



Automatic Generation of Flood Reports for Twitter

Allan Borra, Zyra Rae de Grano, Jonathan Sun, Ma. Alexandra Valencia

Software Technology Department

College of Computer Studies

De La Salle University

allan.borra@dlsu.edu.ph; zyradg006@yahoo.com; jonathansun728@yahoo.com.ph; alexa.valencia@yahoo.com

Abstract: This paper discusses the implementation of an automatic report generation of flood reports from computer data from Philippine Flood Hazard Maps (nababaha.com). Data are in the form of crowd-sourced visual map reports of flooding in nababaha.com. These are gathered and automatically converted into sentences that are then broadcasted through Twitter. Template-based approach of Natural Language Generation (NLG) is used to transform the computer data into understandable texts and sentences. The generated texts are evaluated according to its effectiveness and understandability through a survey to measure fluency, accuracy of conveying desired meaning to the respondent, and how well the readers will respond to the generated text (task-based evaluation). From a scale of 1-5, results from 35 respondents show an average rating of 3.85 for accuracy, 3.875 for task-based evaluation, and 4.045 for fluency.

Key Words: Natural Language Generation;

1. INTRODUCTION

Among the 28 Asian countries, Philippines was ranked 2nd place in terms of the reported disasters in the year 2012 according to the United Nations office for Disaster Risk and Reduction (Panela, 2012). It is clear that the Philippines is greatly affected by heavy rainfalls these past few years (i.e. Ondoy, Pepeng, Pablo). Technology and innovation such as NOAH (noah.dost.gov.ph), allowed for better prediction and visualization of weather maps and patterns. This research intended to use Natural Language Generation Techniques to automatically generate flood reports to the public through the Twitter space as another mode of information.

It is asserted that informing citizens regarding current flood levels on certain areas, they will be able to respond effectively by avoiding these flooded places thus possibly reducing casualties.

Disaster awareness will be the focus of the MayBahaBa system. Its goal is to generate flood reports to the public who have access to Twitter on the state of flood in a certain place in the Philippines

2. RELATED SYSTEMS

2.1 Flood Patrol

Flood Patrol is a mobile application developed by the Ateneo Java Wireless Competency Center which monitors and reports floods in the Philippines. The application was made especially for



project NOAH, which has a flood reporting tool called the Philippine Flood Hazard Maps or www.nababaha.com. The web interface of this application is similar to the Philippine Flood Hazard Maps'.

The application basically allows the user to input flood data such as height, location, time and comments if necessary. The data is then uploaded to the web for users to view. The mobile version contains a map of the Philippines that allows the user to zoom and view more specific places. The map contains dots which are color coded. Each dot color represents flood and height of the flood. The mobile application has a feature that allows the user to input an image that contains the flooded area. The images are then validated by authorities. Also, the application has a feature that allows the user to view the reports on the map, and this could be filtered for a certain height and/or date (Evangelista, 2012).

2.2 BahaBa: A Disaster Management Route Generator for Mobile Systems

BahaBa is a route generation system for Lian, Batangas that delivers its output via SMS. The SMS contains the shortest path to the nearest places where flood is not present. The system uses A* algorithm in generating shortest paths from different places in Lian, Batangas. The system only generates routes if a user enters a correct request syntax and received by the server. Globe's SMS/MMS API is used to implement the delivery of the output of the system through SMS. The barangay captains provide the community map, list of hazardous areas, and the list of safe places. The text generated are made understandable through the use of the Template-based approach of Natural Language Generation. The generated texts are written in Filipino language. The places (where the user must go) and the directions (left or right) are contained on the text generated by the system. The delivery of the SMS messages will depend on the users' requests (Haw, Martinez, Santiagué, & Tamayo, n.d.).

2.3 Philippine Flood Hazard Maps (nababaha.com)

The Philippine Flood Hazard Maps (nababaha.com) was developed by the researchers from the University of the Philippines and the IBM. It uses FLO-2D and FEMA (Federal Emergency Management Agency) in simulating and locating flooded places in the Philippines. Flood hazard maps are shown to provide information on where not to be during extreme rainfall. A map where citizens are allowed to report by inputting the current level of flood in an area is also provided (Philippine Flood Hazard Maps: How We Do It, n.d.).

3. SOLUTION ARCHITECTURE: GENERATING TWITTER REPORTS FROM CROWD-SOURCED DATA

The overall architectural design for the system is shown in Fig. 1. Crowd-sourced flood data in a KML file will be extracted from Philippine Hazards Maps. This file contains the date and time of flood, the flood height and its coordinates as reported by citizens. To get the actual address name from the coordinates, the database is checked whether there are address names given the exact coordinates, else, Google Reverse Geocoding Services from Google Maps will be used. The output from Google Reverse Geocoding is a JSON file containing the address name (street or avenue, city, region) given the coordinates. All the data from the KML and JSON files will be stored in the database. Given all the data, template-based form of natural Language Generation is used to generate flood reports. These Reports will then be broadcasted on Twitter using different public accounts. Each account will represent each region in the Philippines this is so that users need only follow the twitter accounts of regions they are concerned with.

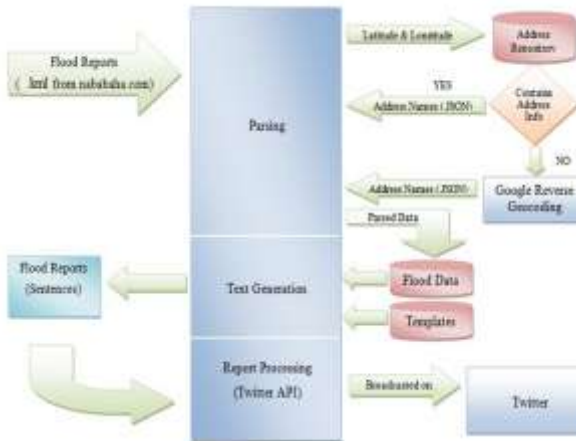


Fig. 1. Maybahaba System Architecture

3.1 Data Requirements and Format

3.1.1 KML File from Philippine Flood Hazard Maps

The Philippine Flood Hazard Maps (nababaha.com) has a Flood Reporting site where users may report a flood on a specific area in the Philippines as well as view existing reports from other people. The flood reports of year 2013 can be downloaded by accessing this url: <http://noah.dost.gov.ph/nababaha-2013-reports.kml>. The reports are updated whenever a user reports a flood on the site. The downloaded .kml file is written in XML format and it contains the time, date, place (represented in coordinates) and the height of the flood. Below is a sample data in the .kml file:

```
<Placemark>
  <name>Knee high</name>
  <description><![CDATA[<div
width="100"><b>Flood Height</b>: Knee
high<br/><b>Flood Date</b>: Jan 10, 2013 6:00
AM<br/><br/><b>Details</b>: G. Flores Avenue going
to<br/> office.. flood is knee</div>]]></description>
  <styleURL>#floodHeightIcon2</styleURL>
</Point>
```

```
<coordinates>125.543241236,8.95058469234,0
</coordinates>
</Point>
</Placemark>
```

Fig. 2. A Snippet or Portion of a Sample KML File

The year on the url can be changed depending on the desired year of the flood reports. The reports are then to be downloaded in .kml file which is named 'nababaha-<year>-reports.kml'.

3.1.2 JSON File from Google Reverse Geocoding

Reverse Geocoding is the process of converting geographic coordinates into an address. This is a special feature from the Google Geocoding Services, and is readily available for the public. For a Reverse Geocode, the request must be in the form: <http://maps.googleapis.com/maps/api/geocode/output?latlng=> where parameter output can either be json or xml, and latlng are latitude-longitude coordinates. Reverse Geocoding will be used in the system to retrieve the names of the places from input coordinates coming from json files of Philippine Flood Hazard Maps.

Below is a sample request URL for the coordinates given in the KML File in Figure 1.

<http://maps.googleapis.com/maps/api/geocode/json?latlng=8.95058469234,125.543241236&sensor=true>

The JSON File that will be produced using this request will contain information about the address of the place, including the region and the formatted address of the coordinates which is needed by the system. Below is a part of the JSON File generated through this request:

```
{
  "results" : [
    {
      "address_components" : [
        {
          "long_name" : "Ester Luna Street",
          "short_name" : "Ester Luna Street",
          "types" : [ "route" ]
        },
        {
          "long_name" : "Butuan City",
          "short_name" : "Butuan City",
          "types" : [ "locality", "political" ]
        }
      ]
    }
  ]
}
```

```

    },
    {
      "long_name": "Agusan Del Norte",
      "short_name": "Agusan Del Norte",
      "types": [ "administrative_area_level_2",
        "political" ]
    },
    {
      "long_name": "Caraga",
      "short_name": "Caraga",
      "types": [ "administrative_area_level_1",
        "political" ]
    },
    {
      "long_name": "Philippines",
      "short_name": "PH",
      "types": [ "country", "political" ]
    }
  ],
  "formatted_address": "Ester Luna Street, Butuan
  City, Philippines",
  "geometry": {
    "bounds": {
      "northeast": {
        "lat": 8.950765199999999,
        "lng": 125.54310270
      },
      "southwest": {
        "lat": 8.9501290,
        "lng": 125.54306540
      }
    }
  }
},

```

Figure 3. A Snippet or Portion of a Sample JSON File

3.2 Modules and Processes

3.2.1 Parsing Module

KML and JSON Files will be analyzed to get the features that will be needed. The data extracted from these files will be stored in a database. Parsing is also called syntactical analysis, it is also the analyzation of strings and symbols in a language.

The data that can be extracted from KML File are the date of flood, height of flood, and the coordinates where the flood occurred. After the KML File is extracted, the coordinates will now be requested to Google Reverse Geocoding. Google Reverse Geocoding will then return a JSON File which will be extracted as well. The data that can be

extracted from the JSON File are formatted_address, street_address, route, intersection, political, country, administrative_area_1, administrative_area_2, administrative_area_3, colloquial_area, locality, sublocality, neighborhood, premise, subpremise, postal_code, natural_feature, airport, park, point_of_interest, post_box, street_number, floor, room. (*The Google Geocoding API, 2012*) However, not all of the components will be used. Only the formatted address and administrative_area1(region) will be stored in the database.

3.2.2 Database

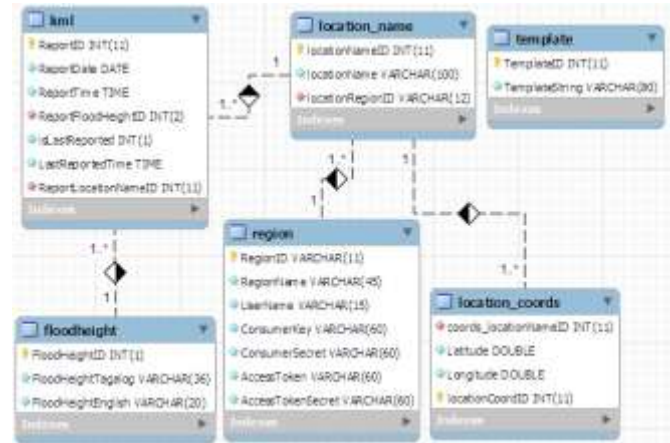


Figure 4. MayBahaBa EERD

The database will contain the extracted data from the KML and the JSON File. Coordinates and its corresponding address will be stored in the database to avoid duplicated requests.

3.2.2.1 KML Table

The KML table contains the data from the downloaded KML file from the Philippine Flood Hazard Maps (nababaha.com). Table 1 explains each column in the kml Table.

Table 1. KML Table

Column Name	Data Type	Description
ReportID	INT	Primary key of the table



ReportDate	DATE	Date when the flood was reported
ReportTime	TIME	Time when the flood was reported
ReportFloodHeightID	INT	Foreign key that describes the height of the flood
ReportLocationID	INT	Foreign key that describes the location of the flood
isLastReported	INT	0 if not reported/ not last reported, 1 if last reported
LastReportedTime	TIME	Time of last report

3.2.2.2 location Table

The location Table contains the information about the location of the flood. Google Reverse Geocoding is used to gather the name of the place. Table 2 explains each column in the location Table.

Table 2. location Table

Column Name	Data Type	Description
LocationID	INT	Primary key of this table
LocationName	VARCHAR	The name of the location
LocationLat	VARCHAR	The latitude value of the location
LocationLong	VARCHAR	The longitude value of the location
LocationRegionID	VARCHAR	A foreign key that describes the region of the location

3.2.2.3 floodheight Table

The floodheight Table contains all the possible heights of the flood. It contains the flood height description and the flood height number (0-8). Table 3 explains each column in the floodheight Table.

Table 3. floodheight Table

Column Name	Data Type	Description
FloodHeightID	INT	Primary key of this table. The primary key serves as the flood height number
FloodHeightTagalog	VARCHAR	Description of the flood in Filipino
FloodHeightEnglish	VARCHAR	Description of the flood in English

The floodheight Table has pre-defined contents provided in Figure 5.

FloodHeightID	FloodHeightTagalog	FloodHeightEnglish
0	Walang baha	No flood
1	Hanggang bukung-bukong	Ankle high
2	Hanggang tuhod	Knee high
3	Hanggang baywang	Waist high
4	Hanggang leeg	Neck high
5	Hanggang tuktok ng ulo	Top of the head high
6	May taas na isang palapag	1 storey high
7	May taas na isa't kalahati...	1.5 storeys high
8	May taas na dalawang p...	2 storeys or higher

Figure 5. floodheight Table Contents

3.2.2.4 region Table

The region Table contains the region name and region number. Table 4 explains each column in the region Table.

Table 4. region Table

Column Name	Data Type	Description
-------------	-----------	-------------



RegionID	VARCHAR	Primary key of this table
RegionName	VARCHAR	The name of the region
UserName	VARCHAR	Twitter username of the region
ConsumerKey	VARCHAR	Consumer key of the region
ConsumerSecret	VARCHAR	Consumer secret of the region
AccessToken	VARCHAR	Access token of the region
AccessTokenSecret	VARCHAR	Access token secret of the region

The region Table has pre-defined contents provided in Figure 6.

RegionID	RegionName	UserName	ConsumerKey	ConsumerSecret	AccessToken	AccessTokenSecret
ARMN	Autonomous Region in Muslim Mindanao	MRE_ARMN	g2fARqg8p...	LR5vndF7c19...	1188172136	z00u3Uz0z0H
CAR	Cordillera Administrative Region	MRE_CAR	oDKD3d8p8...	793D4BdvC...	1188233342	5pRfYUy1p4G
ICR	National Capital Region	MRE_ICR	Zgn7mngp1...	oMe8289K0...	1188236284	2EkgP4W4u6b
Region I	Ilocos Region	MRE_REGION1	WSPRM086...	WJGQ7G7H...	1188147188	W4J3DPGvWf0
Region II	Cagayan Valley	MRE_REGION2	ad1E7e1m...	zF51oc71Q3...	1188211134	3z4V4w3ar5C
Region III	Central Luzon	MRE_REGION3	Z0K7Y1T7g...	81844X0C4T...	1187883881	W2M1Ww3C97
Region IV-A	CALABARZON	MRE_CALABARZON	ar0yP4U0...	oc0Y70yaf2...	1187850588	W1lzmw4E254
Region IV-B	MIMAROPA	MRE_MIMAROPA	Y2w5GmV7R...	Z1w5Y4V1V...	1187834886	D1u4H4E6w4d
Region V	Davao Region	MRE_REGION5	oG2P4hF1L...	1p3D1d4h4c5...	1207087588	u0490K9L4c0
Region VI	Iloilo	MRE_REGION6	w4u4P4w2h...	3H4P487uE...	1187855442	yy0w483G04w
Region VII	Western Visayas	MRE_REGION7	0PL4oc4y4...	3Fz1317u4a3...	1187854738	W7P4054GepU0
Region VIII	Central Visayas	MRE_REGION8	oL3H4D14z...	U4w4G4u4L...	1207081851	4H7C34u4P4d
Region IX	Eastern Visayas	MRE_REGION9	W4AC705P...	4L4w4Z2W...	1207120483	3P4w444W4W
Region X	Northern Mindanao	MRE_REGION10	3w4G4u4Z...	31Y4u4w4P...	1207120488	36454w4w4W
Region XI	Davao Region	MRE_REGION11	Re4G4u4W...	W204G4P4w4E...	1207122336	o3P4w44554Y
Region XII	Southwest	MRE_REGION12	3w4E4V4J...	5z4w4Y4u4N...	1208110367	48P4W442P4d
Region XIII	Cebu	MRE_REGION13	EE4U4k4o...	4W4T43C215...	1208112851	w4P4w44D4w4

Figure 6. region Table Contents

3.2.2.5 template Table

The template Table contains the templates that will be used in generating the reports. Table 5 explains each column in the template Table.

Table 5. template Table

Column Name	Data Type	Description
TemplateID	INT	Primary key of this table
TemplateString	VARCHAR	The template that will be used for the reports

3.2.3 Templates

The following templates will be used for the reports that are newly added in the Philippine Flood Hazard Maps:

Template 1: Sa kasalukuyan, <level of flood> ang baha sa <place/s> . <hashtags>

-This template states the current level of flood in a certain place/s.

Template 2: Bumaba ang baha sa <place/s>, ito'y <level of flood> na lamang . <hashtags>

-This template states the current level of flood that is lower than previous flood report.

Template 3: Tumaas ang baha sa <place/s>, ito'y <level of flood> na . <hashtags>

- This template states the current level of flood is higher than the previous flood report.

Template 4: Sa kasalukuyan, walang baha sa <place/s>. <hashtags>

-This template states that the area has been reported to have no floods.

Template 5: Sa kasalukuyan, walang nang baha sa <place/s>. <hashtags>

-This template states that the flood in an area has subsided.

The following templates will be used to generate sentences for the reports that are not updated after an hour it was reported (It will contain the same flood report with the time it has been previously reported):

Template 6: Noong <time>, <level of flood> ang baha sa <place/s>. <hashtags>

- This template states that a flood of a certain level in an area has been reported on a certain time.

Template 7: Noong <time>, bumaba ang baha sa <place/s>, ito'y <level of flood> na lamang . <hashtags>

- This template states that the flood level on a certain place was lower compared to the previous report.

Template 8: Noong <time>, tumaas ang baha sa <place/s>, ito'y <level of flood> na. <hashtags>



- This template states that the flood level on a certain place was deeper compared to the previous report.

Template 9: Wala nang baha sa <place/s> noong <time> pa. <hashtags>

- This template states that the flood in a certain place was gone on a certain time.
- This template states that the flood in a certain place was gone on a certain time.

3.2.3 Twitter

The Philippines was ranked 10th place for having a big number of Twitter users. The social media Semicast stated that out of the 517M Twitter users, 9.5M came from the Philippines. According to the Smart Communication, Twitter was used as a medium to broadcast and deliver reports by thousands of Filipino users in cases of floods and rains. Orlando Veja, the Smart chief wireless advisor, added that even though places in the Philippines were flooded and did not have electricity, the Filipino netizens still managed to access the internet. Veja also mentioned about how active the Filipino netizens are to the internet and the accessibility of Twitter; mobile phones can now be used to access social networking sites such as Twitter aside from computers (Montecillo, 2012).

Twitter is used as a medium to broadcast the generated flood reports in certain areas in the Philippines. A Twitter account will represent each region in the Philippines, so that users need only to follow the twitter account of regions they are concerned with. The newly generated flood reports, which data will come from the Philippine Flood Hazards Maps, will be broadcasted twice, first immediately and the next after 15 minutes.

Since twitter only allows 140 characters per tweet, TwitLonger is used for tweeting messages beyond the said constraint.

3.2.3.1 Twitter4J: Java API Library

Twitter4J is used as the Java library to allow the system to tweet flood reports without accessing the official Twitter site. To be able to use the library completely, each Twitter account should be registered to the Developers site: <https://dev.twitter.com/apps/new>. The Developers site will then generate the consumer key, consumer secret, access token and access token secret and

these will be used to be able to post on twitter. (Twitter4J, 2013)

4. RESULTS AND OBSERVATIONS

The following are the two types of testing used by the system: (1) Software Testing, specifically Incremental Integration Testing, was done to test the performance of the system's modules. (2) Text Generation Evaluation was also done to test the correctness and grammar of the generated texts.

4.1 Software Testing

Incremental Integration Testing was performed during the implementation of the system. Modules were made and tested individually. These modules are then combined with other modules and tested to work according to the requirements. Testing on the fully integrated system was conducted through the use of different cases and scenarios that may happen during the actual use of the system. Data used for these different scenarios were made by the programmers to ensure that all the different cases will be covered. Tests were conducted for each module

1st Case : The database is initially empty. (No reports)

A user reports a flood level on a certain place. The KML file is then updated and the new entry reads:

```
<Placemark>
  <name>Ankle high</name>
  <description><![CDATA[<div
width=+100+><b>Flood Height</b>: Ankle
high<br/><b>Flood Date</b>: Apr 11, 2013 09:00
AM</div>]]></description>
  <styleUrl>#floodHeightIcon1</styleUrl>
  <Point>

    <coordinates>125.548125302,8.95127341711
,0</coordinates>
  </Point>
</Placemark>
```

This entry states that on 9:00AM, April 11, 2013 (today), there is an Ankle high flood in the coordinates 125.548125302,8.95127341711.

The image below shows that the coordinates first stored in the database for possible future use.

coords_locationNameID	Latitude	Longitude	locationCoordID
58	8.95058469234	125.543241238	81
59	10.707495134	122.948143409	82
60	9.03707529729	125.572727772	83
60	9.0369970921	125.567052216	84
61	8.94814891957	125.547670576	85
62	8.94777797526	125.552533421	86
63	7.10969511066	125.490271303	87
61	8.9470352444	125.546345566	88
61	8.94799729509	125.546372388	89
64	8.9483982342	125.545219038	90
65	14.621882157	121.071483782	91
66	8.94333549627	125.529057715	92
67	17.8042132053	120.456000023	93
68	8.9512329851	125.54094896	94
69	8.94741233874	125.542247149	95
70	8.95127341711	125.548125302	96
71	7.74903215201	126.469350059	97

Figure 7: Location_coords table

Through the use of Google Reverse Geocoding, the system is able to obtain the address name of the provided coordinates. The image below shows the address name of the coordinates 125.548125302, 8.95127341711 retrieved from Google Reverse Geocoding.

locationNameID	locationName	locationRegionID
58	Ester Luna Street, Butuan City	Region XIII
59	Felix Amante Avenue, Bacolod City	Region VI
60	Pan-Philippine Highway, Butuan City	Region XIII
61	Magsaysay Bridge, Butuan City	Region XIII
62	Baan Riverside Road, Butuan City	Region XIII
63	Davao-Bukidnon Road, Davao City	Region XI
64	P. Burgos Street, Butuan City	Region XIII
65	Pasaje Dela Paz, Quezon City	NCR
66	I. Elloso St, Butuan City	Region XIII
67	Manila North Road, Sinait	Region I
68	Montilla Blvd., Butuan City	Region XIII
69	Jose C. Aquino Avenue, Butuan City	Region XIII
70	Magsaysay Viaduct, Butuan City	Region XIII
71	Surigao-Davao Coastal Road, Baganga	Region XI

Figure 8: Location_name table

After all necessary data have extracted from the KML file and the address name has already been obtained, these data are all stored in the database

ReportID	ReportDate	ReportTime	ReportRoadHeightD	WasReported	LastReportedTime	ReportLocationNameID
41	2013-04-11	06:30:00	2	1	17:32:09	58
42	2013-04-11	16:00:00	4	1	17:32:09	59
43	2013-04-11	01:00:00	4	0	00:00:00	60
44	2013-04-11	01:00:00	5	1	17:32:09	60
45	2013-04-11	14:30:00	0	0	00:00:00	61
46	2013-04-11	13:30:00	0	1	17:32:09	62
47	2013-04-11	00:15:00	3	1	17:32:09	63
48	2013-04-11	15:00:00	4	0	00:00:00	61
49	2013-04-11	16:30:00	4	1	17:32:09	61
50	2013-04-11	14:45:00	4	1	17:32:09	64
51	2013-04-11	06:00:00	0	1	17:32:09	65
52	2013-04-11	09:00:00	1	1	17:32:09	66
53	2013-04-11	08:15:00	1	1	17:32:09	67
54	2013-04-11	06:00:00	2	1	17:32:09	68
55	2013-04-11	06:00:00	1	1	17:32:09	69
56	2013-04-11	09:00:00	1	1	17:32:09	70
57	2013-04-11	09:30:00	3	1	17:32:09	71

Figure 9: KML table

After the data has been stored in the database, it is inserted into its appropriate template to create the flood report. Since the flood occurred on the same day but on 9:00 AM, before the program was run, and it is the first time that the flood on that location was reported today, the program will use the 6th template as shown below.

TemplateID	TemplateString
1	Sa kasalukuyan, <level> ang baha sa <place>. <hashtag>
2	Bumaba ang baha sa <place>, ito'y <level> na lamang. <hashtag>
3	Tumaas ang baha sa <place>, ito'y <level> na . <hashtag>
4	Sa kasalukuyan, walang baha sa <place>. <hashtag>
5	Sa kasalukuyan, walang nang baha sa <place>. <hashtag>
6	Noong <time>, <level> ang baha sa <place>. <hashtag>
7	Noong <time>, bumaba ang baha sa <place>, ito'y <level> na lamang. <hashtag>
8	Noong <time>, tumaas ang baha sa <place>, ito'y <level> na. <hashtag>
9	Wala nang baha sa <place> noong <time> pa. <hashtag>

Figure 10: Template Table

Finally after the flood report is constructed, it is reported in its appropriate Twitter account. Since the flood was reported in Region XIII, Maybahaba Region XIII is used to report the flood as shown below.



Figure 11: Flood Reported on Twitter.

In case that there are reports with the same ReportLocationNameID (Address Name) and

ReportTime, the system will choose the report that has the higher flood height as shown in Figure x. isLastReported column contains 1 if it is already reported ones, 0 if its not reported and 2 if it is already reported twice. In Figure x, report 44, is chosen to be reported since it has a higher ReportFloodHeightID.

ReportID	ReportDate	ReportTime	ReportFloodHeightID	isLastReported	LastReportedTime	ReportLocation
41	2013-04-11	06:00:00	2	1	17:32:09	58
42	2013-04-11	15:00:00	4	1	17:32:09	59
43	2013-04-11	01:00:00	4	0	00:00:00	60
44	2013-04-11	01:00:00	5	1	17:32:09	60
45	2013-04-11	14:00:00	0	0	00:00:00	61
46	2013-04-11	13:30:00	0	1	17:32:09	62
47	2013-04-11	00:15:00	3	1	17:32:09	63
48	2013-04-11	15:00:00	4	0	00:00:00	64
49	2013-04-11	15:30:00	4	1	17:32:09	64
50	2013-04-11	14:45:00	4	1	17:32:09	64
51	2013-04-11	06:00:00	0	1	17:32:09	65
52	2013-04-11	09:00:00	1	1	17:32:09	66
53	2013-04-11	08:15:00	1	1	17:32:09	67
54	2013-04-11	06:00:00	2	1	17:32:09	68
55	2013-04-11	06:00:00	1	1	17:32:09	69
56	2013-04-11	08:00:00	1	1	17:32:09	70
57	2013-04-11	05:00:00	3	1	17:32:09	71

Figure 12: KML Report Sample with Same ReportTime, ReportDate, ReportLocationNameID

Figure 13 shows that Report 44 was reported to Twitter.



Figure 13: Sample Report in Twitter

Another case considered is that in the same KML File, there are reports with the same ReportLocationNameID, ReportFloodHeightID and ReportDate. The system will choose the report that has the latest ReportTime. In figure x, Report 49 was reported to Twitter because its ReportTime is 30mins higher than Report 48.

ReportID	ReportDate	ReportTime	ReportFloodHeightID	isLastReported	LastReportedTime	ReportLocationNameID
41	2013-04-11	06:00:00	2	1	17:32:09	58
42	2013-04-11	15:00:00	4	1	17:32:09	59
43	2013-04-11	01:00:00	4	0	00:00:00	60
44	2013-04-11	01:00:00	5	1	17:32:09	60
45	2013-04-11	14:00:00	0	0	00:00:00	61
46	2013-04-11	13:30:00	0	1	17:32:09	62
47	2013-04-11	00:15:00	3	1	17:32:09	63
48	2013-04-11	15:00:00	4	0	00:00:00	64
49	2013-04-11	15:30:00	4	1	17:32:09	64
50	2013-04-11	14:45:00	4	1	17:32:09	64
51	2013-04-11	06:00:00	0	1	17:32:09	65
52	2013-04-11	09:00:00	1	1	17:32:09	66
53	2013-04-11	08:15:00	1	1	17:32:09	67
54	2013-04-11	06:00:00	2	1	17:32:09	68
55	2013-04-11	06:00:00	1	1	17:32:09	69
56	2013-04-11	08:00:00	1	1	17:32:09	70
57	2013-04-11	05:00:00	3	1	17:32:09	71

Figure 14: KML Report Sample with the same ReportFloodHeightID, ReportDate and ReportLocationNameID

Figure 15 shows that Report 49 was reported to Twitter.



Figure 15: Sample Report in Twitter

2nd Case: The database has contents.

No recent update on the flood level of a location:

If a flood has been reported and the system was able to tweet the flood report, if after 1 hour, there had been no reports of flood in that location, the report will simply be retweeted on twitter. The image below shows a previous report being retweeted on twitter.



Figure 16: Retweeted Flood Report

There is an update on the flood level of a location:

If the system detects that there is an update about the level of flood in a certain location. It will tweet the updated report on Twitter. In Figure X, the past report is Report 47 has a reportTime of 00:15:00 which is 12:15 AM, the updated report(Report 62) has a reportTime of 2:00:00 which is 2:00AM. The

system should report Report 62, saying that the flood subsided already.

ReportID	ReportDate	ReportTime	ReportFloodHeightID	isLastReported	LastReportedTime	ReportLocation
45	2013-04-11	14:30:00	0	0	00:00:00	81
46	2013-04-11	13:30:00	0	2	18:41:28	82
47	2013-04-11	01:15:00	3	1	17:32:39	83
48	2013-04-11	15:30:00	4	0	00:00:00	81
49	2013-04-11	15:30:00	4	1	17:32:39	81
50	2013-04-11	14:45:00	4	2	18:41:28	84
51	2013-04-11	06:00:00	0	2	18:41:28	85
52	2013-04-11	09:00:00	1	2	18:41:28	85
53	2013-04-11	02:15:00	1	2	18:41:28	87
54	2013-04-11	06:00:00	2	2	18:41:28	88
55	2013-04-11	06:00:00	1	2	18:41:28	89
56	2013-04-11	09:00:00	1	2	18:41:28	70
57	2013-04-11	09:30:00	3	2	18:41:28	71
58	2013-04-11	07:30:00	4	0	00:00:00	59
59	2013-04-11	14:00:00	0	0	00:00:00	59
60	2013-04-11	02:30:00	6	0	00:00:00	60
61	2013-04-11	14:00:00	2	0	00:00:00	61
62	2013-04-11	01:30:00	0	1	18:41:28	62
63	2013-04-11	14:30:00	7	0	00:00:00	61
64	2013-04-11	15:30:00	1	0	00:00:00	61

Figure 17: KML Sample Data on Updated Report

Figure 17 shows that the tweet says that the flood in Davao-Bukidnon Road subsided.



Figure 18: Sample Tweet

The update on the flood level of a location is not recent.

If the system detects an update on a location, it first checks whether the update is recent, specifically if the time of the update is greater than that of its previous report. The image below shows that the report time of Report 59, which is the update of the previous report, Report 42, is less than the time of the previous report. So instead of reporting the update Report, the system chooses to retweet the previous report.

ReportID	ReportDate	ReportTime	ReportFloodHeightID	isLastReported	LastReportedTime	ReportLocationNameID
41	2013-04-11	06:00:00	2	1	17:32:39	80
42	2013-04-11	15:00:00	4	2	18:41:28	59
43	2013-04-11	01:00:00	4	0	00:00:00	80
44	2013-04-11	01:00:00	5	1	17:32:39	80
45	2013-04-11	14:00:00	0	0	00:00:00	81
46	2013-04-11	13:30:00	0	2	18:41:28	82
47	2013-04-11	01:15:00	3	1	17:32:39	83
48	2013-04-11	15:30:00	4	0	00:00:00	81
49	2013-04-11	15:30:00	4	1	17:32:39	81
50	2013-04-11	14:45:00	4	2	18:41:28	84
51	2013-04-11	06:00:00	0	2	18:41:28	85
52	2013-04-11	09:00:00	1	2	18:41:28	85
53	2013-04-11	02:15:00	1	2	18:41:28	87
54	2013-04-11	06:00:00	2	2	18:41:28	88
55	2013-04-11	06:00:00	1	2	18:41:28	89
56	2013-04-11	09:00:00	1	2	18:41:28	70
57	2013-04-11	09:30:00	3	2	18:41:28	71
58	2013-04-11	07:30:00	4	0	00:00:00	59
59	2013-04-11	14:00:00	0	0	00:00:00	59
60	2013-04-11	02:30:00	6	0	00:00:00	60
61	2013-04-11	14:00:00	2	0	00:00:00	61

Figure 19: KML report sample with report on a location which is not recent

The image below shows the previous tweet, as well as the new tweet which is simply the retweet version of the previous report



Figure 20: Sample Tweets

4.2 Evaluation of Generated Texts

In order to evaluate the overall effectiveness and understandability of the generated texts, it needs to pass certain measures or criteria. And so, the generated texts were evaluated by the respondents using three key criteria. First, in terms of its accuracy or extent to which the generated text conveys the desired meaning to the reader. It was also subjected to task based evaluation, or how well the readers will respond after reading the generated texts. Lastly, it was evaluated in terms of fluency, extent to which the text flows and is readable. (Mellish & Dale, 1998).

The evaluation was successfully completed through conducting a survey with thirty-five (35) respondents. The respondents deemed the five



sample sentences generated by system to be accurate. This indicates that the sentences were clear and comprehensive. The respondents have also evaluated that generated texts were useful in terms of its purpose, which in this case is informing citizens about the state of flood. Lastly, The respondents have agreed that the texts' use of grammar and its fluency is appropriate. To sum up, the evaluation of the respondents implies that the system was successful in producing actual flood reports in text which are understandable, conveys appropriate information, and contains correct choice of words.

5. CONCLUSION AND RECOMMENDATIONS

The MayBahaBa system was able to generate human readable flood reports and these flood reports were also successfully posted on twitter for the public to view. The flood report produced by the system is entirely dependent on the data it gathers from nababaha.com and Google Reverse Geocoding. The speed in which the system would be able to gather its data from these sources would also heavily rely on the Internet speed of the machine where the system will run. The system was also able to compensate for the limited number of request allowed in Google Reverse Geocoding by storing previous requests to the database. The system, however, stops from posting tweets if it reaches the maximum limit of 100 tweets per hour/ 1000 per day. This is done to avoid the Twitter accounts be blocked by Twitter from posting tweets for a long period of time.

The generated texts from the system were evaluated using human rating through conducting a survey. The respondents have evaluated and most have agreed that the five texts which were generated by the system are accurate, proficient in terms of task evaluation and have good fluency and grammar.

The MayBahaBa system's flood data are crowd-sourced. Thus, the accuracy of the data cannot be determined. A recommendation would be to gather flood data from other sources such as the input of Local Government Units (LGUs). Tweets regarding flood could also be collected and analyzed for flood

data. The amount of data gathered by the system gathers continuously, and so these data could be used to find patterns in flooding, and as such, data analytics could play a part through the use of the data collected. Another recommendation would be to use more means to distribute flood reports other than twitter. Other media such as television, SMS, and radio could also be useful as a broadcasting tool. Lastly, It is recommended to test the system further, possibly by having a test deployment of the system. The system was not thoroughly tested on a crowd-by-the-millions scale.

6. REFERENCES (use APA style for citations)

- “Desperate relief effort continues as death toll in Philippines flood rises”. Daily mail reporter. Retrieved December 2011, from <http://www.dailymail.co.uk/news/article-2077014/Philippines-flash-flood-2011-Death-toll-rises-officials-continue-relief-operation.html>.
- “Definition: Twitter jail” Urban Dictionary. Retrieved in 2013 from <http://www.urbandictionary.com/define.php?term=Twitter%20Jail>.
- Evangelista, P. (2012, October). Ateneo develops mobile app for flood monitoring and reporting. <http://www.admu.edu.ph/index.php?p=120&type=2&sec=25&aid=11337>.
- Haw, M., Martinez, A., Santiagué, L., & Tamayo, P. (n.d.). Bahaba: A disaster management route generator for mobile systems.
- Json file. (n.d.). Retrieved from <http://maps.googleapis.com/maps/api/geocode/json?latlng=8.95058469234,125.543241236&sensor=true>
- Mellish, C., & Dale, R. (1998). Natural language generation. <http://clt.mq.edu.au/~rdale/publications/papers/1999/mellishdale1999.pdf>.
- Montecillo, P. (2012, 12). Philippines has 9.5m twitter users, ranks 10th.



Presented at the DLSU Research Congress 2014
De La Salle University, Manila, Philippines
March 6-8, 2014

<http://technology.inquirer.net/15189/philippines-has-9-5m-twitter-users-ranks-10th>

Nababaha.com flood reports in 2013. (n.d.). Retrieved from <http://noah.dost.gov.ph/nababaha-2013-reports.kml>

Panela, S. F. (2012). PHL is 2nd most disaster-prone country in Asia -study. Retrieved from <http://www.gmanetwork.com/news/story/287954/scitech/science/phl-is-2nd-most-disaster-prone-country-in-asia-study>

“Definition of Twitter”. PCMAG.COM Encyclopaedia. Retrieved 2013 from http://www.pcmag.com/encyclopedia_term/0,1237,t=Twitter&i=57880,00.asp

Philippine flood hazard maps: How we do it. (n.d.). <http://nababaha.com/howwedoit.htm>.

The google geocoding api. (2012, October). <https://developers.google.com/maps/documentation/geocoding/>

Twitter4j - a java library. (n.d.). <http://twitter4j.org/en/index.html>.