Extracting and Organizing Disaster-Related Philippine Community Responses for Aiding Nationwide Risk Reduction Planning and Response

Nicco Louis S. Nocon and Charibeth K. Cheng, PhD
De La Salle University
nicco_louis_nocon@dlsu.edu.ph, charibeth.cheng@dlsu.edu.ph

ABSTRACT

Philippines is one of the most disaster-prone countries in the world. Typhoons and floods are normally experienced in the country, for which in every experience, spike of data in all mediums are evident. Given this, it granted researchers the opportunity to analyze and study these data. Attempting to amend the current situation of the country in terms of handling disasters, disaster risk reduction strategies were directly taken from local communities by an online participatory platform, *Malasakit*, in light of the specific needs of the people. Gathering their insights provided numerous ideas on how to prevent and mitigate disasters experienced by the country, however these ideas come in an unstructured form. This research aims to convert unstructured data to structured form by extracting those disaster-related community responses in Filipino, organize the extracted information, and generate a report for decision makers to address. Provided, the overall task collated insights, listing items to "what do people want or need in their community". It was implemented through Part-of-Speech-based Information Extraction (IE) and Clustering techniques. As support, Part-of-Speech Tagging, Word Embeddings, Ranking and Report Generation tasks were included.

Keywords

Information Extraction, Clustering, Word Embeddings, Disaster Data.

1. Overview

Disasters are one of the primary reasons for disrupting the world's social and economic status. From 2006 to 2015, the average occurrence of natural disasters is at 376.4, resulting in 69,827 deaths at the average and about $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 [6].

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* [17], 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 rather 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.

A more direct approach was conducted by *Malasakit* [14], an online participatory tool. It collected responses from local communities with ideas on how to make the country better handle disasters. Existing works analyzed these responses through classification [3] and modeling [5] 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 responses that can directly help disaster risk reduction. Doing so would present many 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.

Since *Malasakit* responses are unstructured, Information Extraction (IE), a method focused on automatically extracting and providing structure to a given unstructured text [7] was implemented. The structured information can be used to disseminate information to an intended target, may it be people or technology (a different program). In relation to *FILIET*, this study adopted the idea of using Part-of-Speech for IE. It is then supplemented by introducing a grouping and ranking technique that organizes the extracted ideas. Furthermore, dissemination in a form of a list report was made as medium to relay and communicate ideas with corresponding decision makers.

Generally, the objective of this research is to extract and organize disaster-related community responses and generate a report for decision makers to address. Specifically, this research aims to:

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

Following this, a discussion about the research's methodology is included, followed by showcasing the experiments, results and analysis, and lastly providing a recap of the research and listing its future directions.

2. Research Methodology

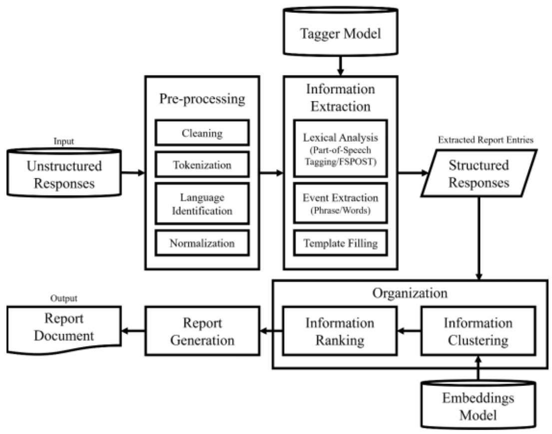

Figure 1. Architectural Diagram of the Extraction Tool.

The following are tasks involved in developing the extraction tool (see Figure 1): Data Acquisition and Preprocessing, Information Extraction and Organization, and Report Generation.

2.1 Data Acquisition

Data gathered came from *Malasakit's* Local Community Responses [14]. It captures insights to an open-ended question, "How could your Barangay help you better prepare for a disaster". It contains 934 qualitative responses, gathered in Filipino and English, with each comprised of *response* and *tag* features. The *response* feature contains ideas and the main target for extracting and organizing information, while *tag* (or response category) represents a general view of the response given used for ranking.

Along this data, a gold standard was produced by an annotator involved in the *Malasakit* project for evaluating the extraction process. There are two kinds of gold standard: one that forms with insight phrases and with word sets containing word tuples that groups the proposed action and its target/s.

2.2 Data Preprocessing

An attempt to correct misspelled (e.g., bagy0 and maayus) and shortened (e.g., brgy, LGU, and kpag) words in *Malasakit* responses, preprocessing tasks such as tokenization, cleaning, and normalization (word standardization) were applied.

Under normalization, an extendable prefix list [15] to find and join all unmerged prefixes in the text, and a *Filipino Normalizer* [13] were utilized to automatically process or overwrite texts. It covers *textspeak*, *swardspeak*, *conyo* (Tagalog-English mix), and *datkilab* (metathesis or reversed words) styles, built from social media sources and Statmt's Moses engine.

In addition to these, a language identifier was implemented, with the intention of directing appropriate natural language to the Part-of-Speech tagger. An off-the-shelf tool called *LangID* [9] was used, built through Numpy's Naïve Bayes classifier and trained on a multi-domain corpus, comprised of various documents (i.e., news, encyclopedia, and internet crawled texts). Applying it, the model was filtered to focus on English (en) and Tagalog (tl).

2.3 Information Extraction

Responses were automatically extracted with insights through Part-of-speech-based approach. In detail, responses were labeled by Filipino Stanford Part-of-Speech Tagger (FSPOST) [4]. 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 in Filipino morphology to determine the appropriate tag for a word.

A Python implementation was done through Natural Language Toolkit (NLTK) [8], which facilitates switching of languages in tagging once the identifier labels a response. Switching involves using NLTK Tagger for English or FSPOST-NLTK for Tagalog.

Equipped with FSPOST, "events" wherein described as verb phrases or verb-noun words that contain actionable suggestions given by the community were extracted. It is done through a pattern of finding an action/verb and traversing through words until a target/noun is found (extended if there is a comma or conjunction *at* 'and' that is followed by another noun). After extraction, template filling provides structure to texts. It works by plugging a set of information into corresponding fields.

2.4 Information Organization

This task is comprised of Clustering and Ranking. There are two ways that it can be set up: one is by organizing all entries, clustering similar entries and ranking solely on frequency; Another is by response categories, clustering under each category, ranking them by category priority and locally based on frequency.

2.4.1 Information Clustering

Entries with similar contents were collated through string and semantic (word embeddings) clustering techniques. For string similarity, an online available Python library called *Strsim*[1] was used, in which the collection contains Sørensen-Dice Coefficient. For word embeddings, pre-trained Wikipedia Tagalog models[2] were used as resources and *Gensim* library [18] for implementing *Word2Vec* [11] and *FastText* [1].

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

Applying a technique results into a computed number representing their similarity. This value is checked with a similarity threshold. A value lower than or equal to the threshold means the pair are not similar to each other, thus there would be no clustering involved. On the other hand, a higher value means the pair are similar and should be joined into a single cluster, appending their information such as IDs, frequencies, actions, and targets. In cases that a pair is exactly similar, only one instance of the word will remain, but their information will still be appended.

2.4.2 Information Ranking

Since organization accommodates two options, ranking can use frequency counts in decreasing order, the given response categories as basis for prioritization, or a combination of both. Moreover, prioritization for categories was arranged based on its characteristic of being actionable by decision makers specific to handling disasters. Having said that, arranging the categories to prioritize 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.

---

[1] https://github.com/luozhouyang/python-string-similarity

[2] https://github.com/Kyubyong/wordvectors

2.5 Report Generation

The ranked information will be transformed into a report, generated in a list format through a template-based approach. The report that is included in this tool can be written in a Microsoft Word (designed for reading and marking items) or Excel (for analysis) file. Report generation's main task is to add details and design the entries on those formats, with consideration to setups toggled through information organization. It produces the final output of this tool, a generated report document which can be relayed to or used by decision makers in handling disasters.

3. Experiments, Results and Analysis

In this part of the paper, experiments were discussed; specifically, changes to utilized data and software modules. Evaluation was done through quantitative and qualitative analysis, for Information extraction (IE) and organizational tasks, respectively.

3.1.1 Quantitative Results

In evaluating the tool, IE's performance was measured through standard, quantitative metrics such as Precision, Recall, Accuracy and F-Measure (see Table 1). Moreover, there were two tests, comparing insight phrases to know the performance in extracting insights and comparing word sets to know the performance in extracting action/verbs and target/nouns.

3.1.1.1 Insight Phrases

Complete Matches (CM) 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 in the results are instances of GS entries without Auxiliary verbs (VBS) at the start of the insight such as *dapat*, *kailangan*, and *be*. Another, OE includes a word or few words after conjunctions 'at' or 'and' which made it longer. Unlike in GS, some ideas are separated into two entries. An example for this 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 causes of having OE, forcing to search for a noun to end an insight added unnecessary words that could have stopped on an adjective or pronoun.

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 for UE is different. Instead of having more words, UE shows necessary details missing in the insight. The phrase "make the barangay gym" is a sample of 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 additional detail such as "ayusin ang drainage" and "do announcements", which could have been "ayusin ang drainage ng barangay" and "do announcements via megaphones".

OVE are partial matches that are equal in length or wherein certain words overlaps with 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 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*).

Table 1. Results of the Information Extraction Task. [3]

	Insight Phrases %			Word Sets %	
	Orig.	**Norm.**		**Orig.**	**Norm.**
CM	**19.59**	18.23	**EM**	**18.75**	18.22
OE	8.35	**8.99**	**PM**	25.6	**26.43**
UE	**25.48**	24.98	**AM**	8.22	**9.32**
OVE	0.3	**0.66**	**TM**	**17.47**	16.08
CMM	**46.27**	47.13	**COM**	**4.02**	2.48
			NMGS	**25.94**	27.46
			NMT	**37.86**	37.97
TP	**53.73 (41.35)**	52.87 (40.24)	**TP**	**51**	50.2
FP	29.69 (17.65)	**29.39 (17.46)**	**FP**	31.09	**30.75**
FN	**16.59 (9.78)**	17.74 (10.6)	**FN**	**17.88**	19.02
TN	0 (31.22)	**0 (31.70)**	**TN**	0	0
P	**64.41 (70.09)**	64.27 (69.74)	**P**	**62.14**	62.03
R	**76.41 (80.87)**	74.87 (79.15)	**R**	**74.06**	72.54
A	**53.73 (72.57)**	52.87 (71.94)	**A**	**51.03**	50.24
F	**69.90 (75.09)**	69.17 (74.14)	**F**	**67.58**	66.88

Complete Mismatch (CMM) are entries that does not match with either the tool's extractions or GS. Under this measure belong the sum of False Positives (FP) which are extractions that are not actual insights nor in GS, and False Negatives (FN) which are not extracted that are actual insights or in 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 tool; unlike in GS, the entries were separated. In this tool's 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.

[3] For the metrics, values in parenthesis are based on word counts, while those outside of it are based on insight counts.

True Negatives (TN) are instances wherein the tool did not extract, as it is not an actual insight nor in GS. Regarding insight counts, the value for this measure is zero, as those that were not extracted and not existent in GS were not included in this type of evaluation. However, for word counts, the value can be computed by counting the total number of words in the data and subtracting word counts in TP, FP, and FN.

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

A normalized version of the insight phrases was also evaluated. Generally, the original version gained better scores compared to the normalized one. The cause of this is mostly on spelling mismatches, specifically over-normalization (e.g., *kits* to *kita*, *kapwa* to *kawpa*, and "ayusin ang mga" to "ayusin ang Metro Rail mga"), annotation normalizations (e.g., *kagamitin* – *kagamitan*, *pundo* – *pondo*, and publicsoundnotifsystem – public sound notif system), and unnormalized typographical errors (e.g., *plansan*, *taraining*, and *alertoat*). Despite this, numerous words in the data was still standardized; examples are *di* to *hindi*, *pls* to *please*, *san* to *saan*, *meron* to *mayroon*, *anu* to *ano*, and more.

3.1.1.2 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 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].

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 GS' missed corrections (e.g., *paguusap* – *pag uusap*). 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 types of 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 targets/nouns.

No matches for Tool (NMT) 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. NMT 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 could not be extracted due to the verb to noun pattern rule.

Based on the results, less than 20% are exact matches, while PM has the highest number of correct instances with more than 25%. Represented with a combined value of 74.06%, extracting action and target fields were effective provided with a straightforward Part-of-Speech pattern-based design. Perfectly extracting both fields, however, still needs work.

Overall, word sets' P, R, A, and F metric values exceeded half, garnering 62.14, 74.06, 51.03, and 67.58, respectively. Comparing with its normalized counterpart, PM, AM, and FP values improved, while everything else were lower than the original.

3.1.2 Qualitative Analysis Setup and Results

Qualitative analysis was 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, Word2Vec, and FastText, with applied lexicalization on insights.

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

In this type of analysis, the following was discovered: top community suggestions, set of words similar to each other, highly composed response category, effectivity of a normalizer, more appropriate clustering approach or threshold, and many more.

3.1.2.1 Community Suggestions

The entirety of the extraction process produced 1,392 insights. Generally, suggestions produced clusters with verbs or topics on *logistics* (e.g., magkaroon, magbigay, ayusin, tanggalin, pagbibigay, and provide), *dissemination* (e.g., pagiinform, ma-inform, and inform), *sanitation* (e.g., maglinis, linisin, and linisan), *preparedness* (e.g., prepare, ihanda, maiwasan, and mkaiwas), and *solidarity* (e.g., gumawa, magsagawa, tumulong, makipagtulongan, and help).

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

Mainly to be able to avert or avoid ramifications such as floods or clogging, spread of tragedies, and getting trapped, community- and self-preparedness through alarms and designating places such as schools for evacuations were suggested. For solidarity, it has been recommended for families and Barangay to continuously help or support each other.

Classifying ideas to a more generalized view, the response categories ordered by frequency (count attached) are as follows: Information Campaign and Capacity Building (393), Early Warning System (259), Preparedness for Emergency (198), Infrastructure Maintenance and Management (169), Filipino Values (114), Community-wide Logistic Support for Disaster Response (80), Local Government Accountability (78), Others (67), and the least suggested Disaster Relief (34).

Formatted in the way that insights are grouped by response categories divided and spread out all entries for which categories the responses are under and produced ideas in a more focused manner. With the assertion of being under a category, entries pertain to needs or actions to be done specific to them.

Information Campaign and Capacity Building focused on actions that the community must have, which includes necessary items or programs, preparations, support, and logistics involved in information dissemination. It also enumerated the intended contents, medium of information, receiving end of the information, recommendations to decision makers. 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 contained ideas that mentioned receiving or providing assistance, goods, food or grocery, medicine, and evacuation option. As a matter of fact, the entries with high frequencies pertain to the same ideas.

Similarly, Community-wide Logistic Support for Disaster Response suggested having safety gears, sirens, and shelter for the operations. It has also been suggested to add more budget, equipment, and volunteers. Expounding on equipment, several responses pairs medical kits and flashlight together. Moreover, boats and storage facility were pointed out to be must-haves. Other ideas mentioned building or buying an area or place, adding budget and community volunteers, and communicate weather predictions.

For Infrastructure Maintenance and Management, most of the suggestions were about sanitation and repairs, specifically on the surroundings such as streets, sewers, rivers, and garbage waste. Cleaning these areas in the minds of the community would ensure prevention of floods and clogs. Furthermore, in avoidance to throwing trash everywhere, one of the top suggestions mentioned proper place or containers for garbage.

As stated for Early Warning System, it has responses with insights that involves information dissemination. There should be proper communication, alert, news and updates, and radio. Entries about information, indicated them to be about the disaster, specifically the typhoon, and announced to the public (people, citizens, and residents). Another with the same idea, indicated alerts regarding calamities, catastrophe, disaster, typhoon, communication, support, and assembly.

Under Preparedness for Emergency, ideas were in a form of a reminder, with verbs grouped such as *maging* 'become', *pagiging*, *magiging*, and *palaging* 'always', or *maayos* 'fix', *malaman* 'know', *magpadala* 'send', *matugunan* 'address', and *maiwasan* 'avoid'. Targets for this set is to alert and be attentive with nature, news, and officials. Another suggests being 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 weather forecasts 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 target ideas that the government should be accountable for: disasters, evacuation, posters, case, and society. It is also expected for them to be ready, cooperative, and able to show and send help. In light of this, there was a suggestion that they should be active and focus on helping people.

Regarding Filipino Values, ideas encouraged solidarity, that is to keep everyone united and participative in helping each other. Help should be observable in preparing families and community; while being cooperative with duties, plans, and preparation. Other than these, separate entries have verbs that mentioned tidiness, readiness, equality, kindheartedness, concern, and order.

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

3.1.2.2 Lexicalization

Lexicalization was applied in all experiments, where one word is designated to represent a cluster (particularly in each of the Target field). Implementing this was successful in labeling joined words related to nouns. Orthographic instances represented tenses and unstandardized variants such as seminars (seminar, seminarsdrill), floods (flooding), *basura* (basurahan), and kapaligiran (paligid). Whereas semantic instances represented a set of words under an idea such as training (community, programs, assembly, government, technology, days, and more). In spite this, there are instances with close string distance but are not exactly related. Examples for this are evacuation-elevation and pagkain-pagkalap.

Primary learning for this is its effectivity highly depends on the performance of clustering approaches. One thing to note of is its potential of displaying a single word that can represent an entire set of words given a good clustering approach. Selecting one word however is another problem to tackle.

3.1.2.3 Normalized Version

Table 2. Normalization Samples

Original-Normalized		
andyan – nandiyan	kase – kasi	s – sa
anung – anong	konting – kaunting	san – saan
aq – ako	kpag – kapag	tas – tapos
cla – sila	meron – mayroon	tsaka – at saka
di – hindi	pano – paano	tv – television
dont – do not	pls – please	xa – siya
facebook – Facebook	pwd– puwede	yung – iyong

Shown on Table 2 are few samples of the corrections using [13] that covered shortcut texts, 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. An instance of this inserted an extra ‘it’ in between ‘as’ and ‘possible’ in the phrase “as early as possible”. Another is “ayusin ang” which has “metro rail” succeeding it.

There were also issues between colloquialisms and interlingual homographs[4]. In the English phrase “seminar for pre or post…”, the prefix *pre-* was found to be a shortcut for the colloquialism *pare* ‘buddy’, instead of leaving it as is. Other mistakes are *kits* in “medical kits” were considered as a typographical error for *kita* ‘you’, *to* into *ito* ‘this’, *non* into *noon* ‘previously’, and *my* into *may* ‘there’. 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*, 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’, ‘dhlan’. ‘magki2ta’, ‘mlman’, ‘gwen’, and ‘magsgwa’.

Observing its effect, frequency counts fluctuated, shifting the order of clusters, but did not affect the top ones. It has been evident that there are 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. One instance, *mag-* prefix was joined with *karoon* which resulted into *magkaroon*, a valid member of a proposed actions. More examples like this found on the experiment are: magplano, nagpeperform, pagpapadala, pagpapaalala, paguusap, pagbigay, paganunsyo, and many more.

3.1.2.4 Threshold Adjustment

Decreasing and increasing the similarity threshold from 50% (default) to 20% and 80% affected the members of clusters. Reducing the threshold signifies a loose acceptance in determining if distance between two words are similar. Since the condition is more lenient, there are more related (variants) words captured in clusters. In fact, using FastText in this experiment dropped 114 entries into only six, with 99% of the responses clustered. Additionally, low threshold value enabled the clustering process to capture variants with far distances that were not clustered in the base experiment. Examples for this are the pairs dagdagan-add, maayos-ayusin, and mabigyan-provide. Although in this adjustment, there were instances wherein clusters mixed up ideas under a vaguely large topic such as medicine-cleanliness, barangay-typhoon, and cause-food.

Increasing the threshold on the other hand, tightens it, thus producing harder but more closely similar clusters. Since the condition is stricter, unrelated words were filtered, producing more accurate and clear manifestation of relationship between them. Positive samples separated evacuation-elevation, drill-drills, training-community, and government-council. Moreover, there were others that even 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-pagbabara*.

However, this restriction is not perfect, as there are instances that slight differences (just a letter in some) can lead to either good, bad, or unaffected clustering. Negative samples for this restriction was unable to cluster *kit-kits*, *training-taraining*, and *basura-basurahan*. In addition, the increase of threshold dispersed cluster members, where several clusters contain a single verb – meaning verbs must be the same to be clustered together.

Comprehensively for this experiment, it proves that increasing or decreasing the threshold could filter out, include, or basically place words in more appropriate clusters. Basically, it 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.

3.1.2.5 Comparison of Clustering Approaches

Three clustering approaches were applied in the experiments, namely Dice’s Coefficient, Word2Vec, and FastText. Dice’s coefficient cluster words with orthographic similarity using n-grams, while Word2Vec uses a vector space to cluster semantically. FastText incorporates the idea of the two to determine if words are related to each other.

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.

Since Dice’s coefficient investigate features of words through a set of characters, one difference between it and Word2Vec is vectors are positioned based on usage – so words that operate the same way are closer together in the space. However, usage does not always mean it produces ideal groupings. There were clusters that produced antonyms like tao-bagay and sakuna-sanhi, and members with relationship indistinguishable from each other.

In word embeddings approaches, both were able to cluster based on usages, thematic similarities, and even those with orthographic similarities. 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*, *improved*, and *improving* were captured; as well as, intra-word code-switching such as inform-mainform and provide-magprovide. Moreover, it also inherits Dice’s capability in clustering shortcuts (e.g., nagbbigay-magbigay, magtulong2-magtulongan), one that is not covered by Word2Vec.

Synthesizing results, characteristics in positive and negative aspects 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. Downside in Dice’s coefficient is its limitation on covering only in the confines of string similarity, where literal character distance matters. Downside in word embeddings approaches, there were still fragments of word variants scattered across different clusters. Particularly obvious instances that could have

[4] words that exists on both English and Filipino language

been included in clusters such as prepare-preparing, linisin-maglinis, giving-pagbibigay, announce-inform, and gamit-bagay. Undoubtedly, all approaches were able to fulfill their purposes, joining similar ideas together and in some instances formed one, interpretable idea in clusters.

3.1.3 Survey on API and Report

For this research, two outputs have been created, namely an Application Programming Interface (API) of the tool and a Report. In evaluating the outputs of this study, survey was performed on both with questions pertaining to the quality of outputs and discussions on user feedback was provided.

3.1.3.1 API Functionalities

API is defined as a collection of functions, intended for researchers and developers alike. It has nine modules, namely Data Utilities, Normalization, Language Identification, Filipino Part-of-Speech Tagger, Information Extraction, Information Organization, Information Clustering, Information Ranking, and Report Generation. As a whole, the functions under each module are intended to be useful in future researches or application not necessarily within Malasakit that involves the tasks related to the modules above.

3.1.3.2 Report Formatting

The report generated is a two-column Microsoft word document, with extracted insights/suggestions organized in two ways based on Information Organization. Each report contains three parts: introduction, insight list, and list of Malasakit Responses.

In the insight list, each entry contains fields such as Sentence ID (of cluster members), Frequency Counts (number of responses), Proposed Action (verbs extracted and clustered), and Target (nouns under verb clusters). Additionally, lexicalizations are formatted by enclosing them in parentheses.

3.1.3.3 Survey Procedure and Results

Prior to answering the questionnaires, an informed consent form was provided. Upon signing, two forms were presented. One intended for API assessment, a form designed from USE Questionnaire [10] and NormAPI's [12] version of the same. The other is intended for Report assessment designed from TAM Model [2]. Both questionnaires have a quantitative and qualitative part, measuring the agreement in certain aspects using numbers and acquiring comments and opinions, respectively.

The survey was conducted on five (5) *Malasakit* team members. In API assessment, the USE Questionnaire has four themes, namely Usefulness, Ease of Use, Ease of Learning, and Satisfaction. Its overall rating in each of the criterion produced a value of 4.3, 3.76, 3.8, and 4.4 out of five, respectively. As a whole, the API was rated with 4.07 out of five score.

In Usefulness, 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.

In Ease of Use, it got the lowest among others. In some counts, it has been described as unintuitive or not user-friendly, lacking enough documentation. Despite this, there are participants that still found the API easy and simple to use. Other distinguishable characteristic pertains to the modularity of the API's functions, where it got most of the participants agreeing that it is flexible.

In Ease of Learning, questions pertain to learning curve involved in using the tool; where four out of five participants agreed it is fast and easy to learn, 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.

In Satisfaction, it is described in the sense that it works as intended and if it would be recommended to others. Majority of the participants were strongly satisfied and positive with it.

In qualitative part of the assessment, most of the functions were tested by the respondents. With their appreciation to the tool, they have pointed out a series of improvements. Aspects to be improved involve better documentation, and potential extensions involve adding Graphical User Interface, packaging using pip install, and commercializing using REST API.

In assessing the generated report, it has been rated 4.25 out of five overall, a value that resides with agreeable to strongly agreeable elements. Based on the results and among the series of questions, 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 this, participants agreed that contents are clear, readable, and sufficient to 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 interpret the information – to conveniently push users to easily learn making use of the report.

Taking opinions regarding organizational preference, most preferred it by response categories. Although, one participant noticed an important element in clustering the entries entirely; 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."

Regarding the participants' preference in clustering approaches, four out of five preferred how the information were clustered through Dice's Coefficient. Due to the fact that it displays a clearer relationship on actions and targets through its orthographic similarities, as compared to semantic similarities which have instances of vaguely related and diverse clusters. Complementing with formatting by categories, the layout looked more presentable and easier to interpret. One participant, however, have a different opinion and pointed out FastText's output to be easier and clearer.

Exploring ideas to improve the report, their comments and suggestions were collected. The reports were judged to be "aesthetically formal and proper to look at." with the information understandable and actionable. However, some initially found it hard, 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. Suggestions to reduce this include the removal of entries with frequency counts of one. For the same purpose, entries could also be limited into the top 5, 10 or 25 entries.

Furthermore, an infographic or visualization that would give an overview in the contents of the report – particularly urgent steps to prevent and mitigate disasters, can be used as an alternative. With regards to content, few indicated refinements in clustering.

3.1.4 Test on News Dataset

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

The News dataset was taken from a Philippine language resource collection [16]. Sources used for this experiment are from Pang-Masa and Pilipino Star Ngayon 2015, both in Filipino with 2,000 sentences. There are four main categories, namely Entertainment (Movies and Showbiz), Sports, World, and Opinion. Having each category with 200 sentences, the highest number of contents is World category with 800 sentences.

In this dataset, sentences were processed through the tool's main tasks. Organization were implemented to process by category with prioritization set by alphabetically. Like previous experiments, all three approaches were used in clustering.

Based on the results, Information Extraction produced 3,503 insights from 1,689 sentences (311 sentences without insights). In Entertainment, phrases such as "Humihingi na siya ng pambili ng pagkain", "Nag-tweet si Vice", and "natuwa ang fans" were extracted. In Opinion, examples are "NAGULAT ang mga kidnaper", "masagot ang nasabing mga tanong", and "nakuha ang plaka". In Sports, extractions include "sinimulan na ang fitness test", "mapapanood ng live", and "sumabak si Pacquiao". In World, "inaprubahan ang bail petitions", "nakaresponde ang mga bumbero", and "patay sa ligaw na bala" are a few of the entries.

Based on 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 actions found on a sentence. 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", and "worked hard for the movie".

However, in some cases, the tagger failed to label 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" '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).

On the condition that extracting information in news domain has been successful to capture actions and targets, potential use in news/media setting has been considered. One of the few of ideas is insight phrases can provide a summary or gist of the contents in an article. By taking parts of the article, specifically 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.

Regarding results for clustering, Dice's coefficient was able to cluster verbs mainly but was ineffective to targets. It is due to a lot of 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 *biktima-biktimang* 'victim', player-players, alas-9-alas-2, and the likes.

Differentiating with Dice's coefficient, Word2Vec produced more but not perfect set of members for both verb and noun fields. It joined semantically related ideas that talks about movement (e.g., lumisan, iniwan, and umalis), topics about franchise (e.g., actor, character, and teleserye), notable names (e.g., Aquino, Bonifacio, and Sharon), sports (e.g., players, team, and championship), crime (e.g., suspek, pulis, biktima), places (e.g., Quezon, city, and mall), family (e.g., mag-asawa, ina, and anak), transportation (e.g., kotse, pasahero, and sakay), and other contextual similarities (e.g., kuwento-buhay, pasahero-sakay, and kaibigan-kapatid).

Applying FastText, results showed more ideas clustered together, significantly lowering the number of clusters – which most of the responses belonged to a cluster with members more than one. 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 present, with more varieties in terminologies as compared to the original dataset. Although this is the case, their negative qualities are also passed down, where it was still unable to combine closely related terminologies and instances that have uncertain relationship with each other were still clustered. Consequently, with the large quantity of sentences that were clustered, it was harder to interpret for its vague relationships.

4. Conclusion and Further Work

In this paper, the development of an automated insight extraction and organization was discussed. Results showed that the use of Part-of-Speech Information Extraction achieved satisfactory marks on standard metrics such as Precision (P), Recall (R), Accuracy (A), and F-Measure (F). Specifically, on insight phrases and word sets, scores as such were recorded as P = 70.09, R = 80.87, A = 72.57, F = 75.09, and P = 62.14, R = 74.06, A = 51.03, F = 67.58, respectively.

Other modules undergone experimentations and through analyses found their advantages and disadvantages. Results showed the effectivity in organizing and clustering, regardless of the approach. In assessing the entirety of the tool and its generated report, they were rated 4.07 and 4.25 out of five, respectively. 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.

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.

5. REFERENCES

[1] Bojanowski, P., Grave, E., Joulin, A. and Mikolov, T. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics, 5*, 135-146.

[2] Davis, F. D., Bagozzi, R. P., and Warshaw, P. R. 1989. User acceptance of computer technology: A comparison of two theoretical models.

[3] De La Cruz, A., Oco, N. and 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),* 1-5. IEEE.

[4] Go, M. P. and 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,* 81-88.

[5] 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),* 293-297. IEEE.

[6] Guha-Sapir, D., Hoyois, P., Wallemacq, P. and Below, R. 2017. Annual disaster statistical review 2016: The numbers and trends. *Brussels: Centre for Research on the Epidemiology of Disasters (CRED)*.

[7] Jurafsky, D. and Martin, J. H. 2018. *Speech and language processing.* Stanford.

[8] Loper, E. and Bird, S. 2002. Nltk: the natural language toolkit. *arXiv preprint cs/0205028*.

[9] Lui, M. and Baldwin, T. 2012. Langid. py: An off-the-shelf language identification tool. In *Proceedings of the ACL 2012 system demonstrations*, 25-30.

[10] Lund, A. M. 2001. Measuring usability with the use questionnaire. *Usability interface, 8*(2), 3-6.

[11] Mikolov, T., Chen, K., Corrado, G. and Dean, J. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

[12] Nocon, N., Cuevas, G., Magat, D., Suministrado, P., and Cheng, C. 2014. Normapi: An api for normalizing filipino shortcut texts. In *Proceedings of the 2014 international conference on asian language processing (IALP)*, 207-210.

[13] Nocon, N., Kho, N. M. and Arroyo, J. 2018. Building a filipino colloquialism translator using sequence-to-sequence model. In *TENCON 2018-2018 IEEE region 10 conference*, 2199-2204.

[14] 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),* 1-6. IEEE.

[15] Oco, N. and Borra, A. 2011. A grammar checker for tagalog using languagetool. In *Proceedings of the 9th workshop on asian language resources*, 2-9.

[16] Oco, N., Syliongka, L. R., Allman, T., and Roxas, R. E. 2016. Philippine language resources: Applications, issues, and directions. In *Proceedings of the 30th pacific asia conference on language, information and computation*, 433-438.

[17] Regalado, R. V., Kalaw, K. M. D., Lu, V., Dela Cruz, K. M. H. and Garcia, J. P. 2015. Filiet: an information extraction system for filipino disaster-related tweets. In *Proceedings of the DLSU research congress vol. 3*.

[18] Rehurek, R. and Sojka, P. 2010. Software framework for topic modelling with large corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 45-50. Valletta, Malta: ELRA. (http://is.muni.cz/publication/884893/en)