



Automated Sentiment Analysis of Commentaries in a HCP-Moderated Page

Rick Jason G. Obrero
Iris Thiele C. Isip-Tan

University of the
Philippines Manila

Contact!



rgobrero@up.edu.ph



@JasonObrero



SENTIMENT ANALYSIS

This refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials.



PROJECT SUMMARY

Project Summary

- Gather comments from the highest commented status in Endocrine Witch.
 - Polycystic Ovary Syndrome (PCOS).
- Perform sentiment analysis on such comments.
 - Classify as either positive, negative or neutral.
 - These are based on certain keywords.

A large teal graphic element consisting of a diagonal line that divides the page into a white upper-left section and a teal lower-right section.

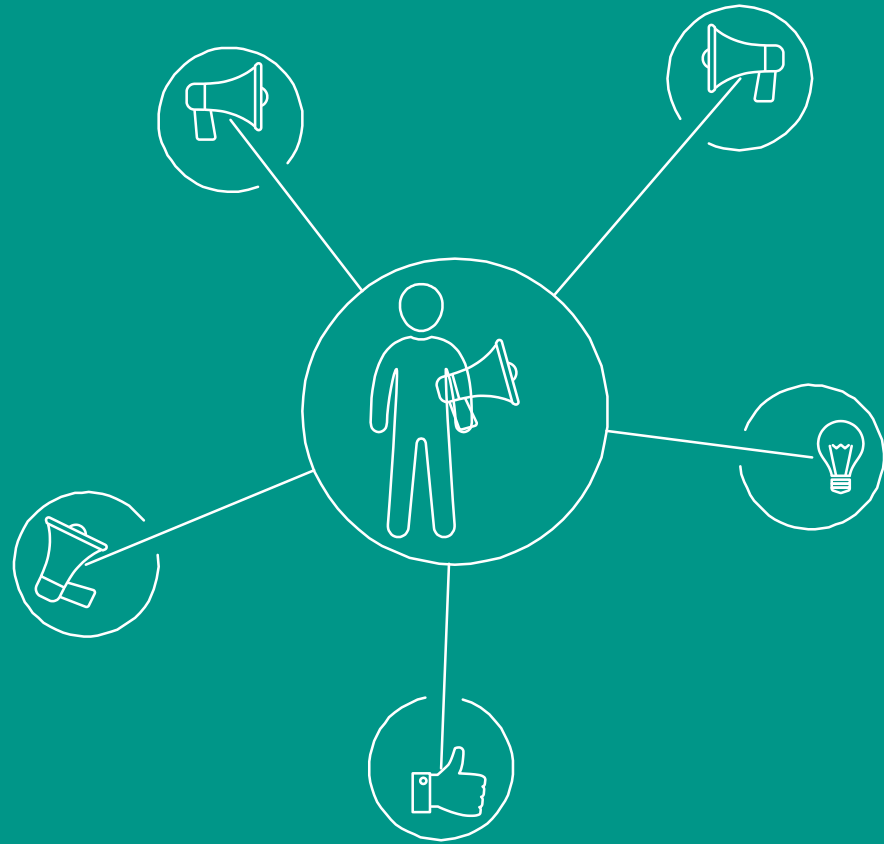
STATEMENT OF THE PROBLEM

A large teal-colored shape on the right side of the slide, consisting of a dark teal triangle pointing towards the top right and a lighter teal trapezoidal shape below it.

“

It is hard to filter all the comments present in Facebook Posts, especially when it is a page involving public service.

Filtering posts

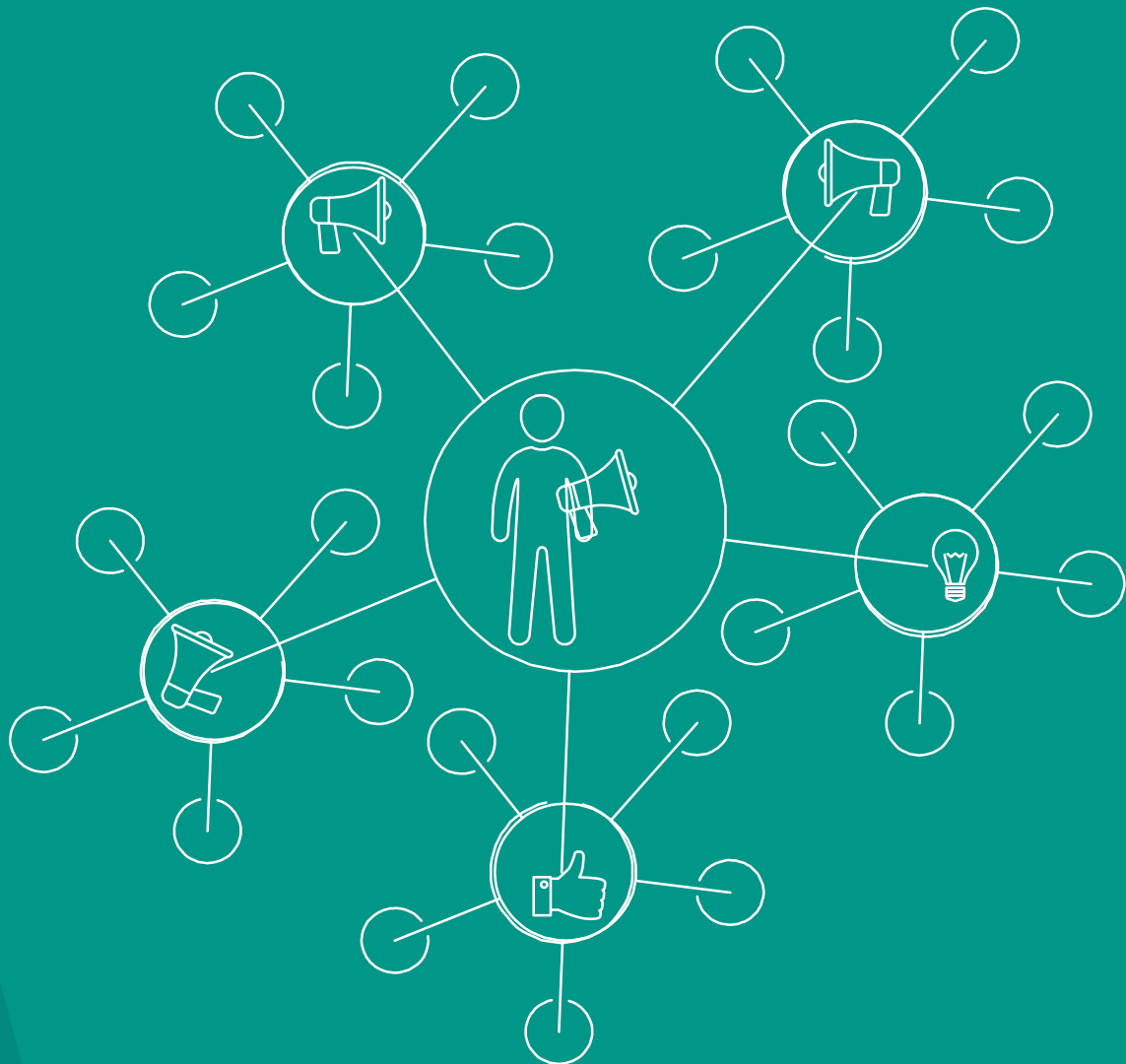


A large teal-colored shape in the top right corner of the slide, consisting of several overlapping diagonal bands.

“

To complicate the situation, existing comments in Facebook can be replied. This creates harder context switching in terms of posts.

Switching contexts



A large teal-colored shape in the top right corner of the slide, consisting of several overlapping diagonal bands.

“

Due to these problems, the lack of time in giving comprehensive comments in important queries is expected.



“

When this problem is resolved, it shall be in a form of public service of giving the commenters to have their queries answered in the fastest possible time.

METHODOLOGY

A thick, bright yellow diagonal stripe runs from the top right corner towards the bottom left, separating the white background on the left from a solid yellow background on the right.

1.

**DATA
COLLECTION**

Data Collection

Gather all of the status updates in Endocrine Witch, upon approval from the Facebook Page Manager.

Among them, find the status update with the most number of comments.

Data Collection

On the side, interview Page Manager for potential words that indicate positive, negative or neutral responses.

A thick, bright yellow diagonal stripe runs from the top right corner towards the bottom left, separating the white background on the left from a solid yellow background on the right.

2.

**DATA
PROCESSING**

Data Processing

Gather the comments and translate them from Filipino to English via Google Translate, and remove all words which were not translated.

Data Processing

Get overall sentiment, as well as the sentiment for each comment.

A large, solid green diagonal shape that starts from the top right and extends towards the bottom left, creating a split background of white and green.

RESULTS AND DISCUSSION

RESULTS AND DISCUSSION

Comments
Collected

Comments Collected

3,443 comments as of 24 March

3,300 unique persons

Comments Collected

Latest of the comments were posted 24 March 2016.

Oldest comment was on 9 August 2015, the day when the post was created.

Comments Collected

The comment with the highest number of likes was posted on 9 August 2015, 5:19pm, with 617, as of comment collection date.

Comment Pre-Processing

Code-switching is prevalent.

Code-switching is the use of different languages or dialects in a single sentence.

Comment Pre-Processing

Some words in the comments are also being placed in shorthand.

This is the act of removing or changing certain letters in the word in place of the correct spelling of the word.

Comment Pre-Processing

These produce the limitations of the Google Translate API.

Some further steps are needed to be included in the preprocessing of the comments.

Comment Pre-Processing

Context is being lost due to the direct translation of such words.

Comment Pre-Processing: Examples

"yr"

A common abbreviation for the word "year".

Comment Pre-Processing: Examples

"q"

A shorthand for the Filipino word "ko", and can be translated to the word "my".

Comment Pre-Processing: Examples

"kc"

A shorthand for the Filipino word "kasi", and can be translated to the word "because".

Comment Pre-Processing: Examples

"dinadatnan"

A Filipino word for the event of menstruation.

Comment Pre-Processing

Context is being lost due to the direct translation of such words.

This is due to the word-per-word nature of the translation. Filipino phrases may have a more direct translation as a single or multiple English word than a phrase formed by translating the words one by one.

Comment Pre-Processing: Examples

"hanggang ngayon"

A Filipino word, that when translated is "up to now".

Comment Pre-Processing

Context is being lost due to the direct translation of such words.

Another reason is the use of medical jargon that was being provided to them by their doctors while not understanding the meaning behind it.

Comment Pre-Processing: Examples

"irreg"

A shorthand for irregular menstruation.

Comment Pre-Processing

Context is being lost due to the direct translation of such words.

Another reason is the spelling of words in different dialects. As a cardinal rule in the Filipino language, on what the pronunciation is, that will be its corresponding spelling.

Comment Pre-Processing: Examples

"minstruition"

A misspell for menstruation.

RESULTS AND DISCUSSION

Processing via
Sentiment
Analysis

Processing via Sentiment Analysis

sentR, an R package was used to do sentiment analysis.

It contains a word dictionary which identifies whether words elicit a positive, negative or neutral response. Also, the subjectivity of the words are identified whether they are strong or weak.

sentR/subjectivity.csv a... x +

GitHub, Inc. (US) | https://github.com/mananshah99/sentR/blob/master/inst/extdata/subjectivity.csv

Search

217	alarming	strongsubj	negative
218	alarmingly	strongsubj	negative
219	alas	strongsubj	negative
220	alienate	weaksbj	negative
221	alienated	weaksbj	negative
222	alienation	weaksbj	negative
223	alloy	strongsubj	positive
224	allegation	strongsubj	negative
225	allegations	strongsubj	negative
226	allege	weaksbj	negative
227	allergic	weaksbj	negative
228	alleviate	weaksbj	positive
229	allow	weaksbj	positive
230	allowable	weaksbj	positive
231	allure	strongsubj	positive
232	alluring	strongsubj	positive
233	alluringly	strongsubj	positive
234	ally	strongsubj	positive
235	almighty	strongsubj	positive
236	aloof	strongsubj	negative
237	altercation	weaksbj	negative
238	although	weaksbj	negative
239	altruist	strongsubj	positive
240	altruistic	strongsubj	positive
241	altruistically	strongsubj	positive
242	amaze	strongsubj	positive
243	amazed	strongsubj	positive
244	amazement	strongsubj	positive
245	amazing	strongsubj	positive
246	amazingly	strongsubj	positive

Word dictionary in sentR

Processing via Sentiment Analysis

sentR does not consider words that are not present in the dictionary, thus the concerns on translation are already directly addressed.

Processing via Sentiment Analysis

Using the dictionary, the comments are classified as either positive, negative or neutral.

Sentiment

Sentiment	Number of Comments
positive	2,834
negative	607
neutral	2
TOTAL	3,443

RESULTS AND DISCUSSION

Generating
Themes

Context for Endocrine Witch

As of 28 April 2016:

- 50,211 unique page likes
- 84% female
- Majority in 25-34 year age bracket at 40%
- Mostly from the Philippines (Manila, Cebu and Davao), while there are present in Saudi Arabia, United Arab Emirates and Hong Kong.



“

To offer reliable information
about endocrine disease from a
board-certified endocrinologist.

RESULTS AND DISCUSSION

Responding to
Stakeholders:
KAP's

Knowledge, Attitudes and Practices

Positive responses are elicited from people who comment when they ask more information about a topic of interest.

Knowledge, Attitudes and Practices

Comments become negative responses when the tone of discussion already becomes a private health issue consultation.

Knowledge, Attitudes and Practices

Comments become negative responses when sensitive information is already disclosed, such as photos of private body parts and laboratory test results.

In this instance, the comment gets hidden from view.

Knowledge, Attitudes and Practices

When a specific brand name is also mentioned, the comment becomes negative in nature.

In this event, when the comment-poster is persistent, the person is banned from interacting in the page.

Knowledge, Attitudes and Practices

Comments are answered in the simplest way possible, as the persons interacting have little to no background in endocrinology.

Knowledge, Attitudes and Practices

Recommendations in seeing a doctor are also being laid out to the persons commenting.

RESULTS AND DISCUSSION

Sentiment
Analysis in
Themes

Sentiment Analysis in Themes

Positive words:

- "Ask"
- "More"
- "Information"

And their synonyms

Sentiment Analysis in Themes

Negative words:

- "Consult"
- "Sex"
- "Menstruation"
- "Endorse"
- "Prescribe"
- "Supplements"
- "Medicine"

And their synonyms

Sentiment Analysis in Themes

Synonyms of such words were searched in an online copy of the WordNet thesaurus.

Table 2. Top five translated positive comments and their respective scores

No	Comment	Score
1	PCOS is the first ever in dysmenorrhoea	38.671884
2	hi ask ko lng anu kc drug may rarely lng ako ngreregla to time a year akng menstruation sometimes months depending ur lng yon pasibble I PCOS single poh me my age	0.057897
3	I dismissed my doc being irregular and the mentreation ngyn diagnose stage I endometrial cancer b	0.040512
4	doc poh aq irregular menstruation q hindi rin poh poh aq nbuntis mnsan times aq ngkaroon gang of men a year ntakot poh poh aq kz may mysakit n aq	0.060241
5	lastyr I did delay me mens almost mos then nagpacheck up in ultrasound I just thought I preggy that I then was not sorted prescribed me medication pamparegla then after a month I nagmems I done now april LMP I till now nothing again my mens what the heck should I do it again I hope I makarecieve message from you thank you very much	3.600339

Sentiment analysis in positive comments

Table 3. Top five translated negative comments and their respective scores

No	Comment	Score
1	hello ask ko lang po n yr old girl who finished continuous means losing the menstruation for the whole month what is the reason and ways of how to treat i thank you	1.768592
2	Hello Dr. I just wanted to seek advice normally po ang <u>meanstration</u> my year Napo My husband and I hndi we still <u>kakaanak</u> and hindi pa po ako <u>nkaranas</u> mg take the pills I want to mg baby we <u>owepayat</u> rising the age I am now nPO nkg wife, I said years old and I should have done for you baby mg n we	75.028647
3	doc before poh regular I dinadatnan but when either pills then q give birth sometime lang and when to further inject aq every month at the health center every month aq dinadatnan done a month before sometimes nude sometimes not be po ba aqng to p check up my PCOS well because it widens well aq aq waist q	37.115834
4	to experience Kona also delay the yr severe and almost months I have terrible this disease <u>ndi</u> greedy but bikis kodin fat meron po ako na <u>magtake</u> of medicine so I tried regular exercise and diet good thing was regular mens but I also I almost stop mons I <u>dinnakapag</u> exercise according ako sa month delay as diet diet so I am not able to correct discipline and of course also check up anyway to put godbless girls and I hard to get pregnant	0.462807
5	before I doc experienced qoh month delay I aqo law once <u>months</u> finished Mibsam days lng menstruation qoh nong june delay please aqo of months <u>thought</u> po namin wife qoh hndi <u>aqo</u> pregnant peru nong nag months <u>aqong</u> delay the mercy of God who loved na poh aq	3.527186

Sentiment analysis in negative comments

Conclusion

Conclusion

There are delays in getting the posts from the Endocrine Witch Facebook Page.

Reasons of which include:

- The controls which the Facebook Graph API introduced recently, such as the limit of getting Facebook posts from an unlimited amount to ten at a time.

Conclusion

There are delays in getting the posts from the Endocrine Witch Facebook Page.

Reasons of which include:

- Getting posts is different than that of getting the comments in the post.

Conclusion

There are delays in getting the posts from the Endocrine Witch Facebook Page.

Reasons of which include:

- The sheer amount of data in which the Endocrine Witch has, including all likes and comments, as well as posts are huge in nature.

Conclusion

The time consumed to get the most number of comments in Endocrine Witch is 35 seconds.

Conclusion

The files which contains the synonyms has a file size of 24 MB. Due to this sheer size, the time it consumes to load the file lasts for about ten seconds.

Conclusion

Reading comments will be faster, but the context of the posts will be lost.

Classification contained false positives and true negatives.

Conclusion

It depends on the Page Manager on what to reply on such instances.

Conclusion

Lack of awareness transcends to those in the persons who ask questions which are answerable in the post. Also, consultations to their physician is also recommended all the time, to ensure personalized prescriptions.

Recommendations

Recommendations

Use a better pre-processing algorithm to accommodate the following:

- 1) Conversion of emojis and emoticons on what appropriate response they actually translate.

Recommendations

Use a better pre-processing algorithm to accommodate the following:

- 2) Transform code-switched words to their appropriate English counterparts.

Recommendations

Use a better pre-processing algorithm to accommodate the following:

- 3) Transform shorthanded text to their spelled out counterparts.

Recommendations

Use a better pre-processing algorithm to ease the translator to convert local texts into the English language.

Recommendations

Use another translator aside from Google Translate.

Google Translate API has a 60-day trial that needs to be used right away.

Recommendations

Better sentiment analysis tools and dictionaries are needed to obtain better analysis.

