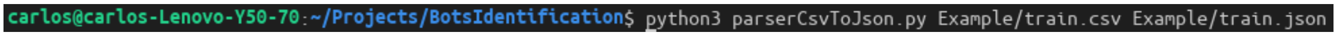
Carlos Augusto Amador Manilla A01329447

Homework 1

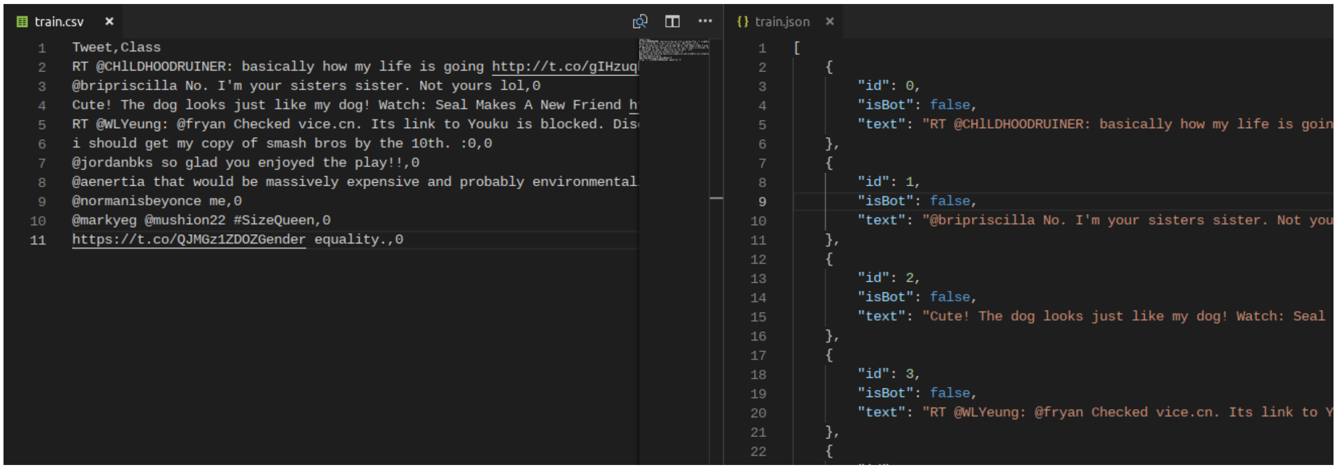
To make the file handling easier, every program receives two parameters via command line, the input and output file. In this way, it makes faster and easier to define which file is going to be used and where the output would be stored.

I decided to use python as it has many libraries that makes reading and writing in files, decoding, encoding and making requests easier.

The first step was to parse the data was to encode the data to a supported encoding, the parserCsvToJson.py program encoded the text of the tweets to latin-1 encoding, stored the class as a boolean 'isBot' and dumped the result into a file with json format. It also checked for errors and stored them in a different file. Three tweets were deleted from the data because they were empty strings.



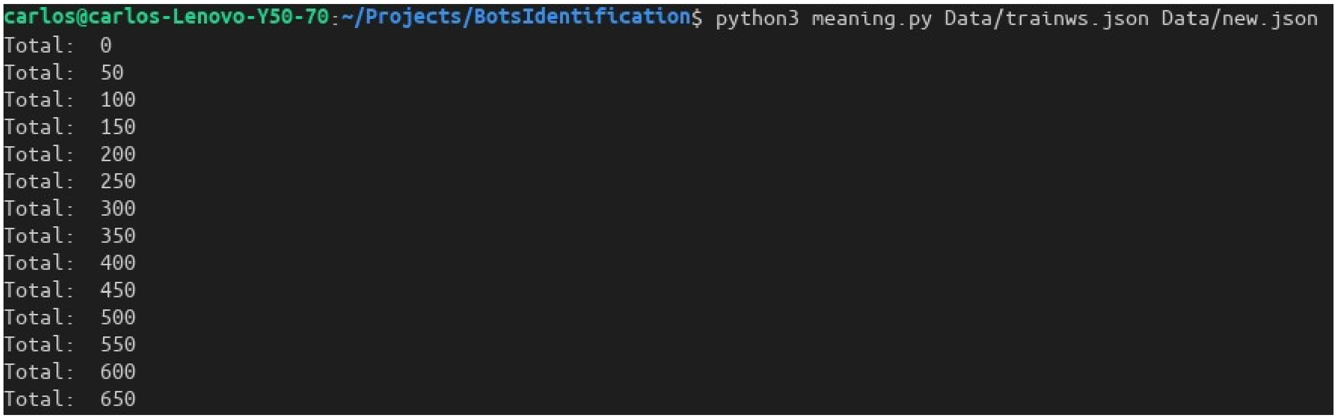
```
carlos@carlos-Lenovo-Y50-70:~/Projects/BotsIdentification$ python3 parserCsvToJson.py Example/train.csv Example/train.json
```



```
Tweet,Class
RT @CHlLDHOODRUINER: basically how my life is going http://t.co/gIHzuq
@bripriscilla No. I'm your sisters sister. Not yours lol,0
Cute! The dog looks just like my dog! Watch: Seal Makes A New Friend h
RT @WLYeung: @fryan Checked vice.cn. Its link to Youku is blocked. Dis
i should get my copy of smash bros by the 10th. :0,0
@jordanbks so glad you enjoyed the play!!,0
@aenertia that would be massively expensive and probably environmental
@normanisbeyonce me,0
@markyeg @mushion22 #SizeQueen,0
https://t.co/QJMGz1ZDOZGender equality.,0
```

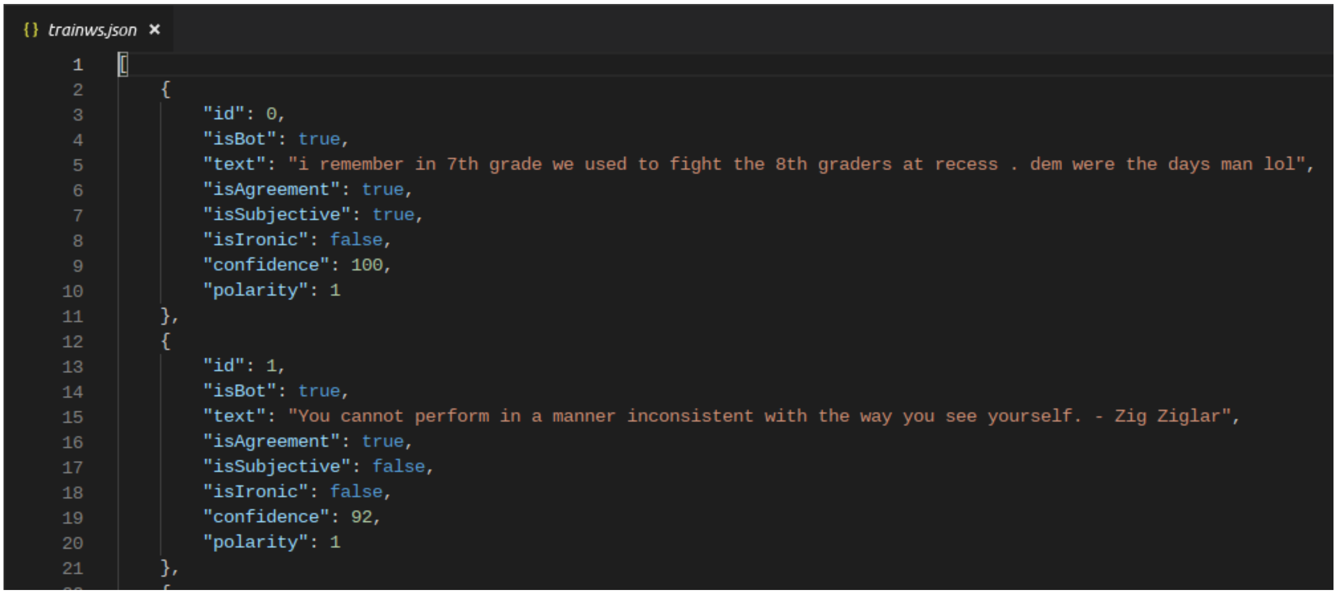
```
[
    {
        "id": 0,
        "isBot": false,
        "text": "RT @CHlLDHOODRUINER: basically how my life is goin
    },
    {
        "id": 1,
        "isBot": false,
        "text": "@bripriscilla No. I'm your sisters sister. Not you
    },
    {
        "id": 2,
        "isBot": false,
        "text": "Cute! The dog looks just like my dog! Watch: Seal
    },
    {
        "id": 3,
        "isBot": false,
        "text": "RT @WLYeung: @fryan Checked vice.cn. Its link to Y
    },
    {
```

Once the data was parsed to a json format and encoded to latin-1, the meaning.py program made a request for each tweet to the meaning cloud API. It waited .2 seconds between each request and stored the errors so I could check what was failing. The program printed the count of tweets processed each 50 tweets and write them on a file in order to record every advance.



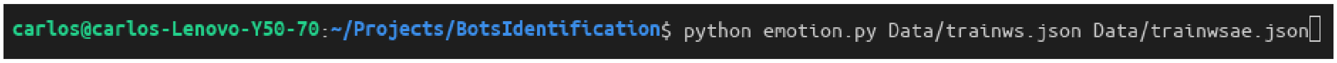
```
carlos@carlos-Lenovo-Y50-70:~/Projects/BotsIdentification$ python3 meaning.py Data/trainws.json Data/new.json
Total:  0
Total:  50
Total:  100
Total:  150
Total:  200
Total:  250
Total:  300
Total:  350
Total:  400
Total:  450
Total:  500
Total:  550
Total:  600
Total:  650
```

It also translated the Polarity value to an int value, as it is a qualitative ordinal value, the scale goes from -2 to 2 where the negative values represent negative polarity just as the positive values represent positive polarities and the 0 is neutral, confidence was stores as provided by the API. The other three values where stored as boolean values: isAgreement, isSubjective and isIronic.



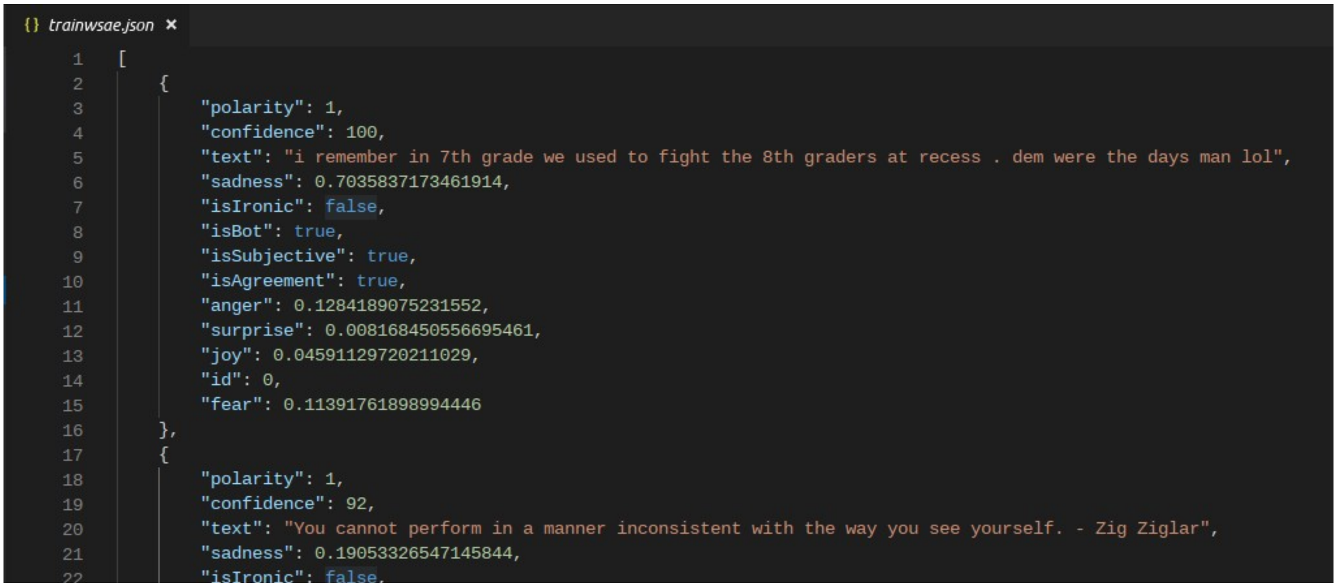
```json
[
    {
        "id": 0,
        "isBot": true,
        "text": "i remember in 7th grade we used to fight the 8th graders at recess . dem were the days man lol",
        "isAgreement": true,
        "isSubjective": true,
        "isIronic": false,
        "confidence": 100,
        "polarity": 1
    },
    {
        "id": 1,
        "isBot": true,
        "text": "You cannot perform in a manner inconsistent with the way you see yourself. - Zig Ziglar",
        "isAgreement": true,
        "isSubjective": false,
        "isIronic": false,
        "confidence": 92,
        "polarity": 1
    },
    {
```

The emotion.py program uses indico python library to request the 5 emotion values, it makes a call to the api for each tweet, also writes them in a file every 100 tweets and again at the end.



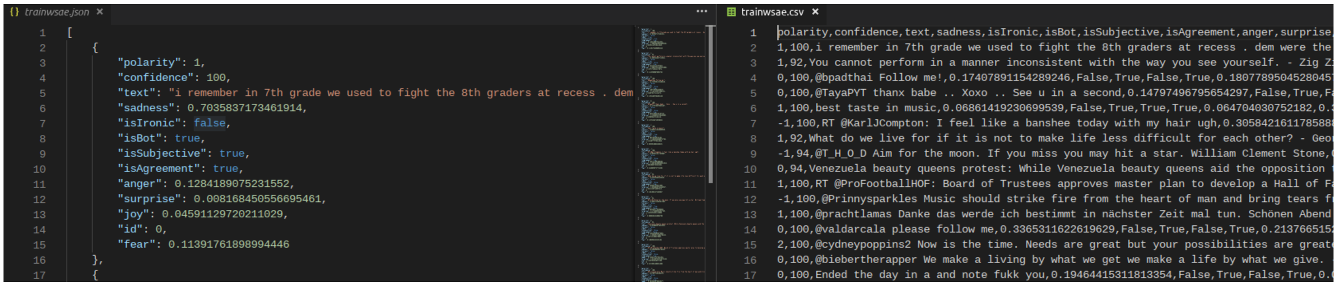
```
carlos@carlos-Lenovo-Y50-70:~/Projects/BotsIdentification$ python emotion.py Data/trainws.json Data/trainwsae.json
```

The values of the emotion API where stored as provided.



```json
[
    {
        "polarity": 1,
        "confidence": 100,
        "text": "i remember in 7th grade we used to fight the 8th graders at recess . dem were the days man lol",
        "sadness": 0.7035837173461914,
        "isIronic": false,
        "isBot": true,
        "isSubjective": true,
        "isAgreement": true,
        "anger": 0.1284189075231552,
        "surprise": 0.008168450556695461,
        "joy": 0.04591129720211029,
        "id": 0,
        "fear": 0.11391761898994446
    },
    {
        "polarity": 1,
        "confidence": 92,
        "text": "You cannot perform in a manner inconsistent with the way you see yourself. - Zig Ziglar",
        "sadness": 0.19053326547145844,
        "isIronic": false,
```

Finally, the parserJsonToCsv.py program re-formats the data from json to csv.



```
[
    {
        "polarity": 1,
        "confidence": 100,
        "text": "i remember in 7th grade we used to fight the 8th graders at recess . dem
        "sadness": 0.7035837173461914,
        "isIronic": false,
        "isBot": true,
        "isSubjective": true,
        "isAgreement": true,
        "anger": 0.1284189075231552,
        "surprise": 0.008168450556695461,
        "joy": 0.04591129720211029,
        "id": 0,
        "fear": 0.11391761898994446
    },
    {
```

```
polarity,confidence,text,sadness,isIronic,isBot,isSubjective,isAgreement,anger,surprise,
1,100,i remember in 7th grade we used to fight the 8th graders at recess . dem were the
1,92,You cannot perform in a manner inconsistent with the way you see yourself. - Zig Z
0,100,@bpadthai Follow me!,0.17407891154289246,False,True,False,True,0.180778950452804
0,100,@TayaPYT thanx babe .. Xoxo .. See u in a second,0.14797496795654297,False,True,F
1,100,best taste in music,0.06861419230699539,False,True,True,True,0.064704030752182,0.
-1,100,RT @KarlJCompton: I feel like a banshee today with my hair ugh,0.305842161178588
1,92,What do we live for if it is not to make life less difficult for each other? - Geo
-1,94,@T_H_O_D Aim for the moon. If you miss you may hit a star. William Clement Stone,
0,94,Venezuela beauty queens protest: While Venezuela beauty queens aid the opposition
1,100,RT @ProFootballHOF: Board of Trustees approves master plan to develop a Hall of F
-1,100,@Prinnysparkles Music should strike fire from the heart of man and bring tears f
1,100,@prachtlamas Danke das werde ich bestimmt in nächster Zeit mal tun. Schönen Abend
0,100,@valdarcala please follow me,0.3365311622619629,False,True,False,True,0.21376651
2,100,@cydneypoppins2 Now is the time. Needs are great but your possibilities are great
0,100,@biebertherapper We make a living by what we get we make a life by what we give.
0,100,Ended the day in a and note fukk you,0.19464415311813354,False,True,False,True,0.
```

The trainwsae.json and trainwsae.csv contain the tweets with the new 10 features.