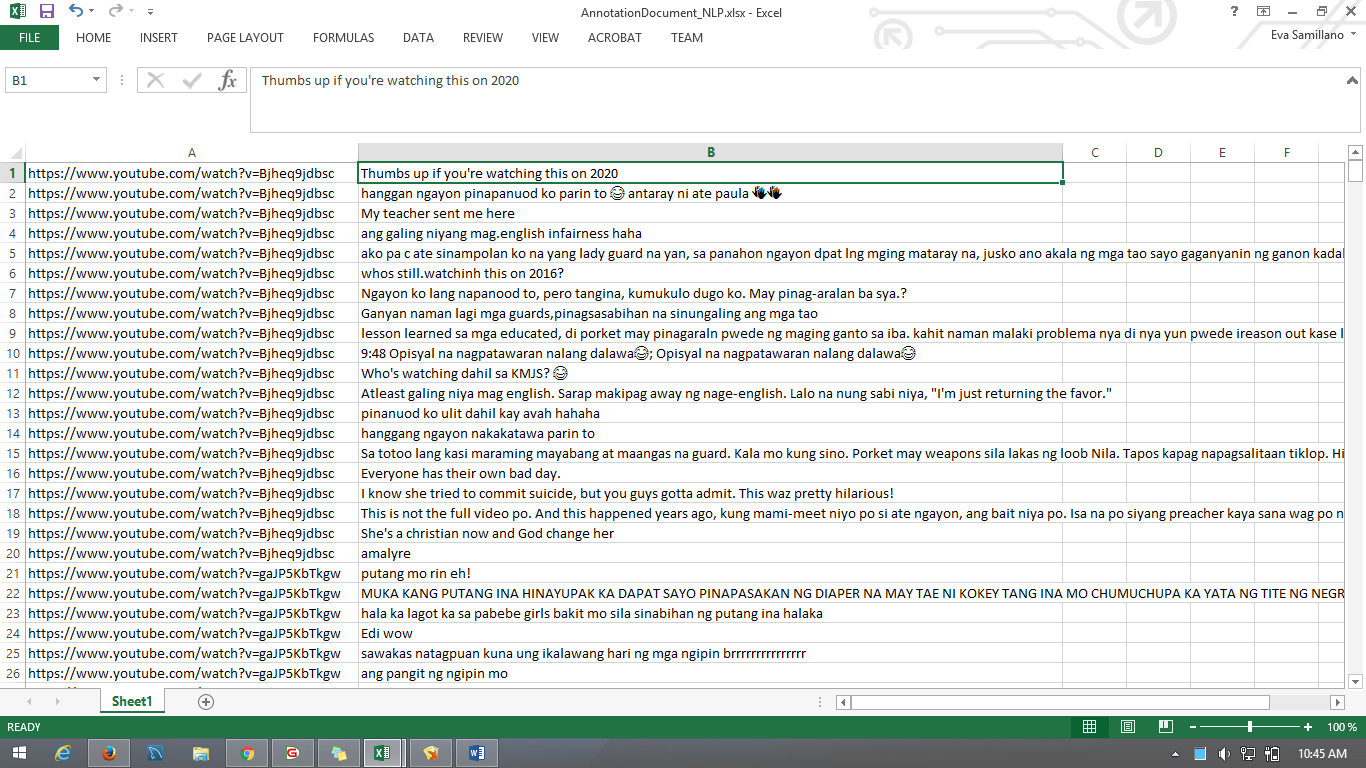
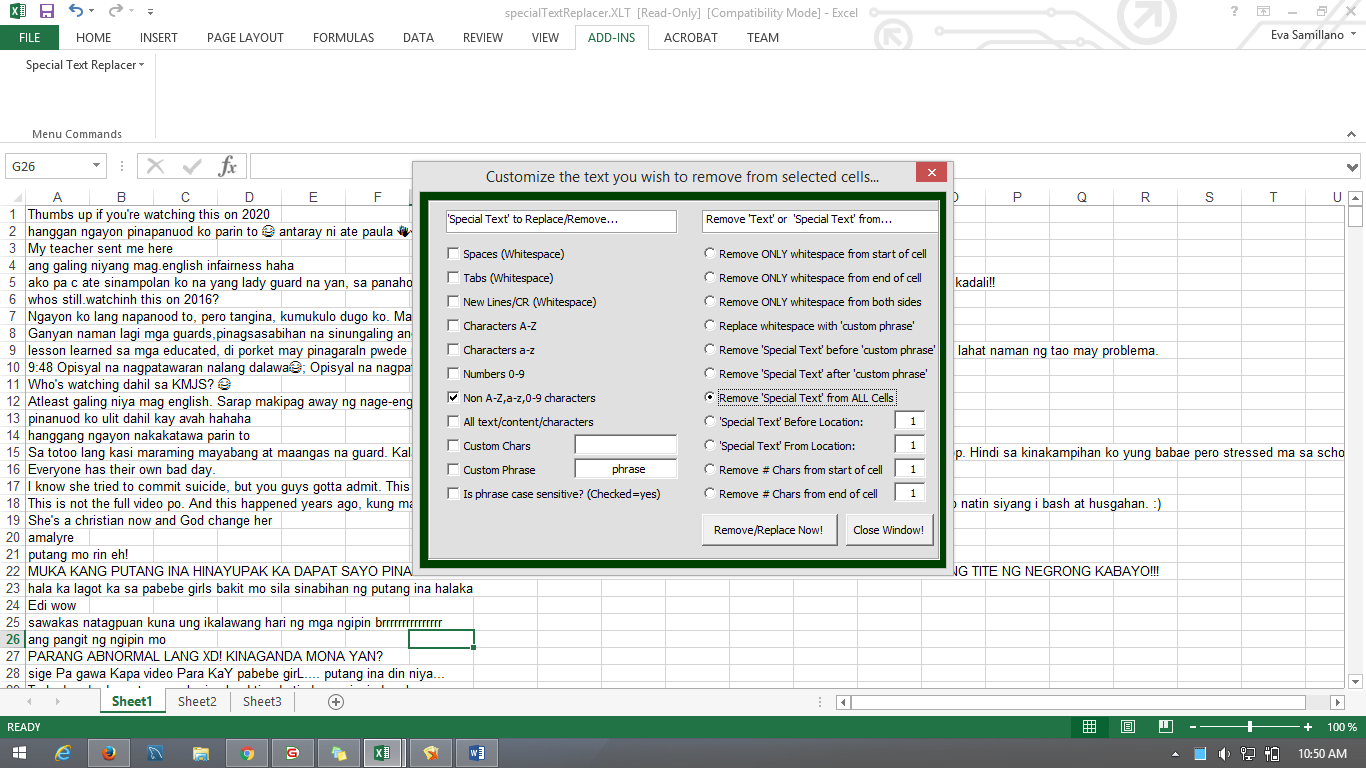
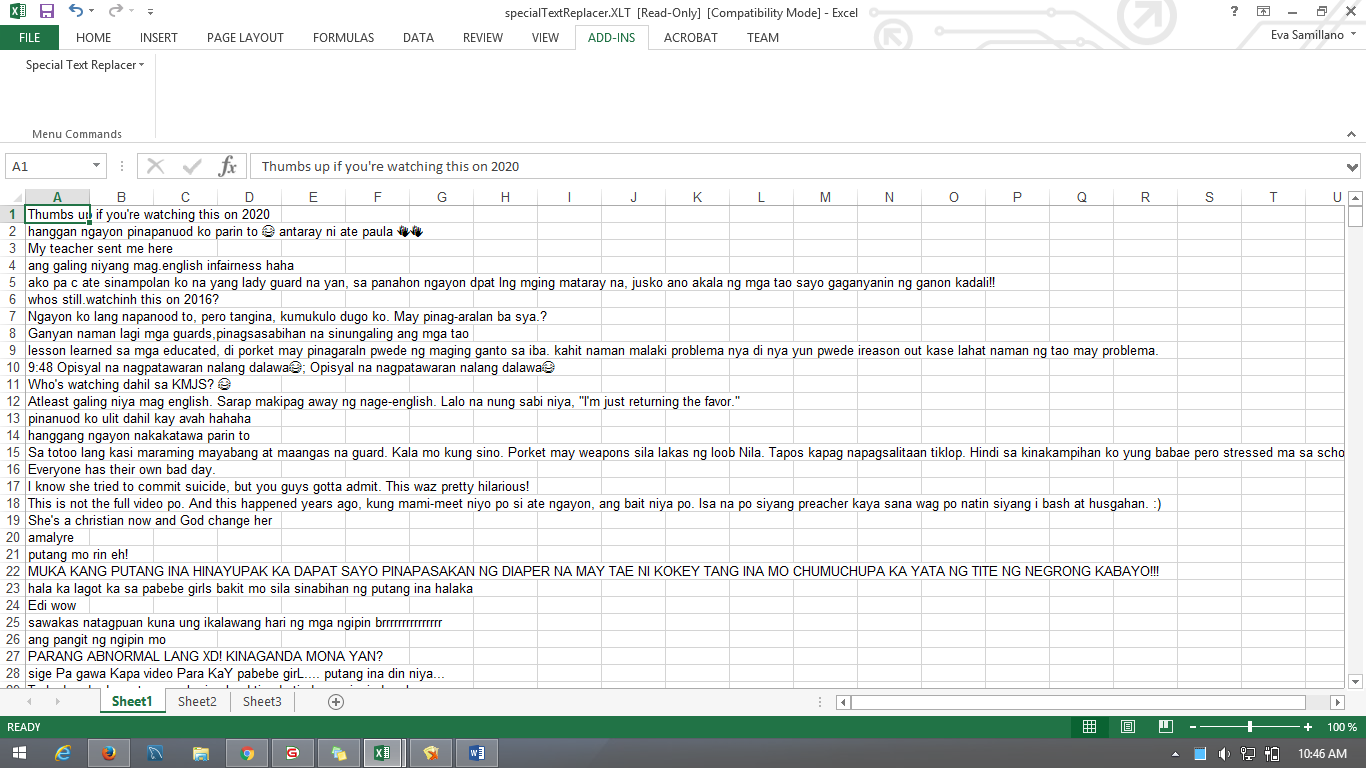
How we annotate?

First we get some data (comments and pots) from different popular social media such as Facebook, twitter, YouTube and ask.fm. **Using Import.io software tool online** we import data from unknown Filipino people, unknown conversation and most recent topics/issue in the Philippines.

After to get the data we compress it in **Normal Excel Form** (Figure1) and Copy it to **Excel Delete Replace and Remove Special Text Spaces and Characters from Multiple Cells Software** (Figure2)**,** then *Add-Ins* Menu we choose *Special Text Replacement,* then click *Remove Special Text Control Panel* from there you can choose what do you want to remove and replace in your excel file.

***Figure1: Normal Excel***

***Figure2: Excel Delete Replace and Remove Special Text Spaces and Characters from Multiple Cells Software***

***Figure: 3 Control Panel to customize which do you want to remove or replace***

**These symbols were removed so they will not be extracted as features later on.**

After arranging the document that we will be using for annotation, we started annotation in Gate – General Architecture Text Engineering. We first annotate how the sentence nearly to a cyberbullying, we categories was it into 3 different level (Figure4) :

HarmnessLevel\_0 – In this level, there is no cyberbullying occur, its typical or just giving opinion in a good way comments and post.

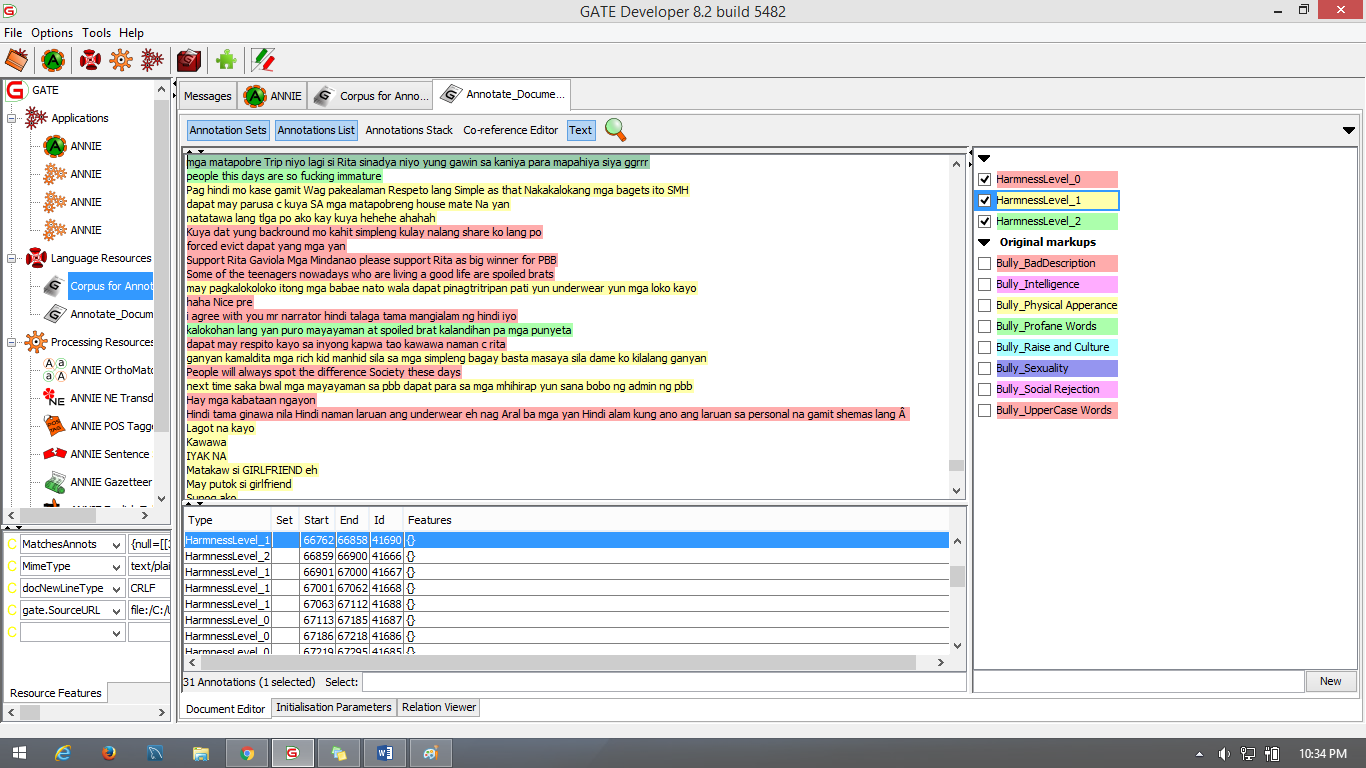
HarmnessLevel\_1 – In this level, some words could be define as cyberbullying but words is not really specify the exact word to define that was a bully like curse words. Example Word for this level are (Edi Wow, Hahahaha, nakakatawa sya, parang abnormal, panget mo, mataba ka).

HarmnessLevel\_2 – In this level, you will really notice the strong occurrence of cyberbullying because curse is already here, different threats and abusing to the person he/she describing is also here.

Cyberbullying Result in Harmness:

Out of 625 Sentence that we imported from different Social Media

|  |  |  |
| --- | --- | --- |
| Harmness Level | Score | Percentage (%) |
| HarmnessLevel\_0 | 215 | 34.4% |
| HarmnessLevel\_1 | 250 | 40% |
| HarmnessLevel\_2 | 160 | 25.6% |



***Figure 4: Using GATE – Annotating the Harmness of Sentence***

After ranking the cyberbullying, we next annotate the different kinds of bully in words. The following are the type we use to classify words in cyberbullying (Figure5):

Bully\_Sexuality – In this category, it contains censored words (puking ina, pepe, burat, chuchupa, etc.)

Bully\_Physical Apperance – In this category, it contains physical look of a person (mataba, payat, panget, mukang unggoy, negro etc.)

Bully\_Intelligence – In this category, it contain about the comprehension of a person (bobo, tanga, shunga, utak talangka, idiot, stupid etc.)

Bully\_Raise and Culture – In this category, it contains about how people bully the culture and religion activities of other religion because for other they see it weird (Ex. Demonyo, antichrist, mga adik, illuminati, kulto etc.)

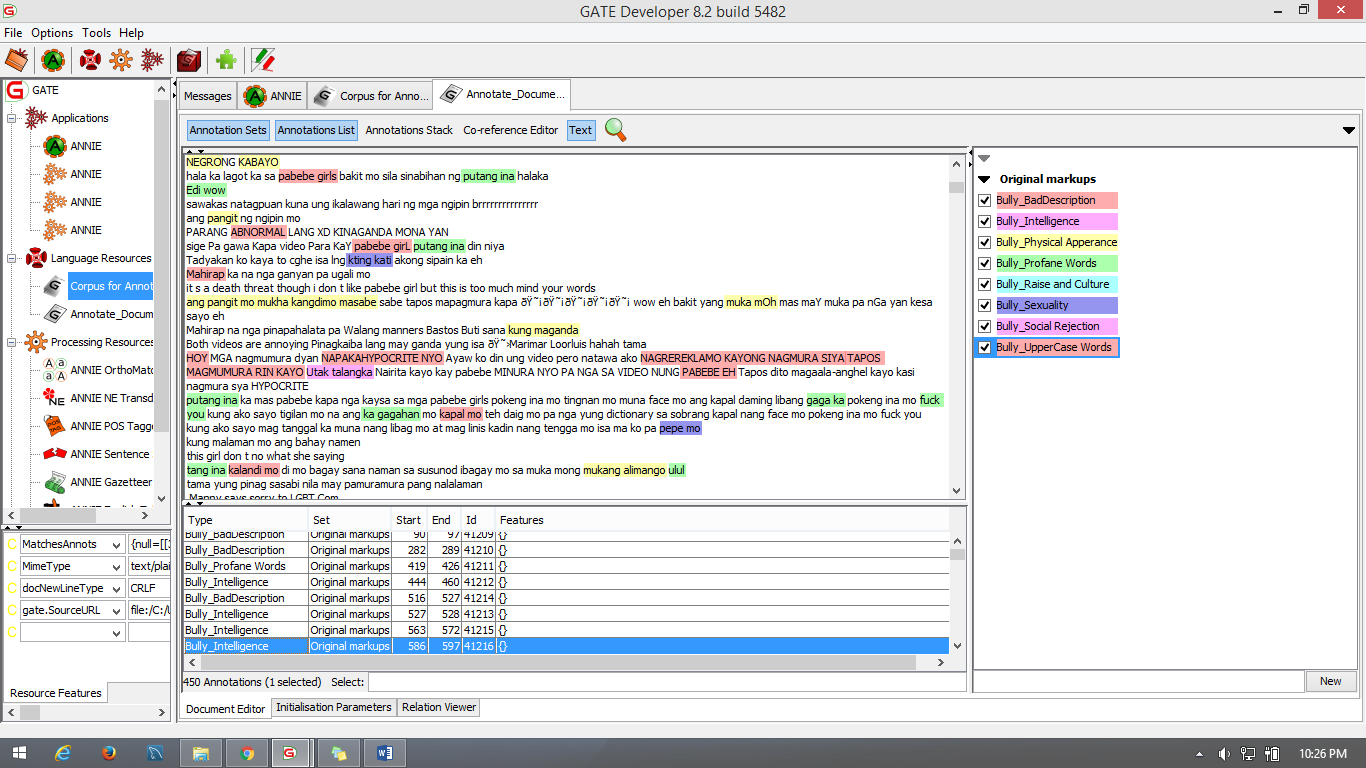
Bully\_Social Rejection – In this category, it contain how people bully other people using the words edi wow, hahahaXD, abnormal, nakakatawa ka, nakakahiya ka and etc. We also include here how people threat other like mamatay ka na, sipain kita, pakulong kita and etc. We include this words in social rejection because it show like people hate this person, or they describe and want to do what they said to this person literally.

Bully\_Bad Description – In this category, it contains how people bully other person using not good descriptions (Ex. Mayabang, ahas ka, plastic, maldita, adik, pokpok) this words could also use in upper case words and physical appearance. Some issue that you may in counter if you’re going to categories the word which is possible to be bad description or physical appearance for example, pokpok is a bad description but it could be also physical appearance if the person really shows identities of being a pokpok.

Cyberbullying Result in Classified Word:

We annotate 261 words out of 69814 words and sets of numbers

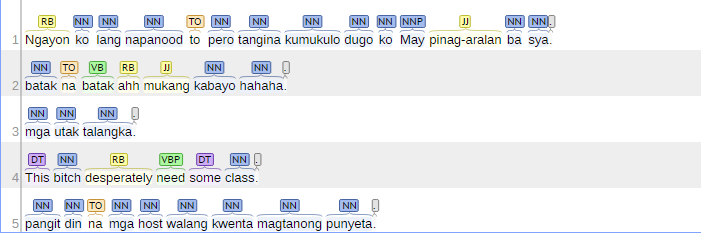
|  |  |  |
| --- | --- | --- |
| Bully Words | Score | Percentage (%) |
| Bully\_Bad Description | 83 | 31.8% |
| Bully\_Intelligence | 41 | 15.7% |
| Bully\_Physical Appearance | 43 | 16.5% |
| Bully\_Raise and Culture | 16 | 6.1% |
| Bully\_Sexuality | 30 | 11.5% |
| Bully\_Social Rejection | 48 | 18.4% |

***Figure 5: Using GATE - Annotating Bully Category***

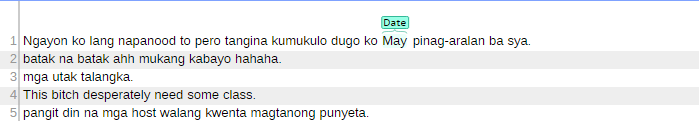
Lastly, we annotate the sentence using the Brat Annotation Tool in Stanford CoreNLP feature of Brat.

Part of Speech - Part of Speech is one of the feature that Stanford CoreNLP can annotate visualize by Brat. Stanford using Penn Tree Bank Part of Speech which is became the entity for this section.

|  |  |  |
| --- | --- | --- |
| Number | Tag | Description |
| 1 | CC | Coordinating conjunction |
| 2 | CD | Cardinal number |
| 3 | DT | Determiner |
| 4 | EX | Existential *there* |
| 5 | FW | Foreign word |
| 6 | IN | Preposition or subordinating conjunction |
| 7 | JJ | Adjective |
| 8 | JJR | Adjective, comparative |
| 9 | JJS | Adjective, superlative |
| 10 | LS | List item marker |
| 11 | MD | Modal |
| 12 | NN | Noun, singular or mass |
| 13 | NNS | Noun, plural |
| 14 | NNP | Proper noun, singular |
| 15 | NNPS | Proper noun, plural |
| 16 | PDT | Predeterminer |
| 17 | POS | Possessive ending |
| 18 | PRP | Personal pronoun |
| 19 | PRP$ | Possessive pronoun |
| 20 | RB | Adverb |
| 21 | RBR | Adverb, comparative |
| 22 | RBS | Adverb, superlative |
| 23 | RP | Particle |
| 24 | SYM | Symbol |
| 25 | TO | *to* |
| 26 | UH | Interjection |
| 27 | VB | Verb, base form |
| 28 | VBD | Verb, past tense |
| 29 | VBG | Verb, gerund or present participle |
| 30 | VBN | Verb, past participle |
| 31 | VBP | Verb, non-3rd person singular present |
| 32 | VBZ | Verb, 3rd person singular present |
| 33 | WDT | Wh-determiner |
| 34 | WP | Wh-pronoun |
| 35 | WP$ | Possessive wh-pronoun |
| 36 | WRB | Wh-adverb |



Name Entity Recognition



Basic Dependency – After identifying the POS Basic Dependency just expanding the relationship of each tag.

