtasan
Entry 41	Entry 49
ID/s: 280	ID/s: 528
Frequency: 1	Frequency: 1
Proposed action: dadaan	Proposed action: pagdating
Target: barangay	Target: kalamidad
Entry 42	Entry 50
ID/s: 284	ID/s: 532
Frequency: 1	Frequency: 1
Proposed action: radio	Proposed action: pagkilos
Target: disaster	Target: kalamidad
Entry 43	Entry 51
ID/s: 315	

ID/s: 533	Entry 59
Frequency: 1	
Proposed action: ilagay	
Target: bgo	
	Entry 52
ID/s: 544	
Frequency: 1	
Proposed action: pumunta	
Target: safe	
	Entry 53
ID/s: 616	
Frequency: 1	
Proposed action: iimbak	
Target: pagkain	
	Entry 54
ID/s: 624	
Frequency: 1	
Proposed action: kakailanganin	
Target: panahon	
	Entry 55
ID/s: 625	
Frequency: 1	
Proposed action: med	
Target: kits	
	Entry 56
ID/s: 652	
Frequency: 1	
Proposed action: tumulong	
Target: pagkalap	
	Entry 57
ID/s: 673	
Frequency: 1	
Proposed action: maka	
Target: alis	
	Entry 58
ID/s: 680	
Frequency: 1	
Proposed action: inform	
Target: barangay	
	Entry 59
ID/s: 680	
Frequency: 1	
Proposed action: need	
Target: proper	
	Entry 60
ID/s: 680	
Frequency: 1	
Proposed action: go	
Target: case	
	Entry 61
ID/s: 704	
Frequency: 1	
Proposed action: have	
Target: disaster, risk, reduction, management, council	
	Entry 62
ID/s: 731	
Frequency: 1	
Proposed action: ibahagi	
Target: nl	
	Entry 63
ID/s: 731	
Frequency: 1	
Proposed action: ibalita	
Target: s	
	Entry 64
ID/s: 748	
Frequency: 1	
Proposed action: magki2ta	
Target: oras	
	Entry 65
ID/s: 748	
Frequency: 1	
Proposed action: maiwasan	
Target: pagka, trap	
	Entry 66
ID/s: 756	
Frequency: 1	

Proposed action: alisto	Frequency: 1
Target: n, balita	Proposed action: meron
Entry 67	Entry 75
ID/s: 756	
Frequency: 1	
Proposed action: ipahatid	
Target: mamayan	
Entry 68	Entry 76
ID/s: 756	
Frequency: 1	
Proposed action: gagawin	
Target: my	
Entry 69	Entry 77
ID/s: 769	
Frequency: 1	
Proposed action: sakin	
Target: kana	
Entry 70	Entry 78
ID/s: 775	
Frequency: 1	
Proposed action: mahalaga	
Target: stocks	
Entry 71	Entry 79
ID/s: 786	
Frequency: 1	
Proposed action: makaiwas	
Target: sakuna	
Entry 72	Entry 80
ID/s: 793	
Frequency: 1	
Proposed action: nasasakupan	
Target: problemang	
Entry 73	Entry 81
ID/s: 816	
Frequency: 1	
Proposed action: lessen	
Target: scale	
Entry 74	Entry 82
ID/s: 892	

ID/s: 924	Target: sakuna, tagapamahala, kapansanan
Frequency: 1	Entry 5
Proposed action: simulan	
Target: baranggay	
Entry 83	
ID/s: 924	ID/s: 457 471
Frequency: 1	Frequency: 3
Proposed action: malaking	Proposed action: maging
Target: problema	Target: barangay, tagapag, tao
Entry 84	Entry 6
ID/s: 924	ID/s: 96 701
Frequency: 1	Frequency: 2
Proposed action: naghihirap	Proposed action: throwing, knowing
Target: bansa (dahilbansa)	Target: trash, disaster
<hr/>	
LOCAL GOVERNMENT ACCOUNTABILITY	
<hr/>	
Entry 1	Entry 7
ID/s: 153 285 318 371 928 934	ID/s: 153 201
Frequency: 7	Frequency: 2
Proposed action: be	Proposed action: are
Target: disasters, evacuation, government, officials, posters, case, society	Target: help, people
Entry 2	Entry 8
ID/s: 457 518 770	ID/s: 201 318
Frequency: 4	Frequency: 2
Proposed action: kailangan, nangaylangan	Proposed action: happen
Target: kahandaan, mgsma, sma, mg, usap, nmin, hrpin, maiparting, dn, tulong	Target: community, news
Entry 3	Entry 9
ID/s: 28 643 732	ID/s: 210 353
Frequency: 3	Frequency: 2
Proposed action: dapat	Proposed action: have
Target: active, ugnayan, barangay	Target: map, incase
Entry 4	Entry 10
ID/s: 64 471 472	ID/s: 471 904
Frequency: 3	Frequency: 2
Proposed action: may	Proposed action: nag
	Target: pamagitna, kalaminad
Entry 11	Entry 11
ID/s: 835 913	ID/s: 471 904
Frequency: 2	Frequency: 2
Proposed action: mag	Proposed action: mag
	Target: mg, pundo
Entry 12	Entry 12
ID/s: 28	ID/s: 28
Frequency: 1	Frequency: 1

Proposed action: mabuti	Frequency: 1
Target: barangay, problema	Proposed action: enforce
Entry 13	Entry 21
ID/s: 28	Target: laws
Frequency: 1	
Proposed action: maari	
Target: performance	
Entry 14	Entry 22
ID/s: 28	Target: place
Frequency: 1	
Proposed action: maipaabot	
Target: serbisyo	
Entry 15	Entry 23
ID/s: 64	Target: times
Frequency: 1	
Proposed action: ayusin	
Target: patakaran	
Entry 16	Entry 24
ID/s: 64	Target: disasters
Frequency: 1	
Proposed action: mabilis	
Target: pagbibigay	
Entry 17	Entry 25
ID/s: 76	Target: awareness
Frequency: 1	
Proposed action: gawin	
Target: barangay	
Entry 18	Entry 26
ID/s: 82	Target: barangay
Frequency: 1	
Proposed action: gumawa	
Target: organisasyon	
Entry 19	Entry 27
ID/s: 82	Target: watcg
Frequency: 1	
Proposed action: magtutulongan	
Target: paglilinis	
Entry 20	Entry 28
ID/s: 96	Target: people

ID/s: 201	Entry 36
Frequency: 1	
Proposed action: make	
Target: officials	
	Entry 29
ID/s: 201	
Frequency: 1	
Proposed action: surrounds	
Target: one	
	Entry 30
ID/s: 201	
Frequency: 1	
Proposed action: hurt	
Target: typhoon	
	Entry 31
ID/s: 202	
Frequency: 1	
Proposed action: respond	
Target: emergency	
	Entry 32
ID/s: 210	
Frequency: 1	
Proposed action: indicates	
Target: altitude	
	Entry 33
ID/s: 210	
Frequency: 1	
Proposed action: know	
Target: part	
	Entry 34
ID/s: 268	
Frequency: 1	
Proposed action: having	
Target: personnel	
	Entry 35
ID/s: 268	
Frequency: 1	
Proposed action: assigned	
Target: news	
	Entry 36
ID/s: 371	
Frequency: 1	
Proposed action: feed	
Target: people	
	Entry 37
ID/s: 441	
Frequency: 1	
Proposed action: sana	
Target: kalamidad	
	Entry 38
ID/s: 441	
Frequency: 1	
Proposed action: puntahan	
Target: opisyal	
	Entry 39
ID/s: 441	
Frequency: 1	
Proposed action: malaman	
Target: lugar	
	Entry 40
ID/s: 451	
Frequency: 1	
Proposed action: maagap	
Target: pagtulong	
	Entry 41
ID/s: 471	
Frequency: 1	
Proposed action: paman	
Target: tao	
	Entry 42
ID/s: 471	
Frequency: 1	
Proposed action: ligtas	
Target: paraan	
	Entry 43
ID/s: 471	
Frequency: 1	
Proposed action: maibahagi	

Target: tao	Proposed action: utilizing
Entry 44	Entry 52
ID/s: 472	ID/s: 870
Frequency: 1	Frequency: 1
Proposed action: bigyan	Proposed action: nakkatulong
Target: pansin	Target: tao, regarding
Entry 45	Entry 53
ID/s: 472	ID/s: 870
Frequency: 1	Frequency: 1
Proposed action: malapit	Proposed action: paparating
Target: tabing	Target: sakuna
Entry 46	Entry 54
ID/s: 477	ID/s: 904
Frequency: 1	Frequency: 1
Proposed action: andyan	Proposed action: punong
Target: bagyo	Target: brgy
Entry 47	Entry 55
ID/s: 577	ID/s: 904
Frequency: 1	Frequency: 1
Proposed action: facilitate	Proposed action: manguna
Target: disaster, response	Target: palaganap
Entry 48	Entry 56
ID/s: 643	ID/s: 904
Frequency: 1	Frequency: 1
Proposed action: maiwasan	Proposed action: mayrong
Target: sakuna	Target: kalamidad
Entry 49	Entry 57
ID/s: 699	ID/s: 904
Frequency: 1	Frequency: 1
Proposed action: pwersahan	Proposed action: dumating
Target: tao	Target: kuminidad
Entry 50	
ID/s: 770	
Frequency: 1	
Proposed action: mlman	
Target: dpt, gwen	
Entry 51	
ID/s: 845	
Frequency: 1	
	FILIPINO VALUES
	Entry 1
	ID/s:
	55 161 433 443 523 531 641 691 694 723
	768 832

Frequency: 12	ID/s: 70 410 415 563 568
Proposed action: tutulong, magtulong2, makinatalong, magtulong, tulongan, maitutulong, tutulongan, magtulongan, makakatulong	Frequency: 5
Target: paghahanda, baranggay (barangay), kalamidad, komunidad, tulong, esatesa, magkakapamilya, hnda	Proposed action: kailangan, nangangailangan, kaylangan
	Target: ka-barangay, tulong, oras, atensyon, isat
Entry 2	Entry 8
ID/s: 2 251 433 459 551 552 641 666 759	ID/s: 31 39 757 888
Frequency: 10	Frequency: 4
Proposed action: mag	Proposed action: dapat
Target: pagkakaiisa, malasakit, samahan, volunteer, barangay, duty, peace, and, order, gaeing, my, kaisa, tahanan, suporta, tulong	Target: pakiki-isa, komunidad, panahon, mgkaisa
Entry 3	Entry 9
ID/s: 21 49 423 525 540 564	ID/s: 41 443 621 721
Frequency: 7	Frequency: 4
Proposed action: tumulong	Proposed action: maging
Target: bahay, pagkain, bagyo, community, volunteer, bata, tulong, barangay, pra	Target: darating, kalagayan, agrisibo, kalamidad
Entry 4	Entry 10
ID/s: 422 486 498 503 567 570 898	ID/s: 367 428 715
Frequency: 7	Frequency: 4
Proposed action: makipag, makikipag, makipagtulungan	Proposed action: matulungan, magtulungan
Target: barangay (barangay), tao, tulunganat	Target: kabbarangay, kalamidad, s, tao
Entry 5	Entry 11
ID/s: 1 17 63 427 508 832	ID/s: 412 520 858
Frequency: 6	Frequency: 3
Proposed action: magkaisa, nagkakaisa	Proposed action: mali, malinis
Target: tao, tungkulin, plano, paghahanda, ren, lng	Target: kelangan, kpalogirantumulong (kpalogiran)
Entry 6	Entry 12
ID/s: 31 426 715 832	ID/s: 168 262
Frequency: 5	Frequency: 2
Proposed action: may	Proposed action: being
Target: sakunang, bagyo, kalamidadwag (kalamidad)	Target: disaster, share, volunteer
Entry 7	Entry 13
	ID/s: 168 861
	Frequency: 2
	Proposed action: have
	Target: announcements, heart
	Entry 14

ID/s: 367	Entry 22
Frequency: 2	
Proposed action: handa	
Target: volunteer, canal	
Entry 15	
ID/s: 538 666	Entry 23
Frequency: 2	
Proposed action: nakarating, karatig	
Target: kalamidad, bayan	
Entry 16	
ID/s: 9	Entry 24
Frequency: 1	
Proposed action: lumawat	
Target: pagmamalasakit	
Entry 17	
ID/s: 21	Entry 25
Frequency: 1	
Proposed action: napinsala	
Target: bagyo	
Entry 18	
ID/s: 31	
Frequency: 1	
Proposed action: dumating	
Target: kalamidad	
Entry 19	
ID/s: 44	Entry 26
Frequency: 1	
Proposed action: nagkakaroon	
Target: lugar	
Entry 20	
ID/s: 53	Entry 27
Frequency: 1	
Proposed action: makiisa	
Target: komunidad	
Entry 21	
ID/s: 260	Entry 28
Frequency: 1	
Proposed action: matutong	
Target: malasakit	
Entry 29	
ID/s: 410	
Frequency: 1	
Proposed action: pantay-pantay	
Target: pamamahagi	
Entry 29	
ID/s: 410	
Frequency: 1	
Proposed action: mabigyang	

Target: solusyon	Proposed action: magbuo
Entry 30	Entry 38
ID/s: 411	
Frequency: 1	
Proposed action: magmalasakit	
Target: kawpa	
Entry 31	Entry 39
ID/s: 412	
Frequency: 1	
Proposed action: wala	
Target: barangay	
Entry 32	Entry 40
ID/s: 430	
Frequency: 1	
Proposed action: ibigay	
Target: tulong	
Entry 33	Entry 41
ID/s: 443	
Frequency: 1	
Proposed action: naperwisyo	
Target: baha	
Entry 34	Entry 42
ID/s: 443	
Frequency: 1	
Proposed action: umasa	
Target: barangaymas	
Entry 35	Entry 43
ID/s: 443	
Frequency: 1	
Proposed action: gawing	
Target: serbisyo, gsyia	
Entry 36	Entry 44
ID/s: 453	
Frequency: 1	
Proposed action: mkikipgtolongan	
Target: programa	
Entry 37	Entry 45
ID/s: 459	
Frequency: 1	
	Entry 45
	ID/s: 691

Frequency: 1	ID/s: 759	
Proposed action: maitayo	Frequency: 1	
Target: kagamitan	Proposed action: maiwasan	
Entry 46		
ID/s: 715	Target: baha	
Frequency: 1	Entry 54	
Proposed action: magtaponi	ID/s: 861	
Target: basura	Frequency: 1	
Entry 47		
ID/s: 720	Proposed action: be	
Frequency: 1	Target: barangay	
Proposed action: magsama	<hr/>	
Target: sama	OTHERS	
Entry 48		
ID/s: 730	Entry 1	
Frequency: 1	ID/s: 455 791 902 923	
Proposed action: maka	Frequency: 4	
Target: sakuna	Proposed action: dapat	
Entry 49		
ID/s: 740	Target: tao, barangay	
Frequency: 1	Entry 2	
Proposed action: tumolng	ID/s: 314 791 826	
Target: n	Frequency: 3	
Entry 50		
ID/s: 740	Proposed action: wala, walang	
Frequency: 1	Target: pakielam, kwenta (kwentang), bagay	
Proposed action: naabutan	Entry 3	
Target: baha	ID/s: 115 118	
Entry 51		
ID/s: 750	Frequency: 2	
Frequency: 1	Proposed action: putting	
Proposed action: makapaghanda	Target: pockets	
Target: s	Entry 4	
Entry 52		
ID/s: 759	ID/s: 184	
Frequency: 1	Frequency: 2	
Proposed action: kalapit	Proposed action: communicate	
Target: bahay	Target: deaf, citizens, responders	
Entry 53		
ID/s: 759	Entry 5	
Frequency: 1	ID/s: 185 192	
Proposed action: kalapit	Frequency: 2	
Target: bahay	Proposed action: is	
Entry 6		
ID/s: 187 197	Target: question, prepare	

Frequency: 2	ID/s: 184
Proposed action: help	Frequency: 1
Target: teach, people	Proposed action: cannot
Entry 7	
ID/s: 192 195	Target: sign
Frequency: 2	ID/s: 188
Proposed action: be	Frequency: 1
Target: aware	Proposed action: has
Entry 8	
ID/s: 479 802	Target: help
Frequency: 2	ID/s: 188
Proposed action: mag	Frequency: 1
Target: kona, tao	Proposed action: please
Entry 9	
ID/s: 501 542	Target: people
Frequency: 2	ID/s: 191
Proposed action: ipagpatuloy	Frequency: 1
Target: barangay, gawain	Proposed action: pls
Entry 10	
ID/s: 513 611	Target: i
Frequency: 2	ID/s: 191
Proposed action: maging	Frequency: 1
Target: pamaraan, agrisibo	Proposed action: are
Entry 11	
ID/s: 637 791	Target: neighborhoods
Frequency: 2	ID/s: 192
Proposed action: sana, sanang	Frequency: 1
Target: negosyo, dhil, khit	Proposed action: calling
Entry 12	
ID/s: 183	Target: barangay, community
Frequency: 1	ID/s: 193
Proposed action: seen	Frequency: 1
Target: typhoon, disaster	Proposed action: suggest
Entry 13	
ID/s: 184	Target: typhoon, amd, flood
Frequency: 1	ID/s: 194
Proposed action: know	Frequency: 1
Target: deaf, train, emergency, responders	Proposed action: want
Entry 14	
	Target: barangay, help
Entry 15	
Entry 16	
Entry 17	
Entry 18	
Entry 19	
Entry 20	
Entry 21	

Entry 22	Target: locals, ulo
ID/s: 194	Entry 30
Frequency: 1	
Proposed action: better	ID/s: 425
Target: world	Frequency: 1
	Proposed action: humingi
	Target: tulung
Entry 23	Entry 31
ID/s: 195	
Frequency: 1	ID/s: 425
Proposed action: cant	Frequency: 1
Target: life	Proposed action: mkaiwas
	Target: baha
Entry 24	Entry 32
ID/s: 241	
Frequency: 1	ID/s: 455
Proposed action: give	Frequency: 1
Target: permission	Proposed action: sumunodkapag
	Target: warning, device
Entry 25	Entry 33
ID/s: 241	
Frequency: 1	ID/s: 501
Proposed action: let	Frequency: 1
Target: people	Proposed action: nakikita
	Target: q
Entry 26	Entry 34
ID/s: 241	
Frequency: 1	ID/s: 501
Proposed action: build	Frequency: 1
Target: houses	Proposed action: maganda
	Target: gawain
Entry 27	Entry 35
ID/s: 241	
Frequency: 1	ID/s: 521
Proposed action: prevent	Frequency: 1
Target: flashfloods	Proposed action: iboto
	Target: eleksyon
Entry 28	Entry 36
ID/s: 314	
Frequency: 1	ID/s: 539
Proposed action: magpalit	Frequency: 1
Target: brgy, chairman, kase	Proposed action: manghingi
	Target: tulong
Entry 29	Entry 37
ID/s: 356	
Frequency: 1	ID/s: 545
Proposed action: makinig	Frequency: 1

Proposed action: pagsasaayos	Frequency: 1
Target: qng, anu	Proposed action: nararapat
Entry 38	Entry 46
ID/s: 572	Target: pondo, pra
Frequency: 1	
Proposed action: minimize	
Target: corruption	
Entry 39	Entry 47
ID/s: 637	ID/s: 791
Frequency: 1	Frequency: 1
Proposed action: gumawa	Proposed action: maayos
Target: paraan	Target: kominidad
Entry 40	Entry 48
ID/s: 637	ID/s: 791
Frequency: 1	Frequency: 1
Proposed action: maka	Proposed action: maliit
Target: tulong	Target: pamilya
Entry 41	Entry 49
ID/s: 695	ID/s: 791
Frequency: 1	Frequency: 1
Proposed action: stay	Proposed action: apektado
Target: home	Target: tang, ina
Entry 42	Entry 50
ID/s: 695	ID/s: 826
Frequency: 1	Frequency: 1
Proposed action: avoid	Proposed action: gamitin
Target: object	Target: pondo
Entry 43	Entry 51
ID/s: 719	ID/s: 923
Frequency: 1	Frequency: 1
Proposed action: iabot	Proposed action: magbigay
Target: barangay	Target: impormasyon
Entry 44	Entry 52
ID/s: 719	ID/s: 923
Frequency: 1	Frequency: 1
Proposed action: maibigay	Proposed action: handa
Target: barangay	Target: tao
Entry 45	Entry 53
ID/s: 791	

ID/s: 923

Frequency: 1

Proposed action: may

Target: bagyo

D.2 Word2Vec, Organized by Response Categories

FILIPINO TEXT ANALYSIS TOOL REPORT

Mar-24-2020 20:47:25

The information below were extracted and organized automatically.

INFORMATION CAMPAIGN AND CAPACITY BUILDING

Entry 1

ID/s:

80|81|120|121|129|130|132|149|156|157|1
60|169|178|181|214|215|218|222|234|245|
249|256|259|266|269|283|287|297|303|305|
|331|337|340|343|347|358|382|385|389|39|
0|400|403|406|485|582|583|585|591|614|6
27|630|647|654|661|677|687|707|794|801|
831|838|869|893|895|929|933

Frequency: 95

Proposed action: be, put, do, been, have, is, think, give, help, know, avoid, having, go, tell, provide, living, make, are, create, contain, needed, using, please, making, being, add

Target: advantage, possibilities, calamities, dont, times (typhoon, disaster, place, training, community, case, news, household), time (effects, help, check, needs, class, drive, people, things, systems), drills, preparedness, programs (assembly, emergency, information, project, knowledge), officials, instructions, orientation, prepare, duty, sirens, disasters, barangay, drill, meeting, seminar, panicking, floods, families, representative, seminars, outcomes, taraining, idea, seminars, safety, tips, evacuation, ideas, risk, orrientation, lgu, majority, dissemination, signage, calamity, house, emergencies, drainage

Entry 2

ID/s:

13|26|208|209|289|339|469|491|507|576|5
80|598|628|649|717|735|744|762|766|771|
797|800|811|813|836|847|852|854|883|885|
|897|903

Frequency: 46

Proposed action: nais, dapat, malaman, gusto, sana, malalaman, kailangan, makakatulong, handa, maaari, nalaman, maganda, naisip, gagawin, maayos

Target: ngmga, barangay, seminar, tao, oras, volunteer, awareness, programs (for, example, about, weekly, center, before, during, and, after, local, government), prepared, kabbarangay, kaalaman (paraan), disiplina, kalamidad, sakuna, posters, paalala, pgdating, lubus, pamamagitan, drill, like (my), tulong, paaralan, evacuation, encourage, of, the, calamity, pagpapatupad, mg, roong, conduct, drmm

Entry 3

ID/s:

13|33|46|68|208|209|236|243|307|310|311|
339|598|618|631|651|696|716|758|760|762|
|766|787|800|854|885|903|925|926|930

Frequency: 34

Proposed action: maging, magkaroon, mabigyan

Target: kapitbahay, tao, disaster (and, during, about, program, emergency, training, weekly), mamayan, aware, kabbarangay, kabaro, kalamidad, darating, beforeafter, pra, bagyo, trahedy, seminars, kaalaman (ideya), drill, seminar, sos, regarding, iinvite, seminarsdrill, kits, progma, kaalam, disiplina, meeting

Entry 4

ID/s:

46|54|236|281|289|462|536|553|579|599|6
04|676|717|736|760|766|800|836|897|931

Frequency: 23

Proposed action: magbigay, mag, magturo, magsagawa, magtayo, talakayin, tamang, gumawa, maiwasan

Target: seminar, kaalaman (impormasyon), bahay, participate, first (aid, disaster, update),

conduct, drills, schedule, pagpupulong, orientation, seminars, grupo, mg, drill, programa, bagyo (oras), epektso, kaukuluan, trahedy, desaster, impormansyon

Entry 5

ID/s: 344|382|390|398|485|700|842|846|860

Frequency: 9

Proposed action: regarding

Target: prevention, disaster, awareness, preparation, disasters, incoming, calamities, preparations, preparedness

Entry 6

ID/s: 26|362|599|764|813|883|926

Frequency: 8

Proposed action: may, walang

Target: pangangailangan, seminar, meetings, kalamidad, conduct, kinalaman, darating, drill

Entry 7

ID/s: 152|200|216|594|627|798|876|927

Frequency: 8

Proposed action: giving

Target: disaster (knowledge, information), drill, seminars, leaflets

Entry 8

ID/s: 207|238|269|332|343|344|382|677

Frequency: 8

Proposed action: conducting

Target: seminars, drills, community (assembly), lot

Entry 9

ID/s: 101|104|337|347|350|386|405

Frequency: 7

Proposed action: inform

Target: consequence, people (community, kind), neighbor, subordinate

Entry 10

ID/s: 101|104|287|343|347

Frequency: 5

Proposed action: preparing

Target: typhoon (disaster)

Entry 11

ID/s: 169|333|396|583|644

Frequency: 5

Proposed action: prepare

Target: disaster (times), seminar

Entry 12

ID/s: 296|364|610|860|886

Frequency: 5

Proposed action: guide, planning, held, w

Target: community (program, level), incase, session, brgy

Entry 13

ID/s: 362|578|797|800

Frequency: 4

Proposed action: ginagawa, ginagamit, nakikita, nangyayari

Target: samin, babala, disiplina, s

Entry 14

ID/s: 15|592|736

Frequency: 3

Proposed action: seminar

Target: bagyo (lindol), disaster (information, drive)

Entry 15

ID/s: 84|88|846

Frequency: 3

Proposed action: organize

Target: events (disaster, talk), drills

Entry 16

ID/s: 598|618|631

Frequency: 3

Proposed action: magsasagawa

Target: seminar, seminars

Entry 17

ID/s: 5|310

Frequency: 2

Proposed action: malawakang, lumawak

Target: information (drive), kaalaman

Entry 18

ID/s: 13|930

Frequency: 2	ID/s: 469 903
Proposed action: ibat	Frequency: 2
Target: paraan, sona	Proposed action: maiging, importanteng
Entry 19	Entry 27
ID/s: 120 287	ID/s: 603 630
Frequency: 2	Frequency: 2
Proposed action: remind	Proposed action: drill
Target: citizens, community (members)	Target: dapat, facebook
Entry 20	Entry 28
ID/s: 130 157	ID/s: 801 831
Frequency: 2	Frequency: 2
Proposed action: gather	Proposed action: hit
Target: citizens, share, ideas	Target: cases, area
Entry 21	Entry 29
ID/s: 130 178	ID/s: 811 910
Frequency: 2	Frequency: 2
Proposed action: educate	Proposed action: pumunta, pumasok
Target: disaster (people)	Target: briefing, bagyo
Entry 22	Entry 30
ID/s: 132 842	ID/s: 26
Frequency: 2	Frequency: 1
Proposed action: join	Proposed action: sanayin
Target: seminar, barangay	Target: lugar
Entry 23	Entry 31
ID/s: 249 254	ID/s: 26
Frequency: 2	Frequency: 1
Proposed action: teach	Proposed action: pagsapit
Target: citizen, disaster	Target: bagyobaha
Entry 24	Entry 32
ID/s: 311 603	ID/s: 33
Frequency: 2	Frequency: 1
Proposed action: palaging, laging	Proposed action: mapaglalaban
Target: gawim, paalala	Target: sakuna
Entry 25	Entry 33
ID/s: 396 595	ID/s: 42
Frequency: 2	Frequency: 1
Proposed action: conduct	Proposed action: pagsabihan
Target: seminars	Target: kaalaman
Entry 26	

Entry 34	Target: damage
ID/s: 42	Entry 42
Frequency: 1	
Proposed action: nasasakupan	
Target: kalamidad	
Entry 35	
ID/s: 54	
Frequency: 1	
Proposed action: maglinis	
Target: kanal, harap	
Entry 36	Entry 43
ID/s: 80	
Frequency: 1	
Proposed action: door	
Target: disaster, awareness	
Entry 37	
ID/s: 80	Entry 44
Frequency: 1	
Proposed action: maximize	
Target: information, dissemination	
Entry 38	
ID/s: 81	
Frequency: 1	
Proposed action: alert	
Target: citizens	
Entry 39	Entry 45
ID/s: 84	
Frequency: 1	
Proposed action: concerning	
Target: disaster, preparedness	
Entry 40	
ID/s: 121	Entry 46
Frequency: 1	
Proposed action: implement	
Target: disaster, drills	
Entry 41	
ID/s: 121	Entry 47
Frequency: 1	
Proposed action: cause	
	Entry 48
	ID/s: 169
	Frequency: 1
	Proposed action: initiating
	Target: seminars
	Entry 49
	ID/s: 169
	Frequency: 1
	Proposed action: comes
	Target: people
	ID/s: 178
	Frequency: 1

Proposed action: throw	Frequency: 1
Target: garbages	Proposed action: handle
Entry 50	Entry 58
ID/s: 179	ID/s: 259
Frequency: 1	Frequency: 1
Proposed action: informs	Proposed action: afformentioned
Target: people	Target: problems
Entry 51	Entry 59
ID/s: 179	ID/s: 259
Frequency: 1	Frequency: 1
Proposed action: save	Proposed action: encounter
Target: disasters	Target: disaster
Entry 52	Entry 60
ID/s: 181	ID/s: 289
Frequency: 1	Frequency: 1
Proposed action: inviting	Proposed action: magkaron
Target: representative	Target: organisasyon
Entry 53	Entry 61
ID/s: 208	ID/s: 296
Frequency: 1	Frequency: 1
Proposed action: nasaabing	Proposed action: doing
Target: lugar	Target: meeting
Entry 54	Entry 62
ID/s: 208	ID/s: 296
Frequency: 1	Frequency: 1
Proposed action: pinamumunuang	Proposed action: talks
Target: barangay	Target: safety
Entry 55	Entry 63
ID/s: 218	ID/s: 307
Frequency: 1	Frequency: 1
Proposed action: informing	Proposed action: turuan
Target: whoe, sitio	Target: pamilya
Entry 56	Entry 64
ID/s: 243	ID/s: 331
Frequency: 1	Frequency: 1
Proposed action: magpasagawa	Proposed action: knowing
Target: disaster, drill	Target: places
Entry 57	Entry 65
ID/s: 249	

ID/s: 336	Entry 73
Frequency: 1	
Proposed action: maihnada	
Target: tao	
	Entry 66
ID/s: 343	
Frequency: 1	
Proposed action: establish	
Target: program	
	Entry 67
ID/s: 344	
Frequency: 1	
Proposed action: disseminating	
Target: information	
	Entry 68
ID/s: 381	
Frequency: 1	
Proposed action: performing	
Target: drill	
	Entry 69
ID/s: 383	
Frequency: 1	
Proposed action: t	
Target: incoming, disaster	
	Entry 70
ID/s: 389	
Frequency: 1	
Proposed action: pay	
Target: visit, house	
	Entry 71
ID/s: 394	
Frequency: 1	
Proposed action: discussing	
Target: safety, protocols	
	Entry 72
ID/s: 394	
Frequency: 1	
Proposed action: simulating	
Target: disasters	
	Entry 73
ID/s: 403	
Frequency: 1	
Proposed action: survive	
Target: disaster	
	Entry 74
ID/s: 462	
Frequency: 1	
Proposed action: natutunan	
Target: disaster (survey)	
	Entry 75
ID/s: 469	
Frequency: 1	
Proposed action: bibigay	
Target: kaalaman	
	Entry 76
ID/s: 469	
Frequency: 1	
Proposed action: dating	
Target: kalamidad, drills, paraan	
	Entry 77
ID/s: 469	
Frequency: 1	
Proposed action: dadating	
Target: kalamidad	
	Entry 78
ID/s: 565	
Frequency: 1	
Proposed action: magsgwa	
Target: kaalaman	
	Entry 79
ID/s: 565	
Frequency: 1	
Proposed action: makaiwas	
Target: skuna, kalamidad	
	Entry 80
ID/s: 578	
Frequency: 1	
Proposed action: linawin	

Target: terminolohiyang	Proposed action: linisin
Entry 81	Entry 89
ID/s: 583	Target: daan
Frequency: 1	
Proposed action: learn	
Target: things	
Entry 82	Entry 90
ID/s: 588	ID/s: 602
Frequency: 1	Frequency: 1
Proposed action: participating	Proposed action: mkaiwas
Target: garbage, disposal	Target: storms
Entry 83	Entry 91
ID/s: 588	ID/s: 610
Frequency: 1	Frequency: 1
Proposed action: adds	Proposed action: provides
Target: disaster	Target: awareness
Entry 84	Entry 92
ID/s: 591	ID/s: 627
Frequency: 1	Frequency: 1
Proposed action: regular	Proposed action: stay
Target: barangay	Target: times
Entry 85	Entry 93
ID/s: 595	ID/s: 628
Frequency: 1	Frequency: 1
Proposed action: requiring	Proposed action: papaalam
Target: family (member)	Target: tao
Entry 86	Entry 94
ID/s: 599	ID/s: 630
Frequency: 1	Frequency: 1
Proposed action: naigawa	Proposed action: suggest
Target: barangay	Target: share
Entry 87	Entry 95
ID/s: 600	ID/s: 644
Frequency: 1	Frequency: 1
Proposed action: tuwingat	Proposed action: undergo
Target: delubyo	Target: series
Entry 88	Entry 96
ID/s: 602	ID/s: 645
Frequency: 1	Frequency: 1
	Proposed action: ibahagi
	Target: paghahanda
	ID/s: 649

Frequency: 1	ID/s: 766
Proposed action: mabigay	Frequency: 1
Target: information	Proposed action: magandang
Entry 97	Target: conduct
ID/s: 651	ID/s: 797
Frequency: 1	Frequency: 1
Proposed action: magpasimuno	Proposed action: ipinatutupad
Target: drill	Target: tagapamuno
Entry 98	Entry 105
ID/s: 658	ID/s: 797
Frequency: 1	Frequency: 1
Proposed action: hold	Proposed action: naglagay
Target: seminars	Target: karatula
Entry 99	Entry 106
ID/s: 658	ID/s: 797
Frequency: 1	Frequency: 1
Proposed action: happen	Proposed action: nagkakabit
Target: beforeduring	Target: tarpaulin
Entry 100	Entry 107
ID/s: 667	ID/s: 797
Frequency: 1	Frequency: 1
Proposed action: leaflets	Proposed action: nagpapaalala
Target: safety, tips	Target: oras
Entry 101	Entry 108
ID/s: 681	ID/s: 797
Frequency: 1	Frequency: 1
Proposed action: focusing	Proposed action: kakailanganin
Target: idea	Target: residente
Entry 102	Entry 109
ID/s: 735	ID/s: 800
Frequency: 1	Frequency: 1
Proposed action: kilangan	Proposed action: maganada
Target: magkroon	Target: programa
Entry 103	Entry 110
ID/s: 751	ID/s: 801
Frequency: 1	Frequency: 1
Proposed action: magkron	Proposed action: starts
Target: pagpupulong	Target: others
Entry 104	Entry 111

Entry 112	Target: damages
ID/s: 804	Entry 120
Frequency: 1	
Proposed action: participate	ID/s: 885
Target: disaster, awareness, activities	Frequency: 1
	Proposed action: karoon
Entry 113	Target: barangay
ID/s: 811	Entry 121
Frequency: 1	
Proposed action: makaligtas	ID/s: 885
Target: taas	Frequency: 1
	Proposed action: paparating
Entry 114	Target: kalamidad
ID/s: 811	Entry 122
Frequency: 1	
Proposed action: magparticape	ID/s: 903
Target: klase	Frequency: 1
	Proposed action: magpatawag
Entry 115	Target: drill, pra
ID/s: 842	Entry 123
Frequency: 1	
Proposed action: cleaning	ID/s: 903
Target: canals	Frequency: 1
	Proposed action: ready
Entry 116	Target: lahatmgipon
ID/s: 842	Entry 124
Frequency: 1	
Proposed action: attending	ID/s: 903
Target: seminars	Frequency: 1
	Proposed action: nalamang
Entry 117	Target: sakuna
ID/s: 847	Entry 125
Frequency: 1	
Proposed action: magkakaroon	ID/s: 910
Target: orientation	Frequency: 1
	Proposed action: nagmemeeting
Entry 118	Target: araw
ID/s: 847	Entry 126
Frequency: 1	
Proposed action: hinahanda	ID/s: 910
Target: tao	Frequency: 1
	Proposed action: mabilis
Entry 119	Target: aksyon
ID/s: 869	Entry 127
Frequency: 1	
Proposed action: reduce	ID/s: 919
	Frequency: 1

Proposed action: providing
Target: evacuation, center

Entry 128

ID/s: 926
Frequency: 1
Proposed action: sakunamaari
Target: araw

Entry 129

ID/s: 929
Frequency: 1
Proposed action: assist
Target: citizen

Entry 130

ID/s: 931
Frequency: 1
Proposed action: rumesponde
Target: sakuna

Target: assistance, goods, pagkain, medisina, evacuation, tulong

Entry 4

ID/s: 418|420
Frequency: 2
Proposed action: matulungan, gawing
Target: bangka, pagkain, pagtulong

Entry 5

ID/s: 421|828
Frequency: 2
Proposed action: giving
Target: relief, goods

Entry 6

ID/s: 420
Frequency: 1
Proposed action: tinamaan
Target: kalamidad

Entry 7

ID/s: 420
Frequency: 1
Proposed action: may
Target: sakit

Entry 8

ID/s: 440
Frequency: 1
Proposed action: mailikasunahin
Target: pagbibigay

Entry 9

ID/s: 446
Frequency: 1
Proposed action: papadala
Target: kunting

Entry 10

ID/s: 497
Frequency: 1
Proposed action: maagap
Target: tulong

Entry 11

ID/s: 526

DISASTER RELIEF

Entry 1

ID/s: 110|261|299|379|795|878
Frequency: 9
Proposed action: provide, are, cannot support, be, help
Target: food (aid, supply, relief, case), kits, goods, assistance, need, afford, response

Entry 2

ID/s: 205|437|440|460|522|548
Frequency: 7
Proposed action: kailangan, handa, mag, dapat, magbibigay, sana
Target: relief, goods, resque, pagkain (gamot, pangangailangan), sakuna, pagkaen, lugar, brgy

Entry 3

ID/s: 12|559|566
Frequency: 3
Proposed action: pagbibigay

Frequency: 1	Proposed action: ibakwisyon	Target: pagbbigay	Entry 12	Proposed action: have, need, needed, create, having, back, know, is, be, cover, make, etc, contains, provide, are, ex, help
ID/s: 548				Target: emergency (place, aid, disaster, areas, food, center, people, technology), supplies, sirens, equipment, funds, evacuation, kit, area, boats, advance (source, info), facilities, publicsoundnotifsystem, barangay, gym, evacuees, electricity, condition, safety, gears, boat, kits, secure, storage, facility, assistance
Frequency: 1	Proposed action: inaasahang			
Target: sakuna				
			Entry 13	Entry 2
ID/s: 776				ID/s: 40 417 488 653
Frequency: 1				Frequency: 4
Proposed action: bibigay				Proposed action: magkaroon, dagdagan, maglagay
Target: grocery, panganga				Target: medical (center), kit, flashlight, budget, evacuation
			Entry 14	Entry 3
ID/s: 776				ID/s: 167 629 777 780
Frequency: 1				Frequency: 4
Proposed action: ilangan				Proposed action: dapat, kailangan
Target: pamilya				Target: evacuation, center, barangay, truck, bangka
			Entry 15	Entry 4
ID/s: 795				ID/s: 242 247 659
Frequency: 1				Frequency: 3
Proposed action: i				Proposed action: build
Target: government				Target: place, evacuation, area
			Entry 16	Entry 5
ID/s: 906				ID/s: 244 556
Frequency: 1				Frequency: 3
Proposed action: magtulong				Proposed action: may
Target: pagbigay				Target: tao, bangka, baha

COMMUNITY-WIDE LOGISTIC SUPPORT FOR DISASTER RESPONSE

Entry 1	Entry 6
ID/s: 106 114 134 135 142 147 154 171 212 221 242 247 270 276 302 392 397 409 675 87 9 932	ID/s: 488 590 810
Frequency: 33	Frequency: 3

Entry 7

ID/s: 71

Frequency: 2	ID/s: 242
Proposed action: karagdagang	Frequency: 1
Target: kagamitan, community, volunteers	Proposed action: serve
Entry 8	Entry 16
ID/s: 95 911	ID/s: 242
Frequency: 2	Frequency: 1
Proposed action: prepare	Proposed action: suffered
Target: place, roadway	Target: typhoons
Entry 9	Entry 17
ID/s: 109 329	ID/s: 244
Frequency: 2	Frequency: 1
Proposed action: providing	Proposed action: magpatayo
Target: shelterarea, emergency, kits	Target: evacuation, center
Entry 10	Entry 18
ID/s: 51	ID/s: 244
Frequency: 1	Frequency: 1
Proposed action: dumating	Proposed action: tamaan
Target: dilubyobagyo	Target: sakuna
Entry 11	Entry 19
ID/s: 51	ID/s: 330
Frequency: 1	Frequency: 1
Proposed action: nakahanda	Proposed action: designed
Target: gamit	Target: calamity
Entry 12	Entry 20
ID/s: 51	ID/s: 378
Frequency: 1	Frequency: 1
Proposed action: mataas	Proposed action: buy
Target: lugar	Target: instructions
Entry 13	Entry 21
ID/s: 114	ID/s: 378
Frequency: 1	Frequency: 1
Proposed action: armed	Proposed action: giving
Target: forces	Target: equipments, incase
Entry 14	Entry 22
ID/s: 170	ID/s: 392
Frequency: 1	Frequency: 1
Proposed action: disseminate	Proposed action: benefit
Target: evacuation, plan	Target: baranggay
Entry 15	

<p>Entry 23</p> <p>ID/s: 392 Frequency: 1 Proposed action: communicating Target: weather, predictions</p>	<p>Target: pasahero</p>
<p>Entry 24</p> <p>ID/s: 417 Frequency: 1 Proposed action: makabii Target: gamit, volunteer</p>	<p>Entry 31</p> <p>ID/s: 810 Frequency: 1 Proposed action: mag-deploy Target: siren</p>
<p>Entry 25</p> <p>ID/s: 434 Frequency: 1 Proposed action: makasakay Target: bangka, tao</p>	<p>Entry 32</p> <p>ID/s: 911 Frequency: 1 Proposed action: transport Target: vehicle</p>
<p>Entry 26</p> <p>ID/s: 484 Frequency: 1 Proposed action: magtapon Target: basura</p>	<p>Entry 33</p> <p>ID/s: 918 Frequency: 1 Proposed action: inform Target: storm, signal</p>
INFRASTRUCTURE MAINTENANCE AND MANAGEMENT	
<p>Entry 27</p> <p>ID/s: 534 Frequency: 1 Proposed action: nakakatulong Target: barangay</p>	<p>Entry 1</p> <p>ID/s: 4 38 43 45 47 48 52 69 231 366 368 442 4 44 458 465 482 504 512 530 541 743 779 812 817 820 825 853 905 Frequency: 33 Proposed action: kailangan, maiwasan, maglagay, ayusin, panatilihin, mabawasan, tanggalin, dapat, maayos, palitan, dagdagan, gusto, ilagay, mag Target: ikot, waste, dsiposal, pagbaha (sakuna, daanan, kalsada, baha), garbage, can, pagbabaha, drainage, kalinisan (basura, paglaganap), kanal, pangongolekta, pagbara, bata, bantay, pagbabara, basira, lage, lgu, pagbabahahindi, dn, obserba, kalamidad, tapunan, paligid</p>
<p>Entry 28</p> <p>ID/s: 534 Frequency: 1 Proposed action: sana Target: gamit</p>	<p>Entry 2</p> <p>ID/s: 57 74 78 166 449 466 493 725 743 819 82 0 822 823</p>
<p>Entry 29</p> <p>ID/s: 656 Frequency: 1 Proposed action: ready Target: evacuation, center, int</p>	
<p>Entry 30</p> <p>ID/s: 777 Frequency: 1 Proposed action: mastrandead</p>	

Frequency: 14	Target: streets, canals, surroundings
Proposed action: linisin	Entry 8
Target: kanal (basura), siguraduhing, pagtatapon, kapaligiran (paligid), especially, ilogwag, imburnal	ID/s: 62 560 822
Entry 3	Frequency: 3
ID/s: 4 62 288 476 482 500 504 537 648 746 79 0 812 822	Proposed action: itapon
Frequency: 13	Target: basura, mayus
Proposed action: wastong, tamang, malinis, iwasan, nakatulong	Entry 9
Target: pagtatapon, pagtapon, lugar, kapaligiran (lalagyan, daanan, pagtaas), tapunan, kanal	ID/s: 62 537 560
Entry 4	Frequency: 3
ID/s: 98 198 282 308 574 601 688 796 841 843	Proposed action: magbara
Frequency: 12	Target: kanal (ilog)
Proposed action: have, be, put, is, avoid, keep, help	Entry 10
Target: rules, elevation, garbage, plan, factor, floods, cleaning, dumpster, bins, flooding, garbagetrash, disposal	ID/s: 107 282 635
Entry 5	Frequency: 3
ID/s: 14 368 478 482 779 809 820 823	Proposed action: lessen
Frequency: 8	Target: probability, garbage, chances
Proposed action: magkaroon, magbigay, gumawa, maging	Entry 11
Target: basurahan, sariling, pagtitipon (programa), training, barangay, lubid, sakuna (sanhi)	ID/s: 127 617
Entry 6	Frequency: 3
ID/s: 52 480 543 761 829	Proposed action: implement
Frequency: 5	Target: cleaning, programs, segregation
Proposed action: maglinis	Entry 12
Target: kanal, kapaligiran (paligid), pausok, sakapaligiran	ID/s: 478 779 807
Entry 7	Frequency: 3
ID/s: 107 584 808 834	Proposed action: regular
Frequency: 4	Target: basura (paglilinis), pagtatapon, garbagewaste, collections
Proposed action: clean	Entry 13
	ID/s: 4 482
	Frequency: 2
	Proposed action: bantayan
	Target: gamit, kabataang
Entry 14	Entry 14
	ID/s: 57 905
	Frequency: 2
	Proposed action: walang
	Target: baha
	Entry 15
	ID/s: 69 372

Frequency: 2	Entry 23
Proposed action: ilabas, pick	
Target: basura, trashes	
Entry 16	
ID/s: 94 601	
Frequency: 2	
Proposed action: improve	
Target: drainage, system (areas)	
Entry 17	
ID/s: 504 589	
Frequency: 2	
Proposed action: itapun	
Target: basura, basurahan	
Entry 18	
ID/s: 574 635	
Frequency: 2	
Proposed action: perform	
Target: community, cleanups, clean-up, session	
Entry 19	
ID/s: 812 817	
Frequency: 2	
Proposed action: malaki, nagiging	
Target: kanal (basura), dahilan	
Entry 20	
ID/s: 14	
Frequency: 1	
Proposed action: pangunahing	
Target: sanhi	
Entry 21	
ID/s: 32	
Frequency: 1	
Proposed action: mapabilis	
Target: paglilinis	
Entry 22	
ID/s: 38	
Frequency: 1	
Proposed action: kaylangan	
Target: kanal	
Entry 23	
ID/s: 48	
Frequency: 1	
Proposed action: nakabara	
Target: kanal	
Entry 24	
ID/s: 60	
Frequency: 1	
Proposed action: linisan	
Target: paligid	
Entry 25	
ID/s: 69	
Frequency: 1	
Proposed action: dumating	
Target: truck	
Entry 26	
ID/s: 117	
Frequency: 1	
Proposed action: fixing	
Target: drainage, system	
Entry 27	
ID/s: 127	
Frequency: 1	
Proposed action: install	
Target: drainage, systems	
Entry 28	
ID/s: 127	
Frequency: 1	
Proposed action: ensure	
Target: flow	
Entry 29	
ID/s: 163	
Frequency: 1	
Proposed action: secure	
Target: drainage	
Entry 30	
ID/s: 198	
Frequency: 1	
Proposed action: fix	

Target: drainage, system	Proposed action: nagtatapon
Entry 31	Entry 39
ID/s: 279	ID/s: 458
Frequency: 1	Frequency: 1
Proposed action: tumulong	Proposed action: daming
Target: kagamitan	Target: tindahan
Entry 32	Entry 40
ID/s: 279	ID/s: 458
Frequency: 1	Frequency: 1
Proposed action: panglinis	Proposed action: gitna
Target: kalamidad	Target: kalsada
Entry 33	Entry 41
ID/s: 288	ID/s: 465
Frequency: 1	Frequency: 1
Proposed action: nagtulung	Proposed action: maipon
Target: kapaligiran	Target: tubig
Entry 34	Entry 42
ID/s: 308	ID/s: 474
Frequency: 1	Frequency: 1
Proposed action: mentioned	Proposed action: maimprove
Target: survey	Target: drainage, system
Entry 35	Entry 43
ID/s: 361	ID/s: 474
Frequency: 1	Frequency: 1
Proposed action: cleaning	Proposed action: mainform
Target: canals	Target: tao
Entry 36	Entry 44
ID/s: 449	ID/s: 480
Frequency: 1	Frequency: 1
Proposed action: magtapon	Proposed action: magtanim
Target: basura	Target: puno, basura
Entry 37	Entry 45
ID/s: 454	ID/s: 482
Frequency: 1	Frequency: 1
Proposed action: ngbbgay	Proposed action: pasaway
Target: dhlan	Target: basura
Entry 38	Entry 46
ID/s: 458	ID/s: 482
Frequency: 1	

Frequency: 1	ID/s: 601
Proposed action: gagawing	Frequency: 1
Target: hagdan	Proposed action: eg
	Target: rivers
	Entry 47
ID/s: 482	Entry 55
Frequency: 1	ID/s: 617
Proposed action: makasagip	Frequency: 1
Target: bata	Proposed action: cause
	Target: flood
	Entry 56
ID/s: 493	Entry 56
Frequency: 1	ID/s: 617
Proposed action: makaiwas	Frequency: 1
Target: pgaha	Proposed action: thrice
	Target: week
	Entry 57
ID/s: 537	Entry 57
Frequency: 1	ID/s: 635
Proposed action: maghada	Frequency: 1
Target: basura	Proposed action: flooding
	Target: rains
	Entry 58
ID/s: 569	Entry 58
Frequency: 1	ID/s: 779
Proposed action: maintain	Frequency: 1
Target: kalinisan	Proposed action: pumping
	Target: station
	Entry 59
ID/s: 589	Entry 59
Frequency: 1	ID/s: 779
Proposed action: pagtulungan	Frequency: 1
Target: lugar	Proposed action: pinondohan
	Target: taong, bayan
	Entry 60
ID/s: 589	Entry 60
Frequency: 1	ID/s: 790
Proposed action: ikakabuti	Frequency: 1
Target: kanal	Proposed action: bumara
	Target: basura, baha
	Entry 61
ID/s: 589	Entry 61
Frequency: 1	ID/s: 799
Proposed action: iwas	Frequency: 1
Target: baha	Proposed action: magsawa
	Target: seminars
	Entry 54

<p>Entry 62</p> <p>ID/s: 807 Frequency: 1 Proposed action: delay Target: cause</p> <p>Entry 63</p> <p>ID/s: 812 Frequency: 1 Proposed action: khit Target: lakas</p> <p>Entry 64</p> <p>ID/s: 817 Frequency: 1 Proposed action: naipapataasan Target: karanasan</p> <p>Entry 65</p> <p>ID/s: 817 Frequency: 1 Proposed action: tanging Target: kalsada</p> <p>Entry 66</p> <p>ID/s: 817 Frequency: 1 Proposed action: namen Target: epekto</p> <p>Entry 67</p> <p>ID/s: 817 Frequency: 1 Proposed action: napupunta Target: lugar</p> <p>Entry 68</p> <p>ID/s: 823 Frequency: 1 Proposed action: taasan Target: kalsada</p> <p>Entry 69</p> <p>ID/s: 843 Frequency: 1 Proposed action: ensuring</p>	<p>Target: surrounding</p> <p>Entry 70</p> <p>ID/s: 843 Frequency: 1 Proposed action: prevent Target: blockage</p> <hr/> <p style="text-align: center;">EARLY WARNING SYSTEM</p> <hr/> <p>Entry 1</p> <p>ID/s: 85 102 105 111 112 125 137 139 140 146 175 176 224 235 290 291 292 300 304 309 313 373 519 573 575 619 684 708 859 90 9 917</p> <p>Frequency: 48 Proposed action: warning, do, make, be, have, coming, drive, is, are, has, eat, visit, update, give, avoid, said, watch, provide</p> <p>Target: signal, announcements, announcement, training (technology, community, disaster, warning, sign, typhoon, days, people, house, food, events, emergency), discipline, plenty, beforehand, floods, incoming, time (need), barangay, sms, alert, constituents, everyone, theeveryone, damage, devices, calamities, megaphones, tips, announcements, news, loss, happenings, exits</p> <p>Entry 2</p> <p>ID/s: 3 6 7 8 11 20 27 50 58 66 77 335 419 436 438 467 473 535 550 555 557 558 685 749 788 789</p> <p>Frequency: 32 Proposed action: magkaroon, maging, pagkakaroon, makita, mag, magbibigay, mabigyan, bigyang, malaman, makatulong, magbigay</p> <p>Target: komunikasyon (impormasyon, pangangailangan), early (warning, maintenance, members, notice, report), system (device), quarterly, council, tao, speaker, sitwasyon (detalye), coordinate,</p>
---	--

babala (bagyo), paglaki, signal, kalamidad, bahaan, announce

Entry 3

ID/s: 58|90|204|306|436|448|495|529|554|821|8
91|894

Frequency: 17

Proposed action: dapat, handa

Target: sakuna, eg, komunikasyon, kita, mamamayan, pamamagitan, anunsyo, baragay, alert, barangay (baranggay), balita, radio (update), abiso, saknila, my, darating, talad

Entry 4

ID/s: 20|263|450|481|698|712|774|814|833|857|
914

Frequency: 14

Proposed action: may, walang

Target: kalamidad, sakunang, paalam, bagyo (baha), assembly (disaster), iparating, calamidad, bagyono, anu, komunikasyon, sakuna, suporta, pra, sawang

Entry 5

ID/s: 50|58|67|75|263|445|547|727|814|844

Frequency: 10

Proposed action: malinaw, tamang, epektibong, kailangan, inaasahan, malaki, malakas, makakatulong

Target: koordinasyon, panahon, pasabi, tahanan (tulong), anunsyo, tubig, n, bagyo, better (way), barangay

Entry 6

ID/s: 128|189|223|309|313|354|619|782|837|874

Frequency: 10

Proposed action: inform

Target: citizens, typhoon (disaster, public, community), preparedness, residents, announce, na

Entry 7

ID/s: 6|65|547|554|724

Frequency: 6

Proposed action: dumating, lumikas, tumira

Target: bagyo, kalamidad, sakuna, bgyan, talang, evacuation, center

Entry 8

ID/s: 67|263|697|727|837

Frequency: 5

Proposed action: paparating, darating

Target: sakuna, kalamidad, na, bigay, lugar

Entry 9

ID/s: 137|313|619|664|921

Frequency: 5

Proposed action: prepare

Target: community (disaster, things), alarm, weeks

Entry 10

ID/s: 278|697|727|772|856

Frequency: 5

Proposed action: maagang, maaga

Target: announcement, if (ever), impormasyon, kalamidad, anunsyo, palang

Entry 11

ID/s: 272|573|671|686

Frequency: 4

Proposed action: alert

Target: people, citizens, everyone, residents

Entry 12

ID/s: 306|490|724|833

Frequency: 4

Proposed action: gagawin, lagi, nag, magiging

Target: disaster, atpagbigay, anouse, merng, aware

Entry 13

ID/s: 139|176|237

Frequency: 3

Proposed action: announcing

Target: time (people), news

Entry 14

ID/s: 272|708

Frequency: 3	Entry 22
Proposed action: giving	ID/s: 413 419
Target: announcements, newsinformation, beforehand, updates	Frequency: 2
Entry 15	Proposed action: wang
ID/s: 788 821 844	Target: wang
Frequency: 3	Entry 23
Proposed action: nasasakupan	ID/s: 413 419
Target: announce, kalamidad, tulong	Frequency: 2
Entry 16	Proposed action: ipagpatuluy
ID/s: 58	Target: barangay, pagbigay, babala
Frequency: 2	Entry 24
Proposed action: paglikas	ID/s: 547 844
Target: pamilya, gamit	Frequency: 2
Entry 17	Proposed action: iinform
ID/s: 97 388	Target: tao, earlier
Frequency: 2	Entry 25
Proposed action: regarding	ID/s: 712 724
Target: disaster, disasters	Frequency: 2
Entry 18	Proposed action: matigas, umiikot
ID/s: 102 105	Target: ulo, tao
Frequency: 2	Entry 26
Proposed action: roam	ID/s: 724 789
Target: barangay	Frequency: 2
Entry 19	Proposed action: mababang, mataas
ID/s: 189 277	Target: lugar
Frequency: 2	Entry 27
Proposed action: texting	ID/s: 887 914
Target: barangay, community (person)	Frequency: 2
Entry 20	Proposed action: mabilis, naapektuhan
ID/s: 264 619	Target: mensahe, bagyo
Frequency: 2	Entry 28
Proposed action: updated	ID/s: 3
Target: palagi, weather, changes	Frequency: 1
Entry 21	Proposed action: magkikita
ID/s: 300 304	Target: panahon
Frequency: 2	Entry 29
Proposed action: suggest	ID/s: 8
Target: barangay, brgy	Frequency: 1
	Proposed action: pagtibayin

Target: early (warning), system (device)	Proposed action: ma-inform
Entry 30	Entry 38
ID/s: 8	ID/s: 103
Frequency: 1	Frequency: 1
Proposed action: masigurong	Proposed action: improving
Target: kalamidad	Target: imformation, dissemenation
Entry 31	Entry 39
ID/s: 29	ID/s: 112
Frequency: 1	Frequency: 1
Proposed action: bahay-bahay	Proposed action: improved
Target: pag-inform	Target: communication
Entry 32	Entry 40
ID/s: 30	ID/s: 112
Frequency: 1	Frequency: 1
Proposed action: pagiinform	Proposed action: minimize
Target: chairman	Target: theft
Entry 33	Entry 41
ID/s: 30	ID/s: 128
Frequency: 1	Frequency: 1
Proposed action: mapalaganap	Proposed action: messaging
Target: inpmasyon	Target: media (etc)
Entry 34	Entry 42
ID/s: 65	ID/s: 133
Frequency: 1	Frequency: 1
Proposed action: paalalahanan	Proposed action: providing
Target: barangay	Target: safety, places
Entry 35	Entry 43
ID/s: 66	ID/s: 164
Frequency: 1	Frequency: 1
Proposed action: maiaannounce	Proposed action: galingan
Target: balita	Target: gid
Entry 36	Entry 44
ID/s: 75	ID/s: 172
Frequency: 1	Frequency: 1
Proposed action: pagtawag	Proposed action: reaching
Target: anak	Target: troubles
Entry 37	Entry 45
ID/s: 90	ID/s: 175
Frequency: 1	

Frequency: 1	ID/s: 300
Proposed action: telling	Frequency: 1
Target: people	Proposed action: caused
Entry 46	
ID/s: 176	Target: typhoon
Frequency: 1	ID/s: 313
Proposed action: built	Frequency: 1
Target: speakers	Proposed action: prepared
Entry 47	
ID/s: 250	Target: way
Frequency: 1	ID/s: 313
Proposed action: incoming	Frequency: 1
Target: disasters	Proposed action: accommodate
Entry 48	
ID/s: 257	Target: barangay
Frequency: 1	ID/s: 313
Proposed action: nagkulang	Frequency: 1
Target: pagbibigay, babala	Proposed action: allow
Entry 49	
ID/s: 273	Target: residents
Frequency: 1	ID/s: 313
Proposed action: alerting	Frequency: 1
Target: people	Proposed action: participating
Entry 50	
ID/s: 286	Target: needs
Frequency: 1	ID/s: 319
Proposed action: maitutulong	Frequency: 1
Target: barangay	Proposed action: dealing
Entry 51	
ID/s: 286	Target: disaster
Frequency: 1	ID/s: 334
Proposed action: ipamahagi	Frequency: 1
Target: impormasyon	Proposed action: spread
Entry 52	
ID/s: 286	Target: news
Frequency: 1	ID/s: 413
Proposed action: mangyaring	Frequency: 1
Target: sakuna	Proposed action: kpag
Entry 53	
	Target: tubig
Entry 54	
Entry 55	
Entry 56	
Entry 57	
Entry 58	
Entry 59	
Entry 60	

Entry 61	Target: paqdating
ID/s: 413	Entry 69
Frequency: 1	
Proposed action: nagbbigay	ID/s: 529
Target: babala	Frequency: 1
	Proposed action: makausap
Entry 62	Target: tulong
ID/s: 431	Entry 70
Frequency: 1	
Proposed action: ialarm	ID/s: 529
Target: barangay	Frequency: 1
	Proposed action: mailikas
Entry 63	Target: tubig
ID/s: 431	Entry 71
Frequency: 1	
Proposed action: namemegaphone	ID/s: 535
Target: ilog	Frequency: 1
	Proposed action: sabihin
Entry 64	Target: bagyong
ID/s: 431	Entry 72
Frequency: 1	
Proposed action: ingat	ID/s: 554
Target: level	Frequency: 1
	Proposed action: mapaghandaan
Entry 65	Target: comunidad
ID/s: 445	Entry 73
Frequency: 1	
Proposed action: magalarm	ID/s: 557
Target: sirena	Frequency: 1
	Proposed action: makipag
Entry 66	Target: ugnayan
ID/s: 448	Entry 74
Frequency: 1	
Proposed action: magtawag	ID/s: 571
Target: baranggay	Frequency: 1
	Proposed action: facebook
Entry 67	Target: twitter
ID/s: 481	Entry 75
Frequency: 1	
Proposed action: kapit	ID/s: 573
Target: bahay	Frequency: 1
	Proposed action: upcoming
Entry 68	Target: disasters
ID/s: 490	Entry 76
Frequency: 1	
Proposed action: paqaabiso	ID/s: 619
	Frequency: 1

Proposed action: evacuate	Frequency: 1
Target: area	Proposed action: ipabatid
Entry 77	Entry 85
ID/s: 683	
Frequency: 1	
Proposed action: knowing	
Target: disaster	
Entry 78	Entry 86
ID/s: 685	
Frequency: 1	
Proposed action: aga	
Target: respondihan	
Entry 79	Entry 87
ID/s: 698	
Frequency: 1	
Proposed action: pagpapa-alala	
Target: sakuna	
Entry 80	Entry 88
ID/s: 710	
Frequency: 1	
Proposed action: utilize	
Target: media	
Entry 81	Entry 89
ID/s: 710	
Frequency: 1	
Proposed action: disseminate	
Target: awareness	
Entry 82	Entry 90
ID/s: 713	
Frequency: 1	
Proposed action: makinig	
Target: advisory	
Entry 83	Entry 91
ID/s: 724	
Frequency: 1	
Proposed action: masabihan	
Target: tao, n	
Entry 84	Entry 92
ID/s: 739	

ID/s: 894
Frequency: 1
Proposed action: mgbibigay
Target: signal

Entry 93

ID/s: 914
Frequency: 1
Proposed action: maagap
Target: impormasyon

PREPAREDNESS FOR EMERGENCY

Entry 1

ID/s:
22|36|72|73|122|148|151|203|211|341|342|
506|510|528|532|533|562|612|642|673|726|
|748|767|775|793|803|818|850|851|890|89
2|896|899|900|901|908|922

Frequency: 43

Proposed action: maayos, magamit, malaman, dapat, magpadala, matugunan, kailangan, pagkilos, handa, mabilis, maiwasan, mahalaga, sapat

Target: paghahanda (pangangailangan, kaligtasan, pagkain), comunicasion, panahan, barangay (baranggay), kalagayan (sakuna), disaster (official, program), balita (ulat, programa), aware, tao, oras (araw), updated, alirto, radyo, kalamidad, ata, always, alert, membro, i, bahay, pagka, trap, stocks, disciplina, mangyare, karapat, pra

Entry 2

ID/s:
16|19|35|321|341|365|439|447|502|616|62
6|660|690|714|728|731|747|748|752|753|7
83|785|786|806|815|824|830|889

Frequency: 29

Proposed action: maging, magkaroon, pagiging, magiging

Target: bagay (kalikasan), bagyobaha, oras, sakuna, pra, alertoat, handa (darating), atentibo, balita, daratinga, cla, brgys, pkiking, kalamidad, kalamid, kagamatin, opisyal,

kasanyan, xa, participation, sariling, disiplina, uusap, emergency (plan), prepared

Entry 3

ID/s:
123|145|158|159|165|199|233|248|324|380|
|680|704|738|755|816|840|863|868|916|92
0

Frequency: 23

Proposed action: be, check, know, is, help, being, need, go, have, keep, take

Target: news (disaster, typhoon, week, times), officials, happenings, evacuation, plan (things, management, unit, community, time, proper, case, council, emergency), flood, weather, solution, preparedness, risk, reduction, calamity, kit, seminar

Entry 4

ID/s: 61|122|365|492|506|510|613|748|753

Frequency: 9

Proposed action: may

Target: sakuna (kapaligiran), bagyo, sakunang, sakona, pomoang, barangay, kabbaranggay, nkalagay, darating

Entry 5

ID/s: 23|56|528|533|544|818

Frequency: 7

Proposed action: dumating, pagdating, pumunta

Target: sakuna, kalamidad, obserbatibo, safe

Entry 6

ID/s: 315|452|463|613|624|729|850

Frequency: 7

Proposed action: mag

Target: food, supplies, kagamitang (gamot), monitor, samahan, plano, laan, imbak

Entry 7

ID/s: 61|722|742|908

Frequency: 4

Proposed action: maghanda

Target: gamit, anu, bagay, mangyyari, sakuna, pagkain

Entry 8

ID/s: 690 729 890 912	Entry 16
Frequency: 4	ID/s: 203 896
Proposed action: alerto	Frequency: 2
Target: bagay, kalamidada, tv, darating	Proposed action: nangyayari, nagaganap
Entry 9	Target: paligid, barangay
ID/s: 10 756 892	Entry 17
Frequency: 3	ID/s: 248 864
Proposed action: sinasabi, gagawin, meron	Frequency: 2
Target: sakuna, my, paghahanda, pra	Proposed action: comes
Entry 10	Target: time, funds
ID/s: 73 315 769	Entry 18
Frequency: 3	ID/s: 312 672
Proposed action: wala, pwede	Frequency: 2
Target: kalamidad, oras	Proposed action: prepare
Entry 11	Target: disaster (needs)
ID/s: 159 228 284	Entry 19
Frequency: 3	ID/s: 622 626
Proposed action: watch, radio	Frequency: 2
Target: weather, forecast, news (disaster)	Proposed action: makinig
Entry 12	Target: anunyo, balita
ID/s: 211 533 924	Entry 20
Frequency: 3	ID/s: 623 915
Proposed action: tignan, ilagay, simulan	Frequency: 2
Target: bagyo, bgo, baranggay	Proposed action: canned
Entry 13	Target: food, batteries, goods
ID/s: 23 924	Entry 21
Frequency: 2	ID/s: 722 818
Proposed action: maagang, malaking	Frequency: 2
Target: paghahanda, problema	Proposed action: mangyare
Entry 14	Target: darateng, bagyo
ID/s: 37 56	Entry 22
Frequency: 2	ID/s: 16
Proposed action: kaylangan	Frequency: 1
Target: aware, mamamayan	Proposed action: magtulong
Entry 15	Target: tulong
ID/s: 72 924	Entry 23
Frequency: 2	ID/s: 16
Proposed action: nakatuon, laganap	Frequency: 1
Target: weather, forecast, kurapsyon	Proposed action: opisyal

Target: brgy	Proposed action: dadaan
Entry 24	Entry 32
ID/s: 36	ID/s: 284
Frequency: 1	Frequency: 1
Proposed action: ihanda	Proposed action: watching
Target: pweding	Target: news (television)
Entry 25	Entry 33
ID/s: 99	ID/s: 312
Frequency: 1	Frequency: 1
Proposed action: doing	Proposed action: alert
Target: flood, prone, areas	Target: calamity
Entry 26	Entry 34
ID/s: 122	ID/s: 447
Frequency: 1	Frequency: 1
Proposed action: palaging	Proposed action: makipagtutulungan
Target: balita	Target: barangay
Entry 27	Entry 35
ID/s: 165	ID/s: 452
Frequency: 1	Frequency: 1
Proposed action: listen	Proposed action: binabaha
Target: radio	Target: tao
Entry 28	Entry 36
ID/s: 165	ID/s: 492
Frequency: 1	Frequency: 1
Proposed action: regarding	Proposed action: kumplitohin
Target: weather, condition	Target: gamit
Entry 29	Entry 37
ID/s: 280	ID/s: 616
Frequency: 1	Frequency: 1
Proposed action: bumuo	Proposed action: iimbak
Target: emergency, plansan	Target: pagkain
Entry 30	Entry 38
ID/s: 280	ID/s: 624
Frequency: 1	Frequency: 1
Proposed action: pupunta	Proposed action: kakailanganin
Target: oras	Target: panahon
Entry 31	Entry 39
ID/s: 280	ID/s: 625
Frequency: 1	

Frequency: 1	ID/s: 752
Proposed action: med	Frequency: 1
Target: kits	Proposed action: ipagbigay
Entry 40	Target: mamamayan
ID/s: 626	ID/s: 755
Frequency: 1	Frequency: 1
Proposed action: mapagbigay-alam	Proposed action: helps
Target: kapit-bahay	Target: community
Entry 41	Entry 49
ID/s: 652	ID/s: 756
Frequency: 1	Frequency: 1
Proposed action: tumulong	Proposed action: alisto
Target: pagkalap	Target: n, balita
Entry 42	Entry 50
ID/s: 673	ID/s: 756
Frequency: 1	Frequency: 1
Proposed action: maka	Proposed action: ipahatid
Target: alis	Target: mamayan
Entry 43	Entry 51
ID/s: 680	ID/s: 769
Frequency: 1	Frequency: 1
Proposed action: inform	Proposed action: sakin
Target: barangay	Target: kana
Entry 44	Entry 52
ID/s: 731	ID/s: 786
Frequency: 1	Frequency: 1
Proposed action: ibahagi	Proposed action: makaiwas
Target: nl	Target: sakuna
Entry 45	Entry 53
ID/s: 731	ID/s: 793
Frequency: 1	Frequency: 1
Proposed action: ibalita	Proposed action: nasasakupan
Target: s	Target: problemang
Entry 46	Entry 54
ID/s: 748	ID/s: 806
Frequency: 1	Frequency: 1
Proposed action: magki2ta	Proposed action: mapaghandaan
Target: oras	Target: tulong
Entry 47	

<p>Entry 55</p> <p>ID/s: 816 Frequency: 1 Proposed action: lessen Target: scale</p> <p>Entry 56</p> <p>ID/s: 824 Frequency: 1 Proposed action: nangangailangang Target: interaksyon</p> <p>Entry 57</p> <p>ID/s: 881 Frequency: 1 Proposed action: paghahanda Target: kalamidad</p> <p>Entry 58</p> <p>ID/s: 896 Frequency: 1 Proposed action: darating Target: barangay</p> <p>Entry 59</p> <p>ID/s: 912 Frequency: 1 Proposed action: sakunadapat Target: ee, check</p> <p>Entry 60</p> <p>ID/s: 916 Frequency: 1 Proposed action: eg Target: goods, battery</p> <p>Entry 61</p> <p>ID/s: 916 Frequency: 1 Proposed action: clothing Target: etc</p> <p>Entry 62</p> <p>ID/s: 920 Frequency: 1 Proposed action: encouraging</p>	<p>Target: barangay, people</p> <p>Entry 63</p> <p>ID/s: 924 Frequency: 1 Proposed action: naghihirap Target: bansa, dahilbansa</p> <hr/> <p style="text-align: center;">LOCAL GOVERNMENT ACCOUNTABILITY</p> <hr/> <p>Entry 1</p> <p>ID/s: 150 153 196 201 210 268 285 318 353 371 928 934 Frequency: 18 Proposed action: help, be, are, want, working, make, have, know, having, feed Target: disasters, help (people, map, part, government, case), barangay, watcg, officials, personnel, evacuation, incase, posters, society</p> <p>Entry 2</p> <p>ID/s: 28 64 76 441 457 471 472 643 732 770 83 5 913 Frequency: 15 Proposed action: mabuti, dapat, ayusin, gawin, sana, malaman, kailangan, ligtas, bigyan, mag Target: barangay, problema (patakaran, paraan, pansin), active, kalamidad, lugar, kahandaan, ugnayan, mgsma, sma, mg, usap, nmin, hrpin, maiparting, dn, pundo</p> <p>Entry 3</p> <p>ID/s: 64 471 472 Frequency: 3 Proposed action: may Target: sakuna (kapansanan), tagapamahala</p> <p>Entry 4</p> <p>ID/s: 457 471 Frequency: 3 Proposed action: maging</p>
---	--

Target: barangay, tagapag, tao	Proposed action: enforce
Entry 5	Entry 13
ID/s: 64 643	Target: laws
Frequency: 2	
Proposed action: mabilis, maiwasan	
Target: pagbibigay, sakuna	
Entry 6	Entry 14
ID/s: 201 318	Target: trash
Frequency: 2	
Proposed action: happen	
Target: community, news	
Entry 7	Entry 15
ID/s: 471 904	Target: place
Frequency: 2	
Proposed action: nag	
Target: pamagitna, kalaminad	
Entry 8	Entry 16
ID/s: 28	Target: times
Frequency: 1	
Proposed action: maari	
Target: performance	
Entry 9	Entry 17
ID/s: 28	Target: awareness
Frequency: 1	
Proposed action: maipaabot	
Target: serbisyo	
Entry 10	Entry 18
ID/s: 82	Target: people
Frequency: 1	
Proposed action: gumawa	
Target: organisasyon	
Entry 11	Entry 19
ID/s: 82	Target: surrounds
Frequency: 1	
Proposed action: magtutulongan	
Target: paglilinis	
Entry 12	Entry 20
ID/s: 96	Target: one
Frequency: 1	
	ID/s: 201
	Frequency: 1
	Proposed action: hurt
	Target: typhoon
	ID/s: 202

Frequency: 1	ID/s: 477
Proposed action: respond	Frequency: 1
Target: emergency	Proposed action: andyan
Entry 21	Entry 29
ID/s: 210	ID/s: 518
Frequency: 1	Frequency: 1
Proposed action: indicates	Proposed action: nangaylangan
Target: altitude	Target: tulong
Entry 22	Entry 30
ID/s: 268	ID/s: 577
Frequency: 1	Frequency: 1
Proposed action: assigned	Proposed action: facilitate
Target: news	Target: disaster, response
Entry 23	Entry 31
ID/s: 441	ID/s: 699
Frequency: 1	Frequency: 1
Proposed action: puntahan	Proposed action: pwersahan
Target: opisyal	Target: tao
Entry 24	Entry 32
ID/s: 451	ID/s: 701
Frequency: 1	Frequency: 1
Proposed action: maagap	Proposed action: knowing
Target: pagtulong	Target: disaster
Entry 25	Entry 33
ID/s: 471	ID/s: 770
Frequency: 1	Frequency: 1
Proposed action: paman	Proposed action: mlman
Target: tao	Target: dpt, gwen
Entry 26	Entry 34
ID/s: 471	ID/s: 845
Frequency: 1	Frequency: 1
Proposed action: maibahagi	Proposed action: utilizing
Target: tao	Target: tanod
Entry 27	Entry 35
ID/s: 472	ID/s: 870
Frequency: 1	Frequency: 1
Proposed action: malapit	Proposed action: nakkatulong
Target: tabing	Target: tao, regarding
Entry 28	

<p>Entry 36</p> <p>ID/s: 870 Frequency: 1 Proposed action: paparating Target: sakuna</p>	<p>Entry 2</p> <p>ID/s: 21 49 423 433 525 540 564 759 Frequency: 9 Proposed action: tumulong, makatulong, maiwasan Target: bahay, pagkain, bagyo (baha), community, volunteer, bata, tulong, barangay, pra, komunidad</p>
<p>Entry 37</p> <p>ID/s: 904 Frequency: 1 Proposed action: punong Target: brgy</p>	<p>Entry 3</p> <p>ID/s: 168 262 393 861 Frequency: 7 Proposed action: being, have, needs, cannot be Target: disaster, share, volunteer, announcements, reminders, trick, heart, barangay</p>
<p>Entry 38</p> <p>ID/s: 904 Frequency: 1 Proposed action: manguna Target: palaganap</p>	<p>Entry 4</p> <p>ID/s: 31 39 538 564 757 888 Frequency: 6 Proposed action: dapat, inaasahan, maayos Target: pakiki-isa, komunidad, panahon, mgkaisa, tulong, kalamidad</p>
<p>Entry 39</p> <p>ID/s: 904 Frequency: 1 Proposed action: mayrong Target: kalamidad</p>	<p>Entry 5</p> <p>ID/s: 31 426 715 832 Frequency: 5 Proposed action: may Target: sakunang, bagyo, kalamidadwag, kalamidad</p>
<p>Entry 40</p> <p>ID/s: 904 Frequency: 1 Proposed action: dumating Target: kuminidad</p>	<p>Entry 6</p> <p>ID/s: 41 443 621 721 Frequency: 5 Proposed action: maging, gawing Target: darating, kalagayan, agrisibo, kalamidad, serbisyo, gsy</p>
<p>FILIPINO VALUES</p>	<p>Entry 7</p> <p>ID/s: 410 487 494 520 858 Frequency: 5 Proposed action: nangangailangan, dagdagan, palagiang, malinis Target: atensyon, pakikisalamuha (kapaligiran), barangay, kpaligirantumulong</p>

Entry 8	Proposed action: nagkakaisa Target: ren, lng
ID/s: 1 17 63 427 Frequency: 4 Proposed action: magkaisa Target: tao, tungkulin (plano), paghahanda	Entry 16
Entry 9	ID/s: 567 898 Frequency: 2 Proposed action: makikipag Target: baranggay, tulunganat
ID/s: 422 486 498 503 Frequency: 4 Proposed action: makipag Target: barangay (baranggay), tao	Entry 17
Entry 10	ID/s: 666 759 Frequency: 2 Proposed action: karatig, kalapit Target: bayan, bahay
ID/s: 21 44 410 Frequency: 3 Proposed action: napinsala, nagkakaroon, apektado Target: bagyo, lugar, kalamidad, pansi	Entry 18
Entry 11	ID/s: 9 Frequency: 1 Proposed action: lumawat Target: pagmamasakit
ID/s: 443 723 768 Frequency: 3 Proposed action: magtulong Target: tulong	Entry 19
Entry 12	ID/s: 53 Frequency: 1 Proposed action: makiisa Target: komunidad
ID/s: 31 538 Frequency: 2 Proposed action: dumating, nakarating Target: kalamidad	Entry 20
Entry 13	ID/s: 161 Frequency: 1 Proposed action: magtulong2 Target: kalamidad
ID/s: 55 523 Frequency: 2 Proposed action: tutulong Target: paghahanda, baranggay	Entry 21
Entry 14	ID/s: 260 Frequency: 1 Proposed action: matutong Target: malasakit
ID/s: 428 715 Frequency: 2 Proposed action: magtulungan Target: s, tao	Entry 22
Entry 15	ID/s: 298 Frequency: 1 Proposed action: helping Target: others
ID/s: 508 832 Frequency: 2	Entry 23
	ID/s: 355

Frequency: 1	ID/s: 443
Proposed action: matutunang	Frequency: 1
Target: oras	Proposed action: naperwisyo
Entry 24	Target: baha
ID/s: 367	ID/s: 443
Frequency: 1	Frequency: 1
Proposed action: nangyayaring	Proposed action: umasa
Target: barangay	Target: barangaymas
Entry 25	Entry 32
ID/s: 410	ID/s: 443
Frequency: 1	Frequency: 1
Proposed action: pantay-pantay	Proposed action: mkikipgtolongan
Target: pamamahagi	Target: programa
Entry 26	Entry 33
ID/s: 410	ID/s: 453
Frequency: 1	Frequency: 1
Proposed action: mabigyang	Proposed action: magbuo
Target: solusyon	Target: samahan
Entry 27	Entry 34
ID/s: 411	ID/s: 459
Frequency: 1	Frequency: 1
Proposed action: magmalasakit	Proposed action: makatulong
Target: kawpa	Target: lugar
Entry 28	Entry 35
ID/s: 412	ID/s: 483
Frequency: 1	Frequency: 1
Proposed action: wala	Proposed action: tulongan
Target: barangay	Target: esatesa
Entry 29	Entry 36
ID/s: 412	ID/s: 531
Frequency: 1	Frequency: 1
Proposed action: mali	Proposed action: makipagtulungan
Target: kelangan	Target: barangay
Entry 30	Entry 37
ID/s: 415	ID/s: 570
Frequency: 1	Frequency: 1
Proposed action: kaylangan	Proposed action: maitutulong
Target: isat	Target: barangay
Entry 31	Entry 38
ID/s: 415	ID/s: 641
Frequency: 1	Frequency: 1
Proposed action: kaylangan	Proposed action: maitutulong
Target: barangay	Target: barangay

<p>Entry 39</p> <p>ID/s: 641 Frequency: 1 Proposed action: nasiraan Target: bahay</p> <p>Entry 40</p> <p>ID/s: 691 Frequency: 1 Proposed action: tutulongan Target: barangay</p> <p>Entry 41</p> <p>ID/s: 691 Frequency: 1 Proposed action: maitayo Target: kagamitan</p> <p>Entry 42</p> <p>ID/s: 694 Frequency: 1 Proposed action: magtulongan Target: magkakapamilya</p> <p>Entry 43</p> <p>ID/s: 715 Frequency: 1 Proposed action: magtaponi Target: basura</p> <p>Entry 44</p> <p>ID/s: 720 Frequency: 1 Proposed action: magsama Target: sama</p> <p>Entry 45</p> <p>ID/s: 730 Frequency: 1 Proposed action: maka Target: sakuna</p> <p>Entry 46</p> <p>ID/s: 740 Frequency: 1 Proposed action: tumolng</p>	<p>Target: n</p> <p>Entry 47</p> <p>ID/s: 740 Frequency: 1 Proposed action: naabutan Target: baha</p> <p>Entry 48</p> <p>ID/s: 750 Frequency: 1 Proposed action: makapaghanda Target: s</p> <hr/> <p style="text-align: center;">OTHERS</p> <hr/> <p>Entry 1</p> <p>ID/s: 184 185 187 188 192 194 195 197 241 695 Frequency: 16 Proposed action: know, cannot, is, help, has, please, be, are, want, better, give, let, avoid Target: deaf, train, emergency (people), responders, sign (question, help, object), teach, aware, prepare, neighborhoods, barangay, world, permission</p> <p>Entry 2</p> <p>ID/s: 455 791 902 923 Frequency: 6 Proposed action: dapat, nararapat, maayos Target: tao, barangay, pondo, pra, kominidad</p> <p>Entry 3</p> <p>ID/s: 425 501 542 923 Frequency: 5 Proposed action: humingi, ipagpatuloy, magbigay, handa Target: tulung, barangay, gawain, impormasyon, tao</p> <p>Entry 4</p> <p>ID/s: 479 637 791 802 Frequency: 4 Proposed action: mag, sana, sanang</p>
---	--

Target: kona, tao, negosyo, dhil, khit	Frequency: 1
Entry 5	
ID/s: 314 791 826	Proposed action: calling
Frequency: 3	Target: barangay, community
Proposed action: wala, walang	
Target: pakielam, kwenta, kwentang, bagay	
Entry 6	
ID/s: 501 791	Frequency: 1
Frequency: 3	Proposed action: suggest
Proposed action: nakikita, maganda, apektado	Target: typhoon, amd, flood
Target: q, gawain, tang, ina	
Entry 7	
ID/s: 115 118	Frequency: 1
Frequency: 2	Proposed action: cant
Proposed action: putting	Target: life
Target: pockets	
Entry 8	
ID/s: 184	Frequency: 1
Frequency: 2	Proposed action: build
Proposed action: communicate	Target: houses
Target: deaf, citizens, responders	
Entry 9	
ID/s: 513 611	Frequency: 1
Frequency: 2	Proposed action: prevent
Proposed action: maging	Target: flashfloods
Target: pamaraan, agrisibo	
Entry 10	
ID/s: 183	Frequency: 1
Frequency: 1	Proposed action: magpalit
Proposed action: seen	Target: brgy, chairman, kase
Target: typhoon (disaster)	
Entry 11	
ID/s: 191	Frequency: 1
Frequency: 1	Proposed action: makinig
Proposed action: pls	Target: locals, ulo
Target: i	
Entry 12	
ID/s: 193	Frequency: 1
	Proposed action: mkaiwas
	Target: baha
	Entry 13
ID/s: 193	
Frequency: 1	
Proposed action: suggest	
Target: barangay, community	
	Entry 14
ID/s: 195	
Frequency: 1	
Proposed action: cant	
Target: life	
	Entry 15
ID/s: 241	
Frequency: 1	
Proposed action: build	
Target: houses	
	Entry 16
ID/s: 241	
Frequency: 1	
Proposed action: prevent	
Target: flashfloods	
	Entry 17
ID/s: 314	
Frequency: 1	
Proposed action: magpalit	
Target: brgy, chairman, kase	
	Entry 18
ID/s: 356	
Frequency: 1	
Proposed action: makinig	
Target: locals, ulo	
	Entry 19
ID/s: 425	
Frequency: 1	
Proposed action: mkaiwas	
Target: baha	
	Entry 20

ID/s: 455	Entry 28
Frequency: 1	
Proposed action: sumunodkapag	
Target: warning (device)	
Entry 21	
ID/s: 521	Entry 29
Frequency: 1	
Proposed action: iboto	
Target: elekson	
Entry 22	
ID/s: 539	Entry 30
Frequency: 1	
Proposed action: manghingi	
Target: tulong	
Entry 23	
ID/s: 545	Entry 31
Frequency: 1	
Proposed action: pagsasaayos	
Target: qng, anu	
Entry 24	
ID/s: 572	Entry 32
Frequency: 1	
Proposed action: minimize	
Target: corruption	
Entry 25	
ID/s: 637	Entry 33
Frequency: 1	
Proposed action: gumawa	
Target: paraan	
Entry 26	
ID/s: 637	
Frequency: 1	
Proposed action: maka	
Target: tulong	
Entry 27	
ID/s: 695	
Frequency: 1	
Proposed action: stay	
Target: home	

D.3 FastText, Organized by Response Categories

FILIPINO TEXT ANALYSIS TOOL REPORT

Mar-24-2020 21:24:38

The information below were extracted and organized automatically.

INFORMATION CAMPAIGN AND CAPACITY BUILDING

Entry 1

ID/s:

80|81|101|104|120|121|129|130|132|149|1
56|157|160|169|178|179|181|207|214|215|
218|222|234|238|245|249|254|259|266|269|
|283|287|296|303|305|331|332|333|337|34|
0|343|344|347|350|358|382|385|386|389|3
90|394|396|398|400|403|405|406|485|578|
582|583|585|588|591|595|603|610|614|627|
|630|644|647|654|658|661|667|677|681|68
7|700|707|794|801|804|811|831|838|842|8
46|860|869|893|895|903|919|929|933

Frequency: 159

Proposed action: door, be, maximize, put, alert, inform, remind, do, implement, cause, been, have, thrice, gather, is, require, give, help, prepare, comes, avoid, informs, save, inviting, conducting, having, informing, teach, affermentioned, encounter, provide, doing, talks, make, knowing, are, regarding, create, contain, discussing, conduct, needed, linawin, participating, adds, requiring, drill, held, provides, using, undergo, please, happen, leaflets, making, focusing, being, starts, participate, magparticte, reduce, ready, providing, add

Target: disaster (awareness, advantage, information, dissemination, consequence, disasters, possibilities, drill, calamities, families, typhoon, programs, training, officials, community, safety, outcomes, prevention, preparation, evacuation, orientation, program, calamity, household, activities, preparations, session, emergencies, duty, drainage, systems), sirens (citizens, drills, damage, month, members, dont, times,

share, ideas, prepare, floods, meeting, place, whoe, sitio, preparedness, check, needs, citizen, places, neighbor, subordinate, taraining, instructions, knowledge, kind, seminars, tips, orientation, garbage, disposal, member, things, signage, beforeduring, incoming, others, damages), people (help), assembly (house), barangay, family, seminar (seminars), time, representative, emergency, problems (effects), class, project, idea, lot, protocols, terminolohiyang, drive, risk, ligu, dapat, facebook, news, series, klase, lahatmgipon, center

Entry 2

ID/s:

13|26|208|209|289|339|491|507|576|580|5
98|628|649|717|735|762|766|797|811|813|
836|847|852|854|885|897|903

Frequency: 39

Proposed action: nais, dapat, malaman, gusto, kailangan, makakatulong, handa, maaari, naisip, gagawin

Target: ngmga, barangay (calamity), seminar (paaralan), tao (kalamidad, paraan), oras, volunteer (posters, drill, for, evacuation, encourage, during, after, the, weekly), prepared (like, example, about, before, and, of, conduct), kaalaman, disiplina, sakuna, paalala, pgdating, lubus, pamamagitan, center, my, mg, roong, drmm, local (government)

Entry 3

ID/s:

13|33|46|54|68|208|209|236|243|289|307|3
10|311|339|469|565|598|603|618|631|651|
696|716|758|760|762|766|787|800|854|885|
|903|925|926|930

Frequency: 39

Proposed action: maging, magkaroon, maglinis, magkaron, dadating, magsgwa, laging, mabigyan

Target: kapitbahay, tao (kalamidad), disaster (and, during, drill, program, training), mamamayan, aware (sos, regarding, kits, meeting), kabbarangay (kabaro, disciplina), darating, beforeafter (about), pra, bagyo, tragedya, seminars (seminar, seminarsdrill), kaalaman (kaalam), kanal, harap, organisasyon, iinvite, emergency, paalala, progma, ideya, weekly

Entry 4

ID/s:

129|149|152|157|169|178|200|216|249|297
|303|389|585|594|627|630|658|798|876|92
7

Frequency: 20

Proposed action: wasnt, think, giving, know, throw, handle, living, hold

Target: drill (disaster, meeting, panicking, garbages), barangay, emergency, situations (community, officials, majority, information), seminars, case (leaflets), knowledge

Entry 5

ID/s:

101|104|208|287|343|347|630|842|910|926
|931

Frequency: 12

Proposed action: preparing, pinamumunuang, establish, suggest, cleaning, nagmemeeting, sakunamaari, rumesponde

Target: typhoon (disaster), barangay, program, share (canals), araw, sakuna

Entry 6

ID/s: 33|536|565|579|599|736|811|885|903

Frequency: 9

Proposed action: mapaglalaban, magsagawa, makaiwas, makaligtas, paparating, magpatawag

Target: sakuna (skuna), disaster (drill), drills (orientation), kalamidad, pagpupulong, seminars, grupo, mg, programa, taas, pra

Entry 7

ID/s: 26|126|281|339|389|676|771

Frequency: 7

Proposed action: sanayin, staying, magturo, sana, pay, talakayin

Target: lugar, touch, first (visit), aid, awareness (programs, weekly), house, bagyo, epektu

Entry 8

ID/s: 26|362|764|813|883|926

Frequency: 7

Proposed action: may

Target: pangangailangan (kalamidad), seminar, meetings (conduct), kinalaman, darating

Entry 9

ID/s: 46|54|469|649|736|836|885

Frequency: 7

Proposed action: magbigay, bibigay, mabigay, karoon

Target: seminar, kaalaman, bahay, information, barangay

Entry 10

ID/s: 236|289|462|553|604|760

Frequency: 7

Proposed action: mag, gumawa

Target: participate (conduct, schedule, update), impormasyon, seminars, programa

Entry 11

ID/s: 243|362|598|600|618|631|651

Frequency: 7

Proposed action: magpasagawa, ginagawa, magsasagawa, tuwingat, magpasimuno

Target: disaster (drill), samin, seminar (seminars), delubyo

Entry 12

ID/s: 15|84|344|592|736

Frequency: 5

Proposed action: seminar, concerning, disseminating

Target: bagyo (lindol), disaster (information), drive, preparedness

Entry 13

ID/s: 717 766 883 897 931	Proposed action: magandang, maganda
Frequency: 5	Target: conduct, tulong, seminar
Proposed action: tamang, maayos	Entry 21
Target: kaukuluan (impormansyon, pagpapatupad), oras, desaster	ID/s: 26 394
Entry 14	Frequency: 2
ID/s: 84 88 364 846	Proposed action: pagsapit, simulating
Frequency: 4	Target: bagyobaha, disasters
Proposed action: organize, planning	Entry 22
Target: events (disaster, incase), drills, talk	ID/s: 132 842
Entry 15	Frequency: 2
ID/s: 602 751 800 847	Proposed action: join
Frequency: 4	Target: seminar, barangay
Proposed action: linisin, magkron, maiwasan, hinahanda	Entry 23
Target: daan, pagpulong, trahedyta, tao	ID/s: 208 744
Entry 16	Frequency: 2
ID/s: 13 645 930	Proposed action: nasaabing, nalaman
Frequency: 3	Target: lugar, seminar
Proposed action: ibat, ibahagi	Entry 24
Target: paraan, sona, paghahanda	ID/s: 307 628
Entry 17	Frequency: 2
ID/s: 42 169 903	Proposed action: turuan, papaalam
Frequency: 3	Target: pamilya, tao
Proposed action: pagsabihan, initiating, nalaman	Entry 25
Target: kaalaman, seminars, sakuna	ID/s: 336 800
Entry 18	Frequency: 2
ID/s: 130 178 381	Proposed action: maihnada, maganada
Frequency: 3	Target: tao, programa
Proposed action: educate, performing	Entry 26
Target: disaster (drill), people	ID/s: 462 797
Entry 19	Frequency: 2
ID/s: 218 801 831	Proposed action: natutunan, ipinatutupad
Frequency: 3	Target: disaster (survey), tagapamuno
Proposed action: go, hit	Entry 27
Target: calamities (cases), area	ID/s: 469 797
Entry 20	Frequency: 2
ID/s: 766 800	Proposed action: maiging, nagpapaalala
Frequency: 3	Target: barangay, oras
	Entry 28
	ID/s: 578 602

Frequency: 2	ID/s: 310
Proposed action: ginagamit, mkaiwas	Frequency: 1
Target: babala, storms	Proposed action: lumawak
Entry 29	Entry 37
ID/s: 797 800	ID/s: 311
Frequency: 2	Frequency: 1
Proposed action: nakikita, nangyayari	Proposed action: palaging
Target: disiplina, s	Target: gawim
Entry 30	Entry 38
ID/s: 811 910	ID/s: 383
Frequency: 2	Frequency: 1
Proposed action: pumunta, pumasok	Proposed action: t
Target: briefing, bagyo	Target: incoming, disaster
Entry 31	Entry 39
ID/s: 5	ID/s: 403
Frequency: 1	Frequency: 1
Proposed action: malawakang	Proposed action: survive
Target: information, drive	Target: disaster
Entry 32	Entry 40
ID/s: 42	ID/s: 469
Frequency: 1	Frequency: 1
Proposed action: nasasakupan	Proposed action: dating
Target: kalamidad	Target: kalamidad, drills, paraan
Entry 33	Entry 41
ID/s: 126	ID/s: 469
Frequency: 1	Frequency: 1
Proposed action: respond	Proposed action: malalamian
Target: calamities	Target: kabbarangay
Entry 34	Entry 42
ID/s: 256	ID/s: 583
Frequency: 1	Frequency: 1
Proposed action: tell	Proposed action: learn
Target: information	Target: things
Entry 35	Entry 43
ID/s: 296	ID/s: 591
Frequency: 1	Frequency: 1
Proposed action: guide	Proposed action: regular
Target: community	Target: barangay
Entry 36	

<p>Entry 44</p> <p>ID/s: 599 Frequency: 1 Proposed action: walang Target: drill</p> <p>Entry 45</p> <p>ID/s: 599 Frequency: 1 Proposed action: naigawa Target: barangay</p> <p>Entry 46</p> <p>ID/s: 627 Frequency: 1 Proposed action: stay Target: times</p> <p>Entry 47</p> <p>ID/s: 676 Frequency: 1 Proposed action: magtayo Target: programa</p> <p>Entry 48</p> <p>ID/s: 735 Frequency: 1 Proposed action: kilangan Target: magkroon</p> <p>Entry 49</p> <p>ID/s: 797 Frequency: 1 Proposed action: naglalagay Target: karatula</p> <p>Entry 50</p> <p>ID/s: 797 Frequency: 1 Proposed action: nagkakabit Target: tarpaulin</p> <p>Entry 51</p> <p>ID/s: 797 Frequency: 1 Proposed action: kakailanganin</p>	<p>Target: residente</p> <p>Entry 52</p> <p>ID/s: 842 Frequency: 1 Proposed action: attending Target: seminars</p> <p>Entry 53</p> <p>ID/s: 847 Frequency: 1 Proposed action: magkakaroon Target: orientation</p> <p>Entry 54</p> <p>ID/s: 886 Frequency: 1 Proposed action: w Target: brgy, level</p> <p>Entry 55</p> <p>ID/s: 903 Frequency: 1 Proposed action: importanteng Target: bagaypagkain</p> <p>Entry 56</p> <p>ID/s: 910 Frequency: 1 Proposed action: mabilis Target: aksyon</p> <p>Entry 57</p> <p>ID/s: 929 Frequency: 1 Proposed action: assist Target: citizen</p> <hr style="background-color: #6B8E23; border: none; height: 5px; margin: 10px 0;"/> <p style="text-align: center;">DISASTER RELIEF</p> <hr style="background-color: #6B8E23; border: none; height: 5px; margin: 10px 0;"/> <p>Entry 1</p> <p>ID/s: 110 261 299 379 421 795 828 878 Frequency: 11 Proposed action: provide, are, cannot, support, giving, be, help</p>
--	---

Target: food (kits, supply, goods, assistance), aid (need), relief, afford, response, case	Proposed action: mag Target: pagkaen, gamot
Entry 2	Entry 9
ID/s: 205 418 440 522 776	ID/s: 440
Frequency: 6	Frequency: 1
Proposed action: kailangan, handa, matulungan, dapat, ilangan	Proposed action: mailikasunahin
Target: relief, goods, resque, pagkain, sakuna, bangka, lugar, pamilya	Target: pagbibigay
Entry 3	Entry 10
ID/s: 12 460 559 566 776	ID/s: 497
Frequency: 5	Frequency: 1
Proposed action: pagbibigay, magbibigay, bibigay	Proposed action: maagap
Target: assistance (goods), pagkain (panganga), medisina, evacuation (grocery), tulong, pangangailangan	Target: tulong
Entry 4	Entry 11
ID/s: 446 906	ID/s: 526
Frequency: 2	Frequency: 1
Proposed action: papadala, magtulong	Proposed action: ibakwisyon
Target: kunting, pagbigay	Target: pagbbigay
Entry 5	Entry 12
ID/s: 420	ID/s: 548
Frequency: 1	Frequency: 1
Proposed action: gawing	Proposed action: sana
Target: pagtulong	Target: brgy
Entry 6	Entry 13
ID/s: 420	ID/s: 548
Frequency: 1	Frequency: 1
Proposed action: tinamaan	Proposed action: inaasahang
Target: kalamidad	Target: sakuna
Entry 7	Entry 14
ID/s: 420	ID/s: 795
Frequency: 1	Frequency: 1
Proposed action: may	Proposed action: i
Target: sakit	Target: government
Entry 8	
ID/s: 437	COMMUNITY-WIDE LOGISTIC SUPPORT FOR DISASTER RESPONSE
Frequency: 1	
	Entry 1
	ID/s: 95 106 109 114 134 135 142 147 154 170

Extracting and Organizing Disaster-related Philippine Community Responses for Aiding Nationwide Risk Reduction Planning and Response (N. Nocon, 2020)

171|212|221|242|247|270|276|302|329|330
|392|397|409|656|675|777|879|911|918|93
2

Frequency: 40

Proposed action: prepare, have, providing, need, needed, create, having, know, is, be, disseminate, make, contains, provide, suffered, are, help, designed, benefit, ready, mastrandied, inform

Target: place (sirens, publicsoundnotifsystem, condition, gears, kits), roadway, emergency (facility, assistance), supplies (shelterarea, evacuation, boats, equipment, advance, funds, plan, food, safety, calamity, technology, secure), disaster (facilities, typhoons, storm), area, barangay (barangay), gym, electricity, center, aid (people), kit, source, int, storage, pasahero, signal, info

Entry 2

ID/s: 40|417|488|653|810

Frequency: 5

Proposed action: magkaroon, makabii, maglagay, mag-deploy

Target: medical (evacuation, center, siren), kit, flashlight (volunteer), budget, gamit

Entry 3

ID/s: 71|417|534

Frequency: 4

Proposed action: karagdagang, dagdagan, nakakatulong

Target: kagamitan, community, volunteers, budget, barangay

Entry 4

ID/s: 167|629|777|780

Frequency: 4

Proposed action: dapat, kailangan

Target: evacuation (center), barangay, truck, bangka

Entry 5

ID/s: 242|247|378|659

Frequency: 4

Proposed action: build, buy

Target: place, evacuation (instructions), area

Entry 6

ID/s: 244|434|484

Frequency: 3

Proposed action: magpatayo, makasakay, magtapon

Target: evacuation (center), bangka, tao, basura

Entry 7

ID/s: 244|556

Frequency: 3

Proposed action: may

Target: tao, bangka, baha

Entry 8

ID/s: 488|590|810

Frequency: 3

Proposed action: gamit

Target: food, for (volunteers), darating, bumbero

Entry 9

ID/s: 142|378

Frequency: 2

Proposed action: back, giving

Target: funds (equipments, incase)

Entry 10

ID/s: 51

Frequency: 1

Proposed action: dumating

Target: dilubyobagyo

Entry 11

ID/s: 51

Frequency: 1

Proposed action: nakahanda

Target: gamit

Entry 12

ID/s: 51

Frequency: 1

Proposed action: mataas

Target: lugar

Entry 13

ID/s: 114	Entry 21
Frequency: 1	
Proposed action: armed	
Target: forces	
	Entry 14
ID/s: 154	
Frequency: 1	
Proposed action: cover	
Target: areas	
	Entry 15
ID/s: 171	
Frequency: 1	
Proposed action: etc	
Target: evacuees	
	Entry 16
ID/s: 242	
Frequency: 1	
Proposed action: serve	
Target: place	
	Entry 17
ID/s: 244	
Frequency: 1	
Proposed action: tamaan	
Target: sakuna	
	Entry 18
ID/s: 302	
Frequency: 1	
Proposed action: ex	
Target: boat	
	Entry 19
ID/s: 392	
Frequency: 1	
Proposed action: communicating	
Target: weather (predictions)	
	Entry 20
ID/s: 534	
Frequency: 1	
Proposed action: sana	
Target: gamit	
	INFRASTRUCTURE MAINTENANCE AND MANAGEMENT
	Entry 1
ID/s:	
4 43 47 52 57 69 74 78 166 231 366 442 4	
44 449 465 466 480 493 530 543 725 743	
761 779 812 819 820 822 823 825 829 853	
Frequency: 36	
Proposed action: kailangan, maglagay, ayusin, maglinis, linisin, dapat, maayos, gusto, nakakatulong	
Target: ikot, waste, dsiposal, garbage (drainage, siguraduhing, especially, lage), can, kanal, kalinisan (kapaligiran, sakapaligiran, pagbabahahindi), pagtatapon (tapunan), pangongolekta (ilogwag), daanan (kalsada), basura (basira), paligid, pausok, Igu, dn, pagtaas, imburnal	
	Entry 2
ID/s:	
94 98 107 117 127 163 198 282 308 474 5	
69 574 584 601 617 635 688 796 808 834	
841 843	
Frequency: 34	
Proposed action: improve, have, clean, fixing, implement, ensure, secure, fix, be, put, mentioned, is, maimprove, mainform, maintain, perform, avoid, cause, flooding, keep, help, prevent	
Target: drainage (areas, streets, cleaning, programs, garbage, cleanups, floods, canals, elevation, segregation, clean-up, session, rains, dumpster, bins, flooding, surroundings, garbagetrash, blockage), system, rules (disposal), flow (flood), survey, plan, factor, tao (kalinisan), community	

Entry 3

ID/s:

14|38|45|48|368|478|480|482|504|512|537|
743|779|799|820|823

Frequency: 20

Proposed action: magkaroon, maiwasan, mabawasan, magtanim, magbigay, maghada, magsawa, maging

Target: basurahan (pagbabaha, pagbara, basura, lubid, pagbabara), sariling, pagtitipon, training (seminars), barangay, pagbaha (baha), puno, paglaganap (kalamidad, sanhi), sakuna

Entry 4

ID/s:

60|279|288|493|589|648|746|790|817

Frequency: 9

Proposed action: linisan, panglinis, malinis, makaiwas, iwas, naipapataasan

Target: paligid (kalamidad, kapaligiran), pgbaha (baha), kanal, daanan, karanasan

Entry 5

ID/s:

4|38|62|476|482|504|537|822

Frequency: 8

Proposed action: wastong, kaylangan, tamang

Target: pagtatapon (pagtapon, tapunan), kanal, lugar, kapaligiran, lalagyan

Entry 6

ID/s:

48|62|482|537|560|779|790

Frequency: 8

Proposed action: nakabara, magbara, pasaway, pumping, pinondohan, bumara

Target: kanal (ilog), basura, station, taong, bayan, baha

Entry 7

ID/s:

62|449|458|504|560|589|822

Frequency: 7

Proposed action: itapon, magtapon, nagtatapon, itapun

Target: basura (ngbasura, basurahan), maayus, kanal

Entry 8

Entry 3

ID/s:

52|458|500|541

Frequency: 4

Proposed action: tanggalin, palitan, iwasan, ilagay

Target: basura, sakuna, bata, pagtatapon

Entry 9

ID/s:

57|817|905

Frequency: 3

Proposed action: walang, tanging

Target: baha, kalsada

Entry 10

ID/s:

107|282|635

Frequency: 3

Proposed action: lessen

Target: probability, garbage (chances)

Entry 11

ID/s:

127|361|617

Frequency: 3

Proposed action: install, cleaning, thrice

Target: drainage (systems, canals, week)

Entry 12

ID/s:

478|779|807

Frequency: 3

Proposed action: regular

Target: basura, pagtatapon, paglilinis, garbagewaste (collections)

Entry 13

ID/s:

809|817|905

Frequency: 3

Proposed action: gumawa, mag

Target: programa, obserba, kanal

Entry 14

ID/s:

4|482

Frequency: 2

Proposed action: bantayan

Target: gamit, kabataang

Entry 15

ID/s:

288|589

Frequency: 2

Proposed action: nagtulung, pagtulungan

Target: kapaligiran, lugar	Proposed action: ngebbay
Entry 16	Entry 24
ID/s: 14	ID/s: 458
Frequency: 1	Frequency: 1
Proposed action: pangunahing	Proposed action: daming
Target: sanhi	Target: tindahan
Entry 17	Entry 25
ID/s: 32	ID/s: 458
Frequency: 1	Frequency: 1
Proposed action: mapabilis	Proposed action: gitna
Target: paglilinis	Target: kalsada
Entry 18	Entry 26
ID/s: 48	ID/s: 465
Frequency: 1	Frequency: 1
Proposed action: panatilihin	Proposed action: maipon
Target: kalinisan	Target: tubig
Entry 19	Entry 27
ID/s: 69	ID/s: 482
Frequency: 1	Frequency: 1
Proposed action: ilabas	Proposed action: dagdagan
Target: basura	Target: bantay
Entry 20	Entry 28
ID/s: 69	ID/s: 482
Frequency: 1	Frequency: 1
Proposed action: dumating	Proposed action: gagawing
Target: truck	Target: hagdan
Entry 21	Entry 29
ID/s: 279	ID/s: 482
Frequency: 1	Frequency: 1
Proposed action: tumulong	Proposed action: makasagip
Target: kagamitan	Target: bata
Entry 22	Entry 30
ID/s: 372	ID/s: 589
Frequency: 1	Frequency: 1
Proposed action: pick	Proposed action: ikakabuti
Target: trashes	Target: kanal
Entry 23	Entry 31
ID/s: 454	ID/s: 601
Frequency: 1	

Frequency: 1	ID/s: 843
Proposed action: eg	Frequency: 1
Target: rivers	Proposed action: ensuring
Entry 32	
ID/s: 807	Target: surrounding
Frequency: 1	
Proposed action: delay	
Target: cause	
Entry 33	
ID/s: 812	
Frequency: 1	
Proposed action: khit	
Target: lakas	
Entry 34	
ID/s: 812	
Frequency: 1	
Proposed action: malaki	
Target: kanal, basura	
Entry 35	
ID/s: 817	
Frequency: 1	
Proposed action: namen	
Target: epeko	
Entry 36	
ID/s: 817	
Frequency: 1	
Proposed action: napupunta	
Target: lugar	
Entry 37	
ID/s: 817	
Frequency: 1	
Proposed action: nagiging	
Target: dahilan	
Entry 38	
ID/s: 823	
Frequency: 1	
Proposed action: taasan	
Target: kalsada	
Entry 39	
EARLY WARNING SYSTEM	
Entry 1	
ID/s:	
66 90 97 102 103 105 111 112 125 128 13	
3 137 139 140 146 172 175 176 189 223 2	
24 235 237 250 264 272 273 277 290 291	
292 300 304 309 313 319 334 354 373 388	
431 519 547 573 575 619 664 671 683 68	
4 686 708 710 739 782 814 837 844 857 8	
59 874 909 917 921	
Frequency: 108	
Proposed action: maiaannounce, ma-inform, regarding, do, improving, make, improved, minimize, be, have, inform, messaging, providing, coming, prepare, announcing, reaching, telling, texting, is, incoming, updated, giving, alert, alerting, are, suggest, has, eat, caused, visit, prepared, accommodate, allow, participating, dealing, spread, ialarm, namemegaphone, update, iinform, upcoming, give, evacuate, knowing, avoid, said, utilize, disseminate, ipabatid, ma-spread, magprovide, provide	
Target: balita, dilubyong (discipline, ilog), disaster (announcements, imformation, dissemination, announcement, communication, training, technology, safety, community, warning, typhoon, residents, disasters, food, announce, calamities, public, announcements, newsinformation, updates, awareness, events, exits), theft (plenty, citizens, places, beforehand, floods, incoming, troubles, people, preparedness, sms, alert, constituents, news, person, everyone, theeveryone, damage, way, needs, need, weather, changes, things, loss, early, happenings, weeks), media (devices), etc, time, sign, barangay, palagi, house (earlier), brgy, megaphones, tao, tips, area, alarm, na, emergency	

Entry 2

ID/s:

3|6|7|8|11|20|27|50|66|77|335|413|419|467
|473|550|555|557|558|749|774|789

Frequency:

Proposed action: magkaroon, maging, pagkakaroon, magbibigay, mabigyan, nagbbigay, bigyang, makatulong, magbigay, ipagbigay

Target: komunikasyon (impormasyon), early (warning, quarterly, speaker), system (maintenance), council (members), tao (kalamidad, bahaa), notice (device), babala, paglaki, barangay

Entry 3

ID/s:

50|58|90|204|306|436|448|495|529|554|82
1|891|894

Frequency:

Proposed action: malinaw, dapat, mailikas
Target: koordinasyon, sakuna, eg, komunikasyon, kita, mamamayan, pamamagitan, anunsiyo, baragay (barangay, baranggay), alert (abiso), balita, radio, tubig, saknila, update, my, darating

Entry 4

ID/s:

20|263|450|481|712|774|814|833|857|914

Frequency:

Proposed action: may
Target: kalamidad (calamidad), sakunang (sakuna), paalam, bagyo (baha, bagyono), assembly, iparating, anu, disaster, komunikasyon, suporta, pra

Entry 5

ID/s: 8|30|263|286|431|445|724|833

Frequency:

Proposed action: masigurong, mapalaganap, inaasahan, maitutulong, ingat, magalarm, masabihan, etepibo
Target: kalamidad (tao), inpormasyon, anunsyo, barangay, level, sirena, n, kng

Entry 6

ID/s: 58|436|438|535|557|558|685|788

Frequency:

Proposed action:

mag, makipag
Target: coordinate (warning, announce), signal, detalye, report, bagyo, ugnayan

Entry 7

ID/s: 6|65|547|554|724|765

Frequency:

Proposed action:

dumating, lumikas, tumira, iparating
Target: bagyo, kalamidad, sakuna, bgyan, talang, evacuoation (center), s

Entry 8

ID/s: 8|30|75|448|490|698

Frequency:

Proposed action:

pagtibayin, paginform, pagtawag, magtawag, paqaabiso, pagpapa-alala
Target: early (warning), system, device, chairman, anak, baranggay, paqdating, sakuna

Entry 9

ID/s: 278|697|727|772|856|914

Frequency:

Proposed action:

maagang, maaga, maagap
Target: announcement (if, ever), impormasyon, kalamidad, anunsyo, palang

Entry 10

ID/s: 67|263|697|727|837

Frequency:

Proposed action:

paparating, darating
Target: sakuna, kalamidad, na, bigay, lugar

Entry 11

ID/s: 75|448|550|814|844

Frequency:

Proposed action:

kailangan, handa, malaman, makatulong
Target: tahanan, better (way), talad, pangangailangan, barangay

Entry 12

ID/s: 3|20|257|306

Frequency:

Proposed action: magkikita, makita, nagkulang, gagawin	Target: barangay
Target: panahon, sitwasyon, pagbibigay, babala, disaster	Entry 20
	ID/s: 413 419
	Frequency: 2
	Proposed action: wang
	Target: wang
	Entry 21
	ID/s: 413 419
	Frequency: 2
	Proposed action: ipagpatuluy
	Target: barangay, pagbigay, babala
	Entry 22
	ID/s: 445 727
	Frequency: 2
	Proposed action: malaki, malakas
	Target: tubig, n, bagyo
	Entry 23
	ID/s: 724 789
	Frequency: 2
	Proposed action: mababang, mataas
	Target: lugar
	Entry 24
	ID/s: 85
	Frequency: 1
	Proposed action: warning
	Target: signal
	Entry 25
	ID/s: 164
	Frequency: 1
	Proposed action: galingan
	Target: gid
	Entry 26
	ID/s: 176
	Frequency: 1
	Proposed action: built
	Target: speakers
	Entry 27
	ID/s: 176
	Frequency: 1
	Proposed action: roam

Proposed action: drive	Frequency: 1
Target: barangay, days	Proposed action: aga
Entry 28	Entry 36
ID/s: 286	ID/s: 698
Frequency: 1	Frequency: 1
Proposed action: ipamahagi	Proposed action: walang
Target: impormasyon	Target: sawang
Entry 29	Entry 37
ID/s: 286	ID/s: 713
Frequency: 1	Frequency: 1
Proposed action: mangyaring	Proposed action: makinig
Target: sakuna	Target: advisory
Entry 30	Entry 38
ID/s: 413	ID/s: 724
Frequency: 1	Frequency: 1
Proposed action: kpag	Proposed action: nag
Target: tubig	Target: anouse, merng
Entry 31	Entry 39
ID/s: 481	ID/s: 724
Frequency: 1	Frequency: 1
Proposed action: kapit	Proposed action: umiikot
Target: bahay	Target: tao
Entry 32	Entry 40
ID/s: 490	ID/s: 789
Frequency: 1	Frequency: 1
Proposed action: lagi	Proposed action: papalapit
Target: atpagbigay	Target: bagyo, baranggay
Entry 33	Entry 41
ID/s: 535	ID/s: 833
Frequency: 1	Frequency: 1
Proposed action: sabinin	Proposed action: magiging
Target: bagyong	Target: aware
Entry 34	Entry 42
ID/s: 571	ID/s: 859
Frequency: 1	Frequency: 1
Proposed action: facebook	Proposed action: watch
Target: twitter	Target: news
Entry 35	Entry 43
ID/s: 685	

ID/s: 894
Frequency: 1
Proposed action: mgbibigay
Target: signal

Entry 44

ID/s: 914
Frequency: 1
Proposed action: naapektuhan
Target: bagyo

PREPAREDNESS FOR EMERGENCY

Entry 1

ID/s:
16|19|35|61|72|122|321|341|365|439|447|492|502|616|626|660|673|690|714|722|728|731|742|747|748|752|753|783|785|786|806|815|824|830|889|908|920

Frequency: 40

Proposed action: maging, magtulong, magkaroon, maghanda, malaman, kumplitohin, mapagbigay-alam, maka, pagiging, magiging, encouraging
Target: bagay (kalamidad, kalamid, kalikasan, kasayan, kalagayan, mangyari), bagyobaha, oras, sakuna, pra, alertoat (cla, pkiking, participation, prepared), handa, atentibo (kagamatinan), balita, daratinga (darating), brgy, opisyal, xa, tulong, gamit, panahon, sariling, disciplina, usap, kapitbahay (pagkain), alis, anu, emergency, plan, barangay, people

Entry 2

ID/s:
99|123|145|158|159|165|199|233|248|284|312|324|380|623|625|672|680|704|738|755|816|840|863|864|868|915|916|920

Frequency: 38

Proposed action: doing, be, check, know, is, listen, regarding, help, comes, watching, prepare, alert, being, canned, med, inform, need, go, have, helps, keep, clothing, take

Target: flood (weather, happenings, condition, things, calamity, preparedness, food, batteries, kits, kit), prone (areas, officials, evacuation, plan, television, disaster, solution, needs, proper, case, reduction, management, council, community, funds, goods), news (week, radio, times), typhoon, time, barangay, risk, unit, emergency, etc, seminar

Entry 3

ID/s:
73|148|151|203|211|315|341|342|506|510|612|673|767|793|851|890|892|896|899|900|901|912|922

Frequency: 25

Proposed action: dapat, pwede, sakunadapat
Target: barangay (baranggay), disaster (updated, official, program), balita (ulat), aware, oras, araw, alirto, membro, i (check), pra, programa, sakuna, ee

Entry 4

ID/s: 61|122|365|492|506|510|613|748|753

Frequency: 9

Proposed action: may

Target: sakuna (sakunang), bagyo, kapaligiran, sakona, pomoang, barangay (kabaranggay), nkalagay, darating

Entry 5

ID/s: 22|36|506|528|533|562|803|818

Frequency: 8

Proposed action: maayos, magamit, matugunan, kailangan

Target: paghahanda, comunicasion (alert), pangangailangan (kaligtasan), radyo, ata, always, disciplina, mangyare

Entry 6

ID/s: 37|56|280|624|775|824|850

Frequency: 8

Proposed action: kaylangan, pupunta, dadaan, kakailanganin, mahalaga, nangangailangan, sapat

Target: aware, mamamayan, oras, barangay, panahon, stocks, interaksyon, pagkain, pangangailangan

<p>Entry 7</p> <p>ID/s: 23 56 528 533 544 818 Frequency: 7 Proposed action: dumating, pagdating, pumunta Target: sakuna, kalamidad, obserbatibo, safe</p>	<p>Entry 14</p> <p>ID/s: 652 752 924 Frequency: 3 Proposed action: tumulong, ipagbigay, simulan Target: pagkalap, mamamayan, baranggay</p>
<p>Entry 8</p> <p>ID/s: 36 642 726 786 806 881 899 Frequency: 7 Proposed action: ihanda, handa, makaiwas, mapaghandaan, paghahanda Target: pweding, sakuna, barangay, tulong, kalamidad, karapat</p>	<p>Entry 15</p> <p>ID/s: 23 816 Frequency: 2 Proposed action: maagang, lessen Target: paghahanda, scale</p>
<p>Entry 9</p> <p>ID/s: 315 452 463 613 624 729 850 Frequency: 7 Proposed action: mag Target: food (supplies), kagamitang, monitor, samahan, plano, laan, gamot, imbak</p>	<p>Entry 16</p> <p>ID/s: 72 532 Frequency: 2 Proposed action: nakatuon, pagkilos Target: weather (forecast), kalamidad</p>
<p>Entry 10</p> <p>ID/s: 203 452 722 818 896 Frequency: 5 Proposed action: nangyayari, binabaha, mangvare, nagaganap Target: paligid, tao, darateng, bagyo, barangay</p>	<p>Entry 17</p> <p>ID/s: 159 228 Frequency: 2 Proposed action: watch Target: weather (forecast, news)</p>
<p>Entry 11</p> <p>ID/s: 690 729 890 912 Frequency: 4 Proposed action: alerto Target: bagay (kalamidada), tv, darating</p>	<p>Entry 18</p> <p>ID/s: 622 626 Frequency: 2 Proposed action: makinig Target: anunyo, balita</p>
<p>Entry 12</p> <p>ID/s: 73 769 892 Frequency: 3 Proposed action: wala, meron Target: kalamidad, paghahanda, pra</p>	<p>Entry 19</p> <p>ID/s: 748 908 Frequency: 2 Proposed action: mabilis Target: bahay, sakuna</p>
<p>Entry 13</p> <p>ID/s: 211 447 756 Frequency: 3 Proposed action: magpadala, makipagtutulungan, ipahatid</p>	<p>Entry 20</p> <p>ID/s: 924 Frequency: 2 Proposed action: naghihirap, laganap Target: bansa (dahilbansa), kurapsyon</p>
	<p>Entry 21</p> <p>ID/s: 10</p>

Frequency: 1	ID/s: 731
Proposed action: sinasabi	Frequency: 1
Target: sakuna	Proposed action: ibahagi
Entry 22	Target: nl
ID/s: 16	ID/s: 731
Frequency: 1	Frequency: 1
Proposed action: opisyal	Proposed action: ibalita
Target: brgy	Target: s
Entry 23	Entry 30
ID/s: 122	ID/s: 748
Frequency: 1	Frequency: 1
Proposed action: palaging	Proposed action: magki2ta
Target: balita	Target: oras
Entry 24	Entry 31
ID/s: 211	ID/s: 748
Frequency: 1	Frequency: 1
Proposed action: tignan	Proposed action: maiwasan
Target: bagyo	Target: pagka, trap
Entry 25	Entry 32
ID/s: 280	ID/s: 756
Frequency: 1	Frequency: 1
Proposed action: bumuo	Proposed action: alisto
Target: emergency, plansan	Target: n, balita
Entry 26	Entry 33
ID/s: 284	ID/s: 756
Frequency: 1	Frequency: 1
Proposed action: radio	Proposed action: gagawin
Target: disaster	Target: my
Entry 27	Entry 34
ID/s: 533	ID/s: 756
Frequency: 1	Frequency: 1
Proposed action: ilagay	Proposed action: sakin
Target: bgo	Target: kana
Entry 28	Entry 35
ID/s: 616	ID/s: 769
Frequency: 1	Frequency: 1
Proposed action: iimbak	Proposed action: nasasakupan
Target: pagkain	Target: problemang
Entry 29	Entry 36

<p>Entry 37</p> <p>ID/s: 896 Frequency: 1 Proposed action: darating Target: barangay</p> <p>Entry 38</p> <p>ID/s: 916 Frequency: 1 Proposed action: eg Target: goods (battery)</p> <p>Entry 39</p> <p>ID/s: 924 Frequency: 1 Proposed action: malaking Target: problema</p>	<p>Entry 3</p> <p>ID/s: 82 471 518 643 845 870 Frequency: 6 Proposed action: magtutulongan, ligtas, nangaylangan, maiwasan, utilizing, nakkatulong Target: paglilinis (paraan), tulong, sakuna, tanod, tao, regarding</p> <p>Entry 4</p> <p>ID/s: 64 471 472 904 Frequency: 4 Proposed action: may, mayrong Target: sakuna (kapansanan), tagapamahala, kalamidad</p> <p>Entry 5</p> <p>ID/s: 196 201 210 318 Frequency: 4 Proposed action: want, happen, know Target: barangay, community (part), news</p> <p>Entry 6</p> <p>ID/s: 28 441 471 Frequency: 3 Proposed action: maipaabot, malaman, maibahagi Target: serbisyo, lugar, tao</p> <p>Entry 7</p> <p>ID/s: 82 835 913 Frequency: 3 Proposed action: gumawa, mag Target: organisasyon, mg, pundo</p> <p>Entry 8</p> <p>ID/s: 457 471 Frequency: 3 Proposed action: maging Target: barangay, tagapag, tao</p> <p>Entry 9</p> <p>ID/s: 441 477 Frequency: 2 Proposed action: sana, andyan Target: kalamidad, bagyo</p>
---	---

Entry 10	Target: opisyal
ID/s: 471 904	Entry 18
Frequency: 2	
Proposed action: nag	ID/s: 451
Target: pamagitna, kalaminad	Frequency: 1
	Proposed action: maagap
	Target: pagtulong
Entry 11	Entry 19
ID/s: 770 904	
Frequency: 2	ID/s: 471
Proposed action: mlman, manguna	Frequency: 1
Target: dpt, gwen, palaganap	Proposed action: paman
	Target: tao
Entry 12	Entry 20
ID/s: 28	
Frequency: 1	ID/s: 472
Proposed action: mabuti	Frequency: 1
Target: barangay, problema	Proposed action: bigyan
	Target: pansin
Entry 13	Entry 21
ID/s: 64	
Frequency: 1	ID/s: 472
Proposed action: ayusin	Frequency: 1
Target: patakaran	Proposed action: malapit
	Target: tabing
Entry 14	Entry 22
ID/s: 64	
Frequency: 1	ID/s: 699
Proposed action: mabilis	Frequency: 1
Target: pagbibigay	Proposed action: pwersahan
	Target: tao
Entry 15	Entry 23
ID/s: 201	
Frequency: 1	ID/s: 870
Proposed action: hurt	Frequency: 1
Target: typhoon	Proposed action: paparating
	Target: sakuna
Entry 16	Entry 24
ID/s: 202	
Frequency: 1	ID/s: 904
Proposed action: respond	Frequency: 1
Target: emergency	Proposed action: punong
	Target: brgy
Entry 17	Entry 25
ID/s: 441	
Frequency: 1	ID/s: 904
Proposed action: puntahan	Frequency: 1

Proposed action: dumating
Target: kuminidad

FILIPINO VALUES

Entry 1

ID/s:

21|49|55|161|367|423|428|433|443|523|52
5|531|540|551|564|570|641|691|694|715|
23|768|832

Frequency: 25

Proposed action: tumulong, tutulong, magtulong, matulungan, magtulungan, makatulong, magtulong, tulongan, hikayatin, makipagtulongan, maitutulong, tutulongan, magtulongan, makakatulong

Target: bahay, pagkain (paghahanda), bagyo, community, volunteer, bata (hnda), tulong, barangay (kabbarangay, baranggay, barangay), pra, kalamidad (tao), s, komunidad, esatesa, kapit, n, magkakapamilya

Entry 2

ID/s:

2|251|422|433|459|486|498|503|551|552|6
41|666|759

Frequency: 15

Proposed action: mag, makipag, magbuo

Target: pagkakaiisa (kaisa, tao), malasakit, samahan, volunteer (duty), barangay (baranggay), peace (my), and, order, gaeaing, tahanan, suporta, tulong

Entry 3

ID/s: 31|39|70|367|563|568|757|888

Frequency: 9

Proposed action: dapat, kailangan, handa

Target: pakiki-isa (komunidad), panahon, mgkaisa, ka-barangay, volunteer, canal, tulong, oras

Entry 4

ID/s: 168|262|298|393|443|861

Frequency: 9

Proposed action: being, have, helping, needs, cannot, naperwisyo, be

Target: disaster (volunteer, announcements), share (others, reminders, trick, heart), baha, barangay

Entry 5

ID/s: 1|17|53|63|427|508|720|832

Frequency: 8

Proposed action: magkaisa, makiisa, nagkakaisa, magsama

Target: tao, tungkulin, plano, paghahanda, komunidad, ren, sama, lng

Entry 6

ID/s: 31|44|426|715|832

Frequency: 6

Proposed action: may, nagkakaroon

Target: sakunang, bagyo, kalamidadwag (kalamidad), lugar

Entry 7

ID/s: 41|443|621|721|730|750

Frequency: 6

Proposed action: maging, maka, makapaghanda

Target: darating, kalagayan (kalamidad), agrisibo (sakuna), s

Entry 8

ID/s: 260|355|410|415|483

Frequency: 5

Proposed action: matutong, matutunang, nangangailangan, kaylangan, makatulong

Target: malasakit, oras, atensyon, isat, lugar

Entry 9

ID/s: 21|410|641

Frequency: 3

Proposed action: napinsala, apektado, nasiraan

Target: bagyo, kalamidad, pansiin, bahay

Entry 10

ID/s: 367|411|538

Frequency: 3

Proposed action: nangyayaring, magmalasakit, inaasahan	Entry 18
Target: barangay, kawpa, tulong	
Entry 11	
ID/s: 453 567 898	
Frequency: 3	
Proposed action: mkikipgtolongan, makikipag	
Target: programa, baranggay, tulunganat	
Entry 12	
ID/s: 31 538	
Frequency: 2	
Proposed action: dumating, nakarating	
Target: kalamidad	
Entry 13	
ID/s: 520 858	
Frequency: 2	
Proposed action: malinis	
Target: kpaligirantumulong (kapaligiran)	
Entry 14	
ID/s: 666 759	
Frequency: 2	
Proposed action: karatig, kalapit	
Target: bayan, bahay	
Entry 15	
ID/s: 9	
Frequency: 1	
Proposed action: lumawat	
Target: pagmamalasakit	
Entry 16	
ID/s: 410	
Frequency: 1	
Proposed action: pantay-pantay	
Target: pamamahagi	
Entry 17	
ID/s: 410	
Frequency: 1	
Proposed action: mabigyang	
Target: solusyon	
Entry 18	
ID/s: 412	
Frequency: 1	
Proposed action: wala	
Target: barangay	
Entry 19	
ID/s: 412	
Frequency: 1	
Proposed action: mali	
Target: kelangan	
Entry 20	
ID/s: 430	
Frequency: 1	
Proposed action: ibigay	
Target: tulong	
Entry 21	
ID/s: 443	
Frequency: 1	
Proposed action: umasa	
Target: barangaymas	
Entry 22	
ID/s: 443	
Frequency: 1	
Proposed action: gawing	
Target: serbisyo, gsyia	
Entry 23	
ID/s: 487	
Frequency: 1	
Proposed action: dagdagan	
Target: pakikisalamuha	
Entry 24	
ID/s: 494	
Frequency: 1	
Proposed action: palagiang	
Target: barangay	
Entry 25	
ID/s: 564	
Frequency: 1	
Proposed action: maayos	

Target: klamidad	sign, help, aware, typhoon, amd, barangay, world
Entry 26	Entry 2
ID/s: 691	ID/s: 115 118 183 193 195 695
Frequency: 1	Frequency: 6
Proposed action: maitayo	Proposed action: putting, seen, calling, cant, avoid
Target: kagamitan	Target: pockets (disaster), typhoon, barangay, community, life, object
Entry 27	Entry 3
ID/s: 715	ID/s: 455 791 902 923
Frequency: 1	Frequency: 6
Proposed action: magtaponi	Proposed action: dapat, nararapat, handa
Target: basura	Target: tao, barangay, pondo, pra
Entry 28	Entry 4
ID/s: 740	ID/s: 314 501 542
Frequency: 1	Frequency: 3
Proposed action: tumolng	Proposed action: magpalit, ipagpatuloy
Target: n	Target: brgy, chairman, kase, barangay, gawain
Entry 29	Entry 5
ID/s: 740	ID/s: 314 791 826
Frequency: 1	Frequency: 3
Proposed action: naabutan	Proposed action: wala, walang
Target: baha	Target: pakielam, kwenta (kwentang), bagay
Entry 30	Entry 6
ID/s: 759	ID/s: 425 539 719
Frequency: 1	Frequency: 3
Proposed action: maiwasan	Proposed action: humingi, manghingi, maibigay
Target: baha	Target: tulung (tulong), barangay
<hr/>	
OTHERS	
<hr/>	
Entry 1	Entry 7
ID/s:	ID/s: 479 637 802
184 185 187 188 192 193 194 195 197 241	Frequency: 3
Frequency: 17	Proposed action: mag, gumawa
Proposed action: know, cannot, is, help, has, please, be, are, suggest, want, better, give, let, prevent	Target: kona, tao (paraan)
Target: deaf (question, teach, people, prepare, neighborhoods, flashfloods), train (flood, permission), emergency, responders,	Entry 8
	ID/s: 513 611 637
	Frequency: 3
	Proposed action: maging, maka

Target: pamaraan, agrisibo, tulong	Proposed action: makinig
Entry 9	Entry 17
ID/s: 184	ID/s: 425
Frequency: 2	Frequency: 1
Proposed action: communicate	Proposed action: mkaiwas
Target: deaf (citizens), responders	Target: baha
Entry 10	Entry 18
ID/s: 455 791	ID/s: 521
Frequency: 2	Frequency: 1
Proposed action: sumunodkapag, binuboto	Proposed action: iboto
Target: warning, device, tas (corrupt), nag	Target: eleksyon
Entry 11	Entry 19
ID/s: 501 791	ID/s: 545
Frequency: 2	Frequency: 1
Proposed action: nakikita, apektado	Proposed action: pagsasaayos
Target: q, tang, ina	Target: qng, anu
Entry 12	Entry 20
ID/s: 501 791	ID/s: 572
Frequency: 2	Frequency: 1
Proposed action: maganda, maayos	Proposed action: minimize
Target: gawain, komunidad	Target: corruption
Entry 13	Entry 21
ID/s: 637 791	ID/s: 695
Frequency: 2	Frequency: 1
Proposed action: sana, sanang	Proposed action: stay
Target: negosyo, dhil, khit	Target: home
Entry 14	Entry 22
ID/s: 191	ID/s: 719
Frequency: 1	Frequency: 1
Proposed action: pls	Proposed action: iabot
Target: i	Target: barangay
Entry 15	Entry 23
ID/s: 241	ID/s: 791
Frequency: 1	Frequency: 1
Proposed action: build	Proposed action: maliit
Target: houses	Target: pamilya
Entry 16	Entry 24
ID/s: 356	ID/s: 826
Frequency: 1	

Frequency: 1
Proposed action: gamitin
Target: pondo

Entry 25

ID/s: 923
Frequency: 1
Proposed action: magbigay
Target: impormasyon

Entry 26

ID/s: 923
Frequency: 1
Proposed action: may
Target: bagyo

Appendix E

Informed Consent Form

**EXTRACTING AND ORGANIZING DISASTER-RELATED PHILIPPINE COMMUNITY
RESPONSES FOR AIDING NATIONWIDE RISK REDUCTION PLANNING AND RESPONSE**

Informed Consent Form

Philippine-California Advanced Research Institute: Malasakit Team Members

Nicco Louis S. Nocon
College of Computer Studies, De La Salle University, Manila

PURPOSE OF THE STUDY

You are being invited to take part in a research study. Before you decide to participate in this study, it is important that you understand why the research is being done and what your participation will involve. Please read the following information carefully and feel free to ask the researcher if there is anything that is not clear or if you need more information.

The purpose of the study is to automatically extract key insights from community responses and organize that information to be used as a list recommendations or guide in improving the nation's disaster risk reduction planning and response strategies. As developers of Malasakit, you are chosen as respondents to this research who will assess the content of the generated report and/or the developed software tool.

STUDY PROCEDURES

Initially, materials for assessment will be provided to you – these are the report, tool and supporting document/s such as the tool's function list. In assessing the report, you are to provide your opinions regarding its content quality, design or formatting, usability, satisfaction, and organizational preference. In assessing the tool, questions included are about its usefulness, ease of use, ease of learning, satisfaction, and functionalities used. At the latter portion of both forms, comments and suggestions on how to improve them and negative aspects that you have discovered will be asked.

DURATION

Answering either or both forms are good for less than thirty (30) minutes. Unless necessary, there is no need for a follow-up regarding your answers.

VOLUNTARY PARTICIPATION

Your participation in this study is voluntary. It is up to you whether or not you decide to participate. If you decide to participate, you will be asked to sign this consent form. After you sign this consent form, you are still free to withdraw at any time and without giving a reason. Withdrawing from this study will not affect the relationship you have, if any, with the researcher. If you withdraw from the study before data collection is completed, your data will be destroyed.

RISKS

There are no foreseeable risks involved in answering the questionnaires. Your feedback will not be disclosed with your personal information. However, if there are questions that you may feel uncomfortable with answering, you may decline to answer any or all questions and you may withdraw your participation at any time if you choose.

BENEFITS

Initially, your feedback will determine the impact of the research, specifically its overall usability. By participating in this study, you will be able to provide ideas for improving the report and tool. The final product of this research would be used by decision makers in strategizing and executing the nation's disaster risk reduction prevention and mitigation. Successfully utilizing the report will improve the status of your community in handling disasters and actions will be evident among society as a whole. Moreover, the success of this research will benefit you as it would expand the applications of your work. Having said that, researchers and developers in particular, will be able to discover and make use of the tool in creating more applications.

CONFIDENTIALITY

Your responses in this research will be anonymous. Every effort will be made by the researcher to preserve your confidentiality, including the following: personal identifiable information will not be included in research notes and documents, and as for this consent form, it will be stored and accessed digitally in a device that only the principal investigator has access to.

CONTACT INFORMATION

This study was approved by the Research Ethics Review Committee of De La Salle University. If you have any questions at any time about this study, or if you experience any non-normative sensations as a result of participation, you may contact the researcher whose contact information is on the first page. If you have any questions regarding your rights as a research participant, or if problems arise which you do not feel you can discuss with the Principal Investigator, please feel free to contact the Director of the Research Ethics Office, Dr. Nelson B. Arboleda, Jr., at REO@dlsu.edu.ph or by calling (632) 524-4611 local 513.

CONSENT

I have read the provided information, or it has been read to me. I have had the opportunity to ask questions about it and any questions I have been asked have been answered to my satisfaction. I understand that I will be given a copy of this form, and the researcher will keep another copy on file. I consent voluntarily to be a participant in this study.

Print Name of Participant _____

Signature of Participant _____

Date _____

Day/month/year

Print Name of Researcher Nicco Louis S. Nocon

Signature of Researcher _____

Date _____

Day/month/year

Appendix F

Survey Forms

FILIPINO TEXT ANALYSIS TOOL FOR DISASTERS: REPORT SURVEY					
Value Representation 1- Strongly Disagree, 2- Disagree, 3- Neutral, 4- Agree, 5- Strongly Agree					
1. Aesthetics (design): The report's design is acceptable	<input type="checkbox"/>				
2. Content (quality): The different elements of the report (e.g., title, insights list, and Malasakit response list) is necessary	<input type="checkbox"/>				
3. Content (quality): The information fields are appropriate and enough to make a decision or action	<input type="checkbox"/>				
4. Readability (design): The information is clear and readable	<input type="checkbox"/>				
5. Understandability (format): The information is easy to interpret	<input type="checkbox"/>				
6. Usefulness of Information (extraction quality): The information is useful in my job	<input type="checkbox"/>				
7. Organization (organization quality): The information is displayed in an organized manner	<input type="checkbox"/>				
8. Usability (efficiency): Using the report in my job would enable me to accomplish tasks more quickly	<input type="checkbox"/>				
9. Usability (potential): The report would enhance my effectiveness on the job	<input type="checkbox"/>				
10. Report (overall quality): I can make decision/s based on the information provided	<input type="checkbox"/>				
11. User Satisfaction (overall measurement): I am satisfied with the report	<input type="checkbox"/>				
Organization Preference					
<p><i>Do you prefer organizing the information by frequency (all entries and ranked by frequency only), per response categories (grouped by categories first then ordered by frequency), or something else?</i></p>					
Comments and Suggestions					
Negative Aspects					

FILIPINO TEXT ANALYSIS TOOL FOR DISASTERS: API SURVEY					
Value Representation					
1- Strongly Disagree, 2- Disagree, 3- Neutral, 4- Agree, 5- Strongly Agree					
Usefulness					
1. It is effective partnered with other software tools	1 [] 2 [] 3 [] 4 [] 5 []				
2. It can help raise the productivity rate when used in conjunction with other software tools	1 [] 2 [] 3 [] 4 [] 5 []				
3. It is useful for my tasks	1 [] 2 [] 3 [] 4 [] 5 []				
4. It makes it easier to accomplish my tasks	1 [] 2 [] 3 [] 4 [] 5 []				
5. It helps save time	1 [] 2 [] 3 [] 4 [] 5 []				
6. It does everything I would expect it to do	1 [] 2 [] 3 [] 4 [] 5 []				
Ease of Use					
1. It is easy to use	1 [] 2 [] 3 [] 4 [] 5 []				
2. It is simple to use	1 [] 2 [] 3 [] 4 [] 5 []				
3. It is user friendly	1 [] 2 [] 3 [] 4 [] 5 []				
4. It requires the fewest steps possible to accomplish what I want to do with it	1 [] 2 [] 3 [] 4 [] 5 []				
5. It is flexible	1 [] 2 [] 3 [] 4 [] 5 []				
6. Using it is effortless	1 [] 2 [] 3 [] 4 [] 5 []				
7. I can use it without written instructions	1 [] 2 [] 3 [] 4 [] 5 []				
8. I don't notice any inconsistencies as I use it	1 [] 2 [] 3 [] 4 [] 5 []				
9. Both occasional and regular users would like it	1 [] 2 [] 3 [] 4 [] 5 []				
10. I can recover from mistakes quickly and easily	1 [] 2 [] 3 [] 4 [] 5 []				
11. I can use it successfully every time	1 [] 2 [] 3 [] 4 [] 5 []				
Ease of Learning					
1. I learned to use it quickly	1 [] 2 [] 3 [] 4 [] 5 []				
2. I easily remember how to use it	1 [] 2 [] 3 [] 4 [] 5 []				
3. Learning how to use it is easy	1 [] 2 [] 3 [] 4 [] 5 []				

4. I quickly became skillful with using it	<input type="checkbox"/> 1 [] <input type="checkbox"/> 2 [] <input type="checkbox"/> 3 [] <input type="checkbox"/> 4 [] <input type="checkbox"/> 5 []
Satisfaction	
1. I am satisfied with it	<input type="checkbox"/> 1 [] <input type="checkbox"/> 2 [] <input type="checkbox"/> 3 [] <input type="checkbox"/> 4 [] <input type="checkbox"/> 5 []
2. I would recommend it to be used in the future	<input type="checkbox"/> 1 [] <input type="checkbox"/> 2 [] <input type="checkbox"/> 3 [] <input type="checkbox"/> 4 [] <input type="checkbox"/> 5 []
3. It works the way I want it to work	<input type="checkbox"/> 1 [] <input type="checkbox"/> 2 [] <input type="checkbox"/> 3 [] <input type="checkbox"/> 4 [] <input type="checkbox"/> 5 []
Functions Used (Functions that will most likely be useful to me)	
Comments and Suggestions	
Negative Aspects	

Appendix G

Survey Results and Feedback

FILIPINO TEXT ANALYSIS TOOL FOR DISASTERS: API SURVEY					
Value Representation					
1- Strongly Disagree, 2- Disagree, 3- Neutral, 4- Agree, 5- Strongly Agree					
Usefulness					
1. It is effective partnered with other software tools	1 [] 2 [] 3 [] 4 [] 5 [X]				
2. It can help raise the productivity rate when used in conjunction with other software tools	1 [] 2 [] 3 [] 4 [] 5 [X]				
3. It is useful for my tasks	1 [] 2 [] 3 [] 4 [] 5 [X]				
4. It makes it easier to accomplish my tasks	1 [] 2 [] 3 [] 4 [] 5 [X]				
5. It helps save time	1 [] 2 [] 3 [] 4 [] 5 [X]				
6. It does everything I would expect it to do	1 [] 2 [] 3 [X] 4 [] 5 []				
Ease of Use					
1. It is easy to use	1 [] 2 [] 3 [] 4 [X] 5 []				
2. It is simple to use	1 [] 2 [] 3 [X] 4 [] 5 []				
3. It is user friendly	1 [] 2 [] 3 [] 4 [X] 5 []				
4. It requires the fewest steps possible to accomplish what I want to do with it	1 [] 2 [] 3 [X] 4 [] 5 []				
5. It is flexible	1 [] 2 [] 3 [] 4 [X] 5 []				
6. Using it is effortless	1 [] 2 [] 3 [X] 4 [] 5 []				
7. I can use it without written instructions	1 [] 2 [] 3 [X] 4 [] 5 []				
8. I don't notice any inconsistencies as I use it	1 [] 2 [] 3 [] 4 [X] 5 []				
9. Both occasional and regular users would like it	1 [] 2 [] 3 [] 4 [X] 5 []				
10. I can recover from mistakes quickly and easily	1 [] 2 [] 3 [] 4 [] 5 [X]				
11. I can use it successfully every time	1 [] 2 [] 3 [] 4 [X] 5 []				
Ease of Learning					
1. I learned to use it quickly	1 [] 2 [] 3 [] 4 [X] 5 []				
2. I easily remember how to use it	1 [] 2 [] 3 [] 4 [] 5 [X]				
3. Learning how to use it is easy	1 [] 2 [] 3 [] 4 [X] 5 []				

4. I quickly became skillful with using it	<input type="checkbox"/> 1 [] <input type="checkbox"/> 2 [] <input checked="" type="checkbox"/> 3 [X] <input type="checkbox"/> 4 [] <input type="checkbox"/> 5 []
Satisfaction	
1. I am satisfied with it	<input type="checkbox"/> 1 [] <input type="checkbox"/> 2 [] <input type="checkbox"/> 3 [] <input type="checkbox"/> 4 [] <input checked="" type="checkbox"/> 5 [X]
2. I would recommend it to be used in the future	<input type="checkbox"/> 1 [] <input type="checkbox"/> 2 [] <input type="checkbox"/> 3 [] <input type="checkbox"/> 4 [] <input checked="" type="checkbox"/> 5 [X]
3. It works the way I want it to work	<input type="checkbox"/> 1 [] <input type="checkbox"/> 2 [] <input checked="" type="checkbox"/> 3 [X] <input type="checkbox"/> 4 [] <input type="checkbox"/> 5 []
Functions Used (Functions that will most likely be useful to me)	
<ul style="list-style-type: none"> • Cluster • Data Utilities • POS Taggers • Normalize 	
Comments and Suggestions	
<ol style="list-style-type: none"> 1. Most of the functions in modules are properly documented and readable. Atomic functions are very detailed. 2. The functions can be used externally for other data processing tasks. 3. It did not come with any README file or a requirements.txt file to give users a brief overview on how to run the application and its required Python modules. Having these files would ease the use of the API. 4. Example scripts using a few of the functions provided by the API would help users in understanding how it works. 5. Not necessarily a requirement but it would be nice if a Graphical User Interface (GUI) would be integrated in the future for the API. 6. One module, the lang_id, was not recognized by the thesis_software.py program. 	
Negative Aspects	

FILIPINO TEXT ANALYSIS TOOL FOR DISASTERS: REPORT SURVEY					
Value Representation 1- Strongly Disagree, 2- Disagree, 3- Neutral, 4- Agree, 5- Strongly Agree					
1. Aesthetics (design): The report's design is acceptable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/> 4 [X]	<input type="checkbox"/>
2. Content (quality): The different elements of the report (e.g., title, insights list, and Malasakit response list) is necessary	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/> 4 [X]	<input type="checkbox"/>
3. Content (quality): The information fields are appropriate and enough to make a decision or action	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/> 4 [X]	<input type="checkbox"/>
4. Readability (design): The information is clear and readable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/> 4 [X]	<input type="checkbox"/>
5. Understandability (format): The information is easy to interpret	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/> 4 [X]	<input type="checkbox"/>
6. Usefulness of Information (extraction quality): The information is useful in my job	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/> 5 [X]
7. Organization (organization quality): The information is displayed in an organized manner	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/> 5 [X]
8. Usability (efficiency): Using the report in my job would enable me to accomplish tasks more quickly	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/> 4 [X]	<input type="checkbox"/>
9. Usability (potential): The report would enhance my effectiveness on the job	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/> 4 [X]	<input type="checkbox"/>
10. Report (overall quality): I can make decision/s based on the information provided	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/> 5 [X]
11. User Satisfaction (overall measurement): I am satisfied with the report	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/> 4 [X]	<input type="checkbox"/>
Organization Preference					
<p><i>Do you prefer organizing the information by frequency (all entries and ranked by frequency only), per response categories (grouped by categories first then ordered by frequency), or something else?</i></p> <p>I prefer grouping the information first by category then by frequency. However, having a frequency-only ranking for the information also has its merits since it will give the users an idea on which category is most important followed by the rest. I prefer the layout and formatting of A and A – Category compared to the others.</p>					
Comments and Suggestions					
<ol style="list-style-type: none"> 1. The reports are aesthetically formal and proper to look at. 2. It might take some time for the users to fully obtain actionable results from the report since usually it spans more than 20+ pages. 					

- | |
|---|
| 3. Some simple automatically generated graphs such as word clouds or any frequency-based visualization might help the user in understanding the information extracted in partner with the generated report. |
| Negative Aspects |
| |

FILIPINO TEXT ANALYSIS TOOL FOR DISASTERS: API SURVEY					
Value Representation					
1- Strongly Disagree, 2- Disagree, 3- Neutral, 4- Agree, 5- Strongly Agree					
Usefulness					
1. It is effective partnered with other software tools	1 [] 2 [] 3 [] 4 [X] 5 []				
2. It can help raise the productivity rate when used in conjunction with other software tools	1 [] 2 [] 3 [] 4 [X] 5 []				
3. It is useful for my tasks	1 [] 2 [] 3 [X] 4 [] 5 []				
4. It makes it easier to accomplish my tasks	1 [] 2 [] 3 [X] 4 [] 5 []				
5. It helps save time	1 [] 2 [] 3 [] 4 [X] 5 []				
6. It does everything I would expect it to do	1 [] 2 [] 3 [X] 4 [] 5 []				
Ease of Use					
1. It is easy to use	1 [] 2 [] 3 [X] 4 [] 5 []				
2. It is simple to use	1 [] 2 [] 3 [X] 4 [] 5 []				
3. It is user friendly	1 [] 2 [X] 3 [] 4 [] 5 []				
4. It requires the fewest steps possible to accomplish what I want to do with it	1 [] 2 [] 3 [X] 4 [] 5 []				
5. It is flexible	1 [] 2 [] 3 [X] 4 [] 5 []				
6. Using it is effortless	1 [] 2 [X] 3 [] 4 [] 5 []				
7. I can use it without written instructions	1 [] 2 [X] 3 [] 4 [] 5 []				
8. I don't notice any inconsistencies as I use it	1 [] 2 [] 3 [X] 4 [] 5 []				
9. Both occasional and regular users would like it	1 [] 2 [] 3 [] 4 [X] 5 []				
10. I can recover from mistakes quickly and easily	1 [] 2 [] 3 [X] 4 [] 5 []				
11. I can use it successfully every time	1 [] 2 [] 3 [X] 4 [] 5 []				
Ease of Learning					
1. I learned to use it quickly	1 [] 2 [X] 3 [] 4 [] 5 []				
2. I easily remember how to use it	1 [] 2 [] 3 [X] 4 [] 5 []				
3. Learning how to use it is easy	1 [] 2 [X] 3 [] 4 [] 5 []				

4. I quickly became skillful with using it	<input type="checkbox"/> 1 [] <input type="checkbox"/> 2 [] <input checked="" type="checkbox"/> 3 [X] <input type="checkbox"/> 4 [] <input type="checkbox"/> 5 []
Satisfaction	
1. I am satisfied with it	<input type="checkbox"/> 1 [] <input type="checkbox"/> 2 [] <input checked="" type="checkbox"/> 3 [X] <input type="checkbox"/> 4 [] <input type="checkbox"/> 5 []
2. I would recommend it to be used in the future	<input type="checkbox"/> 1 [] <input type="checkbox"/> 2 [] <input type="checkbox"/> 3 [] <input checked="" type="checkbox"/> 4 [X] <input type="checkbox"/> 5 []
3. It works the way I want it to work	<input type="checkbox"/> 1 [] <input type="checkbox"/> 2 [] <input type="checkbox"/> 3 [] <input checked="" type="checkbox"/> 4 [X] <input type="checkbox"/> 5 []
Functions Used (Functions that will most likely be useful to me)	
API function groups 1 - 8	
Comments and Suggestions	
<p>API could have been a RESTful API using HTTP requests and use API documentation tools for the developer guide. Another option is to make it a Python package (installable through pip) and have proper web developer guide to document each API endpoint with the expected request parameters and request format.</p> <p>However, this can be very useful especially that this is tested on a domain in the Filipino language. Users can now have a toolkit that can be used for preprocessing and information extraction for their Filipino dataset.</p>	
Negative Aspects	

FILIPINO TEXT ANALYSIS TOOL FOR DISASTERS: REPORT SURVEY					
Value Representation 1- Strongly Disagree, 2- Disagree, 3- Neutral, 4- Agree, 5- Strongly Agree					
1. Aesthetics (design): The report's design is acceptable	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Content (quality): The different elements of the report (e.g., title, insights list, and Malasakit response list) is necessary	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
3. Content (quality): The information fields are appropriate and enough to make a decision or action	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
4. Readability (design): The information is clear and readable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
5. Understandability (format): The information is easy to interpret	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Usefulness of Information (extraction quality): The information is useful in my job	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. Organization (organization quality): The information is displayed in an organized manner	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
8. Usability (efficiency): Using the report in my job would enable me to accomplish tasks more quickly	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
9. Usability (potential): The report would enhance my effectiveness on the job	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. Report (overall quality): I can make decision/s based on the information provided	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
11. User Satisfaction (overall measurement): I am satisfied with the report	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Organization Preference					
<p><i>Do you prefer organizing the information by frequency (all entries and ranked by frequency only), per response categories (grouped by categories first then ordered by frequency), or something else?</i></p> <p>Organize by response categories then frequency.</p> <p>A – all A – category</p>					
Comments and Suggestions					

Clustering must be improved. Although most clusters are easily understandable, the proposed action (verbs) are still too broad and not every verb in each cluster can be linked to the target nouns.

It will also be cool if there is a report that lists the top 10 opinions by the community as to what should be done during the disaster, and the steps needed to turn the clusters into this list. This way, developers can see deeper value that the information clusters generated is usable in their own report / software.

Negative Aspects

FILIPINO TEXT ANALYSIS TOOL FOR DISASTERS: API SURVEY					
Value Representation					
1- Strongly Disagree, 2- Disagree, 3- Neutral, 4- Agree, 5- Strongly Agree					
Usefulness					
1. It is effective partnered with other software tools	1 [] 2 [] 3 [x] 4 [] 5 []				
2. It can help raise the productivity rate when used in conjunction with other software tools	1 [] 2 [] 3 [] 4 [x] 5 []				
3. It is useful for my tasks	1 [] 2 [] 3 [x] 4 [] 5 []				
4. It makes it easier to accomplish my tasks	1 [] 2 [] 3 [x] 4 [] 5 []				
5. It helps save time	1 [] 2 [] 3 [] 4 [] 5 [x]				
6. It does everything I would expect it to do	1 [] 2 [] 3 [] 4 [] 5 [x]				
Ease of Use					
1. It is easy to use	1 [] 2 [] 3 [] 4 [] 5 [x]				
2. It is simple to use	1 [] 2 [] 3 [] 4 [x] 5 []				
3. It is user friendly	1 [] 2 [] 3 [x] 4 [] 5 []				
4. It requires the fewest steps possible to accomplish what I want to do with it	1 [] 2 [] 3 [] 4 [] 5 [x]				
5. It is flexible	1 [] 2 [] 3 [] 4 [x] 5 []				
6. Using it is effortless	1 [] 2 [] 3 [x] 4 [] 5 []				
7. I can use it without written instructions	1 [] 2 [] 3 [x] 4 [] 5 []				
8. I don't notice any inconsistencies as I use it	1 [] 2 [] 3 [x] 4 [] 5 []				
9. Both occasional and regular users would like it	1 [] 2 [] 3 [x] 4 [] 5 []				
10. I can recover from mistakes quickly and easily	1 [] 2 [] 3 [x] 4 [] 5 []				
11. I can use it successfully every time	1 [] 2 [] 3 [] 4 [] 5 [x]				
Ease of Learning					
1. I learned to use it quickly	1 [] 2 [] 3 [] 4 [x] 5 []				
2. I easily remember how to use it	1 [] 2 [] 3 [] 4 [x] 5 []				
3. Learning how to use it is easy	1 [] 2 [] 3 [] 4 [x] 5 []				

4. I quickly became skillful with using it	1 [] 2 [] 3 [] 4 [x] 5 []
Satisfaction	
1. I am satisfied with it	1 [] 2 [] 3 [] 4 [] 5 [x]
2. I would recommend it to be used in the future	1 [] 2 [] 3 [] 4 [] 5 [x]
3. It works the way I want it to work	1 [] 2 [] 3 [] 4 [x] 5 []
Functions Used (Functions that will most likely be useful to me)	
Clustering part of noisy data (responses)	
Comments and Suggestions	
I would like to recommend for a graphical user interface for ease of use.	
Negative Aspects	
None.	

FILIPINO TEXT ANALYSIS TOOL FOR DISASTERS: REPORT SURVEY					
Value Representation 1- Strongly Disagree, 2- Disagree, 3- Neutral, 4- Agree, 5- Strongly Agree					
1. Aesthetics (design): The report's design is acceptable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
2. Content (quality): The different elements of the report (e.g., title, insights list, and Malasakit response list) is necessary	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
3. Content (quality): The information fields are appropriate and enough to make a decision or action	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
4. Readability (design): The information is clear and readable	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
5. Understandability (format): The information is easy to interpret	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Usefulness of Information (extraction quality): The information is useful in my job	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
7. Organization (organization quality): The information is displayed in an organized manner	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. Usability (efficiency): Using the report in my job would enable me to accomplish tasks more quickly	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
9. Usability (potential): The report would enhance my effectiveness on the job	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
10. Report (overall quality): I can make decision/s based on the information provided	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
11. User Satisfaction (overall measurement): I am satisfied with the report	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Organization Preference					
<p><i>Do you prefer organizing the information by frequency (all entries and ranked by frequency only), per response categories (grouped by categories first then ordered by frequency), or something else?</i></p> <p>Group by frequency first then ordered by categories. A is better for me.</p>					
Comments and Suggestions					
<p>The output should be formatted in such a way non-technical background can easily understand.</p>					
Negative Aspects					
<p>I had a hard time understanding the output at first.</p>					

FILIPINO TEXT ANALYSIS TOOL FOR DISASTERS: API SURVEY					
Value Representation					
1- Strongly Disagree, 2- Disagree, 3- Neutral, 4- Agree, 5- Strongly Agree					
Usefulness					
1. It is effective partnered with other software tools	1 [] 2 [] 3 [] 4 [] 5 [x]				
2. It can help raise the productivity rate when used in conjunction with other software tools	1 [] 2 [] 3 [] 4 [] 5 [x]				
3. It is useful for my tasks	1 [] 2 [] 3 [] 4 [] 5 [x]				
4. It makes it easier to accomplish my tasks	1 [] 2 [] 3 [] 4 [] 5 [x]				
5. It helps save time	1 [] 2 [] 3 [] 4 [] 5 [x]				
6. It does everything I would expect it to do	1 [] 2 [] 3 [] 4 [] 5 [x]				
Ease of Use					
1. It is easy to use	1 [] 2 [] 3 [] 4 [] 5 [x]				
2. It is simple to use	1 [] 2 [] 3 [] 4 [] 5 [x]				
3. It is user friendly	1 [] 2 [] 3 [] 4 [x] 5 []				
4. It requires the fewest steps possible to accomplish what I want to do with it	1 [] 2 [] 3 [] 4 [x] 5 []				
5. It is flexible	1 [] 2 [] 3 [] 4 [x] 5 []				
6. Using it is effortless	1 [] 2 [] 3 [] 4 [x] 5 []				
7. I can use it without written instructions	1 [] 2 [] 3 [] 4 [x] 5 []				
8. I don't notice any inconsistencies as I use it	1 [] 2 [] 3 [] 4 [] 5 [x]				
9. Both occasional and regular users would like it	1 [] 2 [] 3 [] 4 [] 5 [x]				
10. I can recover from mistakes quickly and easily	1 [] 2 [] 3 [] 4 [x] 5 []				
11. I can use it successfully every time	1 [] 2 [] 3 [] 4 [] 5 [x]				
Ease of Learning					
1. I learned to use it quickly	1 [] 2 [] 3 [] 4 [x] 5 []				
2. I easily remember how to use it	1 [] 2 [] 3 [] 4 [] 5 [x]				
3. Learning how to use it is easy	1 [] 2 [] 3 [] 4 [x] 5 []				

4. I quickly became skillful with using it	1 [] 2 [] 3 [] 4 [x] 5 []
Satisfaction	
1. I am satisfied with it	1 [] 2 [] 3 [] 4 [] 5 [x]
2. I would recommend it to be used in the future	1 [] 2 [] 3 [] 4 [] 5 [x]
3. It works the way I want it to work	1 [] 2 [] 3 [] 4 [x] 5 []
Functions Used (Functions that will most likely be useful to me)	
<ul style="list-style-type: none"> ➤ extract_insights_words ➤ organize_by_response_categories ➤ translate_filipino_colloquialism ➤ normalize_list ➤ normalize_string ➤ read_excel ➤ write_report ➤ Clustering Module Functions ➤ Ranking Module Functions 	
Comments and Suggestions	
<ul style="list-style-type: none"> ➤ The user is aware of what is happening during the execution of the program because the program was able to provide enough updates regarding its progress. 	
Negative Aspects	
<ul style="list-style-type: none"> ➤ When I run the program, I noticed some errors like <i>No insights (words) extracted at Response #: <response_ID></i> then I checked the actual responses and there are some valuable insights found there that might also affect the generated reports. There were more or less 50 of such errors. 	

FILIPINO TEXT ANALYSIS TOOL FOR DISASTERS: REPORT SURVEY					
Value Representation 1- Strongly Disagree, 2- Disagree, 3- Neutral, 4- Agree, 5- Strongly Agree					
1. Aesthetics (design): The report's design is acceptable	<input type="checkbox"/> 1 []	<input type="checkbox"/> 2 []	<input type="checkbox"/> 3 []	<input type="checkbox"/> 4 []	<input checked="" type="checkbox"/> 5 [x]
2. Content (quality): The different elements of the report (e.g., title, insights list, and Malasakit response list) is necessary	<input type="checkbox"/> 1 []	<input type="checkbox"/> 2 []	<input type="checkbox"/> 3 []	<input type="checkbox"/> 4 []	<input checked="" type="checkbox"/> 5 [x]
3. Content (quality): The information fields are appropriate and enough to make a decision or action	<input type="checkbox"/> 1 []	<input type="checkbox"/> 2 []	<input type="checkbox"/> 3 []	<input checked="" type="checkbox"/> 4 [x]	<input type="checkbox"/> 5 []
4. Readability (design): The information is clear and readable	<input type="checkbox"/> 1 []	<input type="checkbox"/> 2 []	<input type="checkbox"/> 3 []	<input type="checkbox"/> 4 []	<input checked="" type="checkbox"/> 5 [x]
5. Understandability (format): The information is easy to interpret	<input type="checkbox"/> 1 []	<input type="checkbox"/> 2 []	<input type="checkbox"/> 3 []	<input checked="" type="checkbox"/> 4 [x]	<input type="checkbox"/> 5 []
6. Usefulness of Information (extraction quality): The information is useful in my job	<input type="checkbox"/> 1 []	<input type="checkbox"/> 2 []	<input type="checkbox"/> 3 []	<input type="checkbox"/> 4 []	<input checked="" type="checkbox"/> 5 [x]
7. Organization (organization quality): The information is displayed in an organized manner	<input type="checkbox"/> 1 []	<input type="checkbox"/> 2 []	<input type="checkbox"/> 3 []	<input type="checkbox"/> 4 []	<input checked="" type="checkbox"/> 5 [x]
8. Usability (efficiency): Using the report in my job would enable me to accomplish tasks more quickly	<input type="checkbox"/> 1 []	<input type="checkbox"/> 2 []	<input type="checkbox"/> 3 []	<input type="checkbox"/> 4 []	<input checked="" type="checkbox"/> 5 [x]
9. Usability (potential): The report would enhance my effectiveness on the job	<input type="checkbox"/> 1 []	<input type="checkbox"/> 2 []	<input type="checkbox"/> 3 []	<input type="checkbox"/> 4 []	<input checked="" type="checkbox"/> 5 [x]
10. Report (overall quality): I can make decision/s based on the information provided	<input type="checkbox"/> 1 []	<input type="checkbox"/> 2 []	<input type="checkbox"/> 3 []	<input checked="" type="checkbox"/> 4 [x]	<input type="checkbox"/> 5 []
11. User Satisfaction (overall measurement): I am satisfied with the report	<input type="checkbox"/> 1 []	<input type="checkbox"/> 2 []	<input type="checkbox"/> 3 []	<input type="checkbox"/> 4 []	<input checked="" type="checkbox"/> 5 [x]
Organization Preference					
<p><i>Do you prefer organizing the information by frequency (all entries and ranked by frequency only), per response categories (grouped by categories first then ordered by frequency), or something else?</i></p> <ul style="list-style-type: none"> ➤ I prefer to have the responses organized per response categories. I prefer approach A – category, even though the proposed action is limited, insights are better gained using this approach by being specific on what the proposed action is. 					
Comments and Suggestions					
<ul style="list-style-type: none"> ➤ Entries with only 1 frequency I think is not anymore necessary to be included in the reports since most of these entries doesn't make sense and is probably included already in other entries with higher frequency. 					

Negative Aspects
➤ The length of the generated reports were kind of overwhelming but given the amount of information provided, it is understandable.

FILIPINO TEXT ANALYSIS TOOL FOR DISASTERS: API SURVEY					
Value Representation					
1- Strongly Disagree, 2- Disagree, 3- Neutral, 4- Agree, 5- Strongly Agree					
Usefulness					
1. It is effective partnered with other software tools	1 [] 2 [] 3 [] 4 [] 5 [x]				
2. It can help raise the productivity rate when used in conjunction with other software tools	1 [] 2 [] 3 [] 4 [] 5 [x]				
3. It is useful for my tasks	1 [] 2 [] 3 [] 4 [x] 5 []				
4. It makes it easier to accomplish my tasks	1 [] 2 [] 3 [] 4 [] 5 [x]				
5. It helps save time	1 [] 2 [] 3 [] 4 [x] 5 []				
6. It does everything I would expect it to do	1 [] 2 [] 3 [] 4 [x] 5 []				
Ease of Use					
1. It is easy to use	1 [] 2 [] 3 [] 4 [x] 5 []				
2. It is simple to use	1 [] 2 [] 3 [] 4 [] 5 [x]				
3. It is user friendly	1 [] 2 [] 3 [] 4 [x] 5 []				
4. It requires the fewest steps possible to accomplish what I want to do with it	1 [] 2 [] 3 [] 4 [x] 5 []				
5. It is flexible	1 [] 2 [] 3 [] 4 [x] 5 []				
6. Using it is effortless	1 [] 2 [] 3 [] 4 [x] 5 []				
7. I can use it without written instructions	1 [] 2 [] 3 [] 4 [x] 5 []				
8. I don't notice any inconsistencies as I use it	1 [] 2 [] 3 [] 4 [x] 5 []				
9. Both occasional and regular users would like it	1 [] 2 [] 3 [] 4 [x] 5 []				
10. I can recover from mistakes quickly and easily	1 [] 2 [] 3 [] 4 [x] 5 []				
11. I can use it successfully every time	1 [] 2 [] 3 [] 4 [x] 5 []				
Ease of Learning					
1. I learned to use it quickly	1 [] 2 [] 3 [] 4 [x] 5 []				
2. I easily remember how to use it	1 [] 2 [] 3 [] 4 [] 5 [x]				
3. Learning how to use it is easy	1 [] 2 [] 3 [] 4 [x] 5 []				

4. I quickly became skillful with using it	<input type="checkbox"/> 1 [] <input type="checkbox"/> 2 [] <input type="checkbox"/> 3 [] <input checked="" type="checkbox"/> 4 [x] <input type="checkbox"/> 5 []
Satisfaction	
1. I am satisfied with it	<input type="checkbox"/> 1 [] <input type="checkbox"/> 2 [] <input type="checkbox"/> 3 [] <input type="checkbox"/> 4 [] <input checked="" type="checkbox"/> 5 [x]
2. I would recommend it to be used in the future	<input type="checkbox"/> 1 [] <input type="checkbox"/> 2 [] <input type="checkbox"/> 3 [] <input type="checkbox"/> 4 [] <input checked="" type="checkbox"/> 5 [x]
3. It works the way I want it to work	<input type="checkbox"/> 1 [] <input type="checkbox"/> 2 [] <input type="checkbox"/> 3 [] <input checked="" type="checkbox"/> 4 [x] <input type="checkbox"/> 5 []
Functions Used (Functions that will most likely be useful to me)	
read_excel, refresh_excel, organize_sublist, identify_language_string_list, format_pos, rank_by_response_categories, extract_insights_words	
Comments and Suggestions	
Instructions must be provided to enable the user to use it to do its job/purpose. Also, functions need some example inputs or arguments to feed into the function.	
Negative Aspects	
No instructions provided if there are requirements needed to install and on how to use it.	

FILIPINO TEXT ANALYSIS TOOL FOR DISASTERS: REPORT SURVEY					
Value Representation 1- Strongly Disagree, 2- Disagree, 3- Neutral, 4- Agree, 5- Strongly Agree					
1. Aesthetics (design): The report's design is acceptable	<input type="checkbox"/> 1 []	<input type="checkbox"/> 2 []	<input type="checkbox"/> 3 []	<input checked="" type="checkbox"/> 4 [x]	<input type="checkbox"/> 5 []
2. Content (quality): The different elements of the report (e.g., title, insights list, and Malasakit response list) is necessary	<input type="checkbox"/> 1 []	<input type="checkbox"/> 2 []	<input type="checkbox"/> 3 []	<input type="checkbox"/> 4 []	<input checked="" type="checkbox"/> 5 [x]
3. Content (quality): The information fields are appropriate and enough to make a decision or action	<input type="checkbox"/> 1 []	<input type="checkbox"/> 2 []	<input type="checkbox"/> 3 []	<input checked="" type="checkbox"/> 4 [x]	<input type="checkbox"/> 5 []
4. Readability (design): The information is clear and readable	<input type="checkbox"/> 1 []	<input type="checkbox"/> 2 []	<input type="checkbox"/> 3 []	<input checked="" type="checkbox"/> 4 [x]	<input type="checkbox"/> 5 []
5. Understandability (format): The information is easy to interpret	<input type="checkbox"/> 1 []	<input type="checkbox"/> 2 []	<input type="checkbox"/> 3 []	<input checked="" type="checkbox"/> 4 [x]	<input type="checkbox"/> 5 []
6. Usefulness of Information (extraction quality): The information is useful in my job	<input type="checkbox"/> 1 []	<input type="checkbox"/> 2 []	<input type="checkbox"/> 3 []	<input checked="" type="checkbox"/> 4 [x]	<input type="checkbox"/> 5 []
7. Organization (organization quality): The information is displayed in an organized manner	<input type="checkbox"/> 1 []	<input type="checkbox"/> 2 []	<input type="checkbox"/> 3 []	<input type="checkbox"/> 4 []	<input checked="" type="checkbox"/> 5 [x]
8. Usability (efficiency): Using the report in my job would enable me to accomplish tasks more quickly	<input type="checkbox"/> 1 []	<input type="checkbox"/> 2 []	<input type="checkbox"/> 3 []	<input checked="" type="checkbox"/> 4 [x]	<input type="checkbox"/> 5 []
9. Usability (potential): The report would enhance my effectiveness on the job	<input type="checkbox"/> 1 []	<input type="checkbox"/> 2 []	<input type="checkbox"/> 3 []	<input type="checkbox"/> 4 []	<input checked="" type="checkbox"/> 5 [x]
10. Report (overall quality): I can make decision/s based on the information provided	<input type="checkbox"/> 1 []	<input type="checkbox"/> 2 []	<input type="checkbox"/> 3 []	<input checked="" type="checkbox"/> 4 [x]	<input type="checkbox"/> 5 []
11. User Satisfaction (overall measurement): I am satisfied with the report	<input type="checkbox"/> 1 []	<input type="checkbox"/> 2 []	<input type="checkbox"/> 3 []	<input type="checkbox"/> 4 []	<input checked="" type="checkbox"/> 5 [x]
Organization Preference					
<p><i>Do you prefer organizing the information by frequency (all entries and ranked by frequency only), per response categories (grouped by categories first then ordered by frequency), or something else?</i></p> <p>I prefer organizing per response categories with C then ordered by frequency to present a clearer insight and easier interpretation on the information shown.</p>					
Comments and Suggestions					
<p>Words like "mag" in the proposed action must contain other information for narrowing down some actions.</p>					

Negative Aspects
The reports generated are too much to interpret and a generalization of the words is needed to provide a broader and faster reports. Either it is full of information that make its purpose done or a generalization of ideas are needed based from the informations from each categories

Appendix H

Resource Persons

Dr. Charibeth Cheng

Adviser & Associate Dean
College of Computer Studies
De La Salle University-Manila
charibeth.cheng@dlsu.edu.ph

Dr. Rachel Roxas

Project Leader
Philippine-California Advanced Research Institutes
National University, Manila
reoroxas@national-u.edu.ph

Dr. Ethel Ong

Graduate Studies Coordinator
College of Computer Studies
De La Salle University-Manila
ethel.ong@dlsu.edu.ph

Dr. Conrado Ruiz Jr.

Assistant Professor
College of Computer Studies
De La Salle University-Manila
cons.ruizjr@delasalle.ph

References

- Ali, F., Kwak, D., Khan, P., Ei-Sappagh, S. H. A., Islam, S. M. R., Park, D., & Kwak, K. (2017). Merged ontology and svm-based information extraction and recommendation system for social robots. *IEEE Access*.
- Appelt, D. E. (1999). Introduction to information extraction. *Ai Communications*, 12(3), 161-172.
- Beduya, L. J., & Espinosa, K. J. (2014). Disaster-related participant tweet identification using svm. *Philippine Computing Journal Dedicated Issue on Natural Language Processing*, 24-33.
- BIPM, I., IFCC, I., IUPAC, I., & ISO, O. (2012). The international vocabulary of metrology basic and general concepts and associated terms (VIM), 3rd edn. JCGM 200: 2012. *JCGM (Joint Committee for Guides in Metrology)*.
- Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5, 135–146.
- Borra, A., Santos, K. D., Gonzales, D., & Reyes, A. (2013). Maternal medical information extraction (mamie) system. In *Proceedings of the 9th national natural language processing research symposium (NNLPRS)*.
- Cheng, C., Cagampan, B., & Lim, C. D. (2016). Organizing news articles and editorials through information extraction and sentiment analysis. In *Proceedings of the pacific asia conference on information systems (PACIS)* (p. 258).
- Cheng, C. K., & Rabo, V. S. (2004). Tpost: A template-based, n-gram part-of-speech tagger for tagalog. *Journal Research in Science, Computing and Engineering (JRSCE)*, 3(1).
- Codebook version 4.7 for the malasakit qualitative responses*. (2017). Philippine California Advanced Research Institutes. (IIID 2015-007: E-Participation 2.0: Connecting Diverse Philippine Populations for Disaster Risk Management with a Toolkit Integrating Text and Speech Analytics)
- Culotta, A., Bekkerman, R., & McCallum, A. (2004). Extracting social networks and contact information from email and the web. *Computer Science Department Faculty Publication Series*, 33.
- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of

- computer technology: A comparison of two theoretical models..
- De La Cruz, A., Oco, N., & Roxas, R. E. (2017). A classifier module for analyzing community responses on disaster preparedness. In *Proceedings of the 2017 IEEE 9th international conference on humanoid, nanotechnology, information technology, communication and control, environment and management (HNICEM)* (p. 1-5). IEEE.
- Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3), 297–302.
- Dietrich, D., Heller, B., & Yang, B. (2015). Data science & big data analytics: discovering, analyzing, visualizing and presenting data. *EMC Education Services*.
- Driver, H. E., & Kroeber, A. L. (1932). *Quantitative expression of cultural relationships* (Vol. 31) (No. 4). University of California Press.
- Gan, G., Ma, C., & Wu, J. (2007). *Data clustering: Theory, algorithms, and applications*. Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104). Retrieved from <https://books.google.com.ph/books?id=HMfJHBW8x\EC>
- Go, M. P., & Nocon, N. (2017). Using stanford part-of-speech tagger for the morphologically-rich filipino language. In *Proceedings of the 31st pacific asia conference on language, information and computation* (pp. 81–88).
- Go, M. P., Nocon, N., & Borra, A. (2017). Gramatika: A grammar checker for the low-resourced filipino language. In *Tencon 2017 ieee region 10 conference* (pp. 471–475).
- Gorro, K., Ancheta, J. R., Capao, K., Oco, N., Roxas, R. E., Sabellano, M. J., ... Goldberg, K. (2017). Qualitative data analysis of disaster risk reduction suggestions assisted by topic modeling and word2vec. In *Proceedings of the 2017 international conference on asian language processing (IALP)* (p. 293-297). IEEE.
- Grishman, R. (1997). Information extraction: Techniques and challenges. *International Summer School on Information Extraction*.
- Guha-Sapir, D., Hoyois, P., Wallemacq, P., & Below, R. (2017). Annual disaster statistical review 2016: The numbers and trends. *Brussels: Centre for Research on the Epidemiology of Disasters (CRED)*.
- Hripcsak, G., & Rothschild, A. S. (2012). Agreement, the f-measure, and reliability in information retrieval. *Journal of the American Medical Informatics Association*, 12(3), 296-298.
- Jurafsky, D., & Martin, J. H. (2018). *Speech and language processing*. Stanford.
- Lagmay, A. M. F. A., Racoma, B. A., Aracan, K. A., Alconis-Ayco, J., & Saddi, I. L. (2017). Disseminating near-real-time hazards information and flood maps in the philippines through Web-GIS. *Journal of Environmental Sciences*, 59, 13-23.
- Lam, A., Paner, I., Macatangay, J. M., & Delos Santos, D. D. (2014). Classifying

- typhoon related tweets. In *Proceedings of the 10th national natural language processing research symposium (NNLPRS)*.
- Livelo, E. D. S., Ver, A. N. O., Chua, J. L., Yao, J. P. S., & Cheng, C. K. (2017). A hybrid agent for automatically determining and extracting the 5ws of filipino news articles. In *Proceedings of ijcai workshop on semantic machine learning (SML 2017)*.
- Loper, E., & Bird, S. (2002). Nltk: the natural language toolkit. *arXiv preprint cs/0205028*.
- Lui, M., & Baldwin, T. (2012). langid. py: An off-the-shelf language identification tool. In *Proceedings of the acl 2012 system demonstrations* (pp. 25–30).
- Lund, A. M. (2001). Measuring usability with the use questionnaire. *Usability interface*, 8(2), 3–6.
- Marcus, M. P., Marcinkiewicz, M. A., & Santorini, B. (1993). Building a large annotated corpus of english: The penn treebank. *Computational linguistics*, 19(2), 313-330.
- Miguel, D., & Roxas, R. E. (2007). Comparative evaluation of tagalog part-of-speech taggers. In *Proceedings of the 4th national natural language processing research symposium NNLPRS*.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Mikolov, T., Grave, E., Bojanowski, P., Puhrsch, C., & Joulin, A. (2017). Advances in pre-training distributed word representations. *arXiv preprint arXiv:1712.09405*.
- Mikolov, T., Yih, W.-t., & Zweig, G. (2013). Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: Human language technologies* (pp. 746–751).
- Nocon, N., & Borra, A. (2016). Smtpost using statistical machine translation approach in filipino part-of-speech tagging. In *Proceedings of the 30th pacific asia conference on language, information and computation (PACLIC)* (p. 391-396).
- Nocon, N., Cuevas, G., Magat, D., Suministrado, P., & Cheng, C. (2014). Normapi: An api for normalizing filipino shortcut texts. In *2014 international conference on asian language processing (ialp)* (pp. 207–210).
- Nocon, N., Kho, N. M., & Arroyo, J. (2018). Building a filipino colloquialism translator using sequence-to-sequence model. In *Tencon 2018-2018 ieee region 10 conference* (pp. 2199–2204).
- Nonnecke, B., Mohanty, S., Lee, A., Lee, J., Beckman, S., Mi, J., ... others (2018). Malasakit 2.0: A participatory online platform with feature phone integration and voice recognition for crowdsourcing disaster risk reduction strategies in the philippines. In *2018 ieee global humanitarian technology conference (ghtc)* (pp. 1–6).

- Nonnecke, B. M., Mohanty, S., Lee, A., Lee, J., Beckman, S., Mi, J., . . . Goldberg, K. (2017). Malasakit 1.0: A participatory online platform for crowdsourcing disaster risk reduction strategies in the philippines. In *Proceedings of the 2017 IEEE global humanitarian technology conference (GHTC)* (p. 1-6). IEEE.
- Oco, N., & Borra, A. (2011). A grammar checker for tagalog using languagetool. In *Proceedings of the 9th workshop on asian language resources* (pp. 2–9).
- Oco, N., Syliongka, L. R., Allman, T., & Roxas, R. E. (2016, October). Philippine language resources: Applications, issues, and directions. In *Proceedings of the 30th pacific asia conference on language, information and computation: Posters* (pp. 433–438). Seoul, South Korea. Retrieved from <https://www.aclweb.org/anthology/Y16-3015>
- Peña, W., & Melgar, A. (2015). Ontology-based information extraction from spanish forum. *Computational Collective Intelligence*.
- Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (emnlp)* (pp. 1532–1543).
- Regalado, R. V., Kalaw, K. M. D., Lu, V., Dela Cruz, K. M. H., & Garcia, J. P. (2015). Filiet: An information extraction system for filipino disaster-related tweets. In *Proceedings of the DLSU research congress vol. 3*.
- Řehůřek, R., & Sojka, P. (2010, May 22). Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks* (pp. 45–50). Valletta, Malta: ELRA. (<http://is.muni.cz/publication/884893/en>)
- Saggion, H., Funk, A., Maynard, D., & Bontcheva, K. (2007). Ontology-based information extraction for business intelligence. *The Semantic Web*, 843–856.
- Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Fam Med*, 37(5), 360-363.