Sentiment Analysis – SentiStrength

Για την ανάλυση συναισθήματος χρησιμοποιήθηκε το Lexicon based εργαλείο SentiStrength. Το SentiStrength είναι διαθέσιμο στην ιστοσελίδα http://sentistrength.wlv.ac.uk/.

Το εργαλείο μετράει το συναίσθημα σε απλό κείμενο με τιμές από -5 έως +5. Όπου το -1 έως -5 είναι Negative Sentiment και το 1 έως 5 είναι Positive Sentiment. Το 0 θεωρείται Neutral. Το εργαλείο SentiStrength μπορεί να επιστρέψει εώς 16.000 texts με Sentiment ανά δευτερόλεπτο.

Το Sentistrength αιτιολογεί την ύπαρξη δύο scores στο γεγονός ότι ο άνθρωπος μπορεί να νιώσει αρνητικά και θετικά συναισθήματα παράλληλα (mixed emotions).

Αρχικά υλοποιήθηκε για την εξαγωγή Sentiment για μικρές προτάσεις στα Αγγλικά. Με τα χρόνια και την περαιτέρω υλοποίηση μπόρεσε επίσης να καλύψει μια μεγάλη γκάμα από ξένες γλώσσες όπως Γερμανικά , ολλανδικά ,Ισπανικά , Ιταλικά κ.α. Απαραίτητη προϋπόθεση όμως για να γίνει αυτό είναι να γίνει αλλαγή στα Input files. Εκτός από την εισαγωγή Data sets υπάρχει και η δυνατότητα quick test με την εξαγωγή συναισθήματος από μια μόνο γραμμή κειμένου

Για την εξαγωγή Sentiment χρησιμοποιήθηκε το dataset της Kaggle όπου είναι διαθέσιμο στην ιστοσελίδα https://www.kaggle.com/c/twitter-sentiment-analysis2/data?fbclid=IwAR1IRcOrTXPxy1EqOJ4vHcUBPmb-XzWFQjIB5c4kL5EQJsgJLMFTrQ5-488 . Το παρόν dataset διαθέτει ένα αρχείο με 100κ tweets και θα εξαχθεί το Sentiment από αυτά (train.csv). Διαθέτει τρεις στήλες με ItemID , Sentiment και Sentiment text.

	ItemID	,Sentiment,SentimentText
1	ItemID	,Sentiment,SentimentText
2	1,0,	is so sad for my APL friend.............
3	2,0,	I missed the New Moon trailer...
4	3,1,	omg its already 7:30 :O
5	4,0,	.. Omgaga. Im sooo im gunna CRy. I've been at this dentist since 11.. I was suposed 2 ju
6	5,0,	i think mi bf is cheating on me!!! T_T
7	6,0,	or i just worry too much?
8	7,1,	Juuuuuuuuuuuuuuuuussssst Chillin!!
9	8,0,	Sunny Again Work Tomorrow :-| TV Tonight
10	9,1,	handed in my uniform today . i miss you already
11	10,1,	hmmmm.... i wonder how she my number @-)
12	11,0,	I must think about positive..
13	12,1,	thanks to all the haters up in my face all day! 112-102
14	13,0,	this weekend has sucked so far
15	14,0,	jb isnt showing in australia any more!
16	15,0,	ok thats it you win.
17	16,0,	<-------- This is the way i feel right now...
18	17,0,"	awhhe man.... I'm completely useless rt now. Funny, all I can do is twitter. http://myloc.me
19	18,1,	Feeling strangely fine. Now I'm gonna go listen to some Semisonic to celebrate
20	19,0,	HUGE roll of thunder just now...SO scary!!!!

Το SentiStrength για να εξάγει συναίσθημα αφού έγινε το απαραίτητο tokenization από το 2ο βήμα από retweets , απαλοιφή άλλων γλωσσών εκτός από αγγλικά αλλά και απαλοιφή html tags. Επιπρόσθετα χρειάζεται το Dataset να μετατραπεί σε .txt αρχείο. Δηλαδή της μορφής.

```
ItemID,Sentiment,SentimentText
1,0,                is so sad for my APL friend.............
2,0,                I missed the New Moon trailer...
3,1,            omg its already 7:30 :O
4,0,        .. Omgaga. Im sooo  im gunna CRy. I've been at this dentist since 11.. I was suposed 2 just get a crown put on (30mins)...
5,0,        i think mi bf is cheating on me!!!        T_T
6,0,        or i just worry too much?
7,1,        Juuuuuuuuuuuuuuuusssssst Chillin!!
8,0,        Sunny Again        Work Tomorrow  :-|        TV Tonight
9,1,        handed in my uniform today . i miss you already
10,1,       hmmm.... i wonder how she my number @-)
11,0,       I must think about positive..
12,1,       thanks to all the haters up in my face all day! 112-102
13,0,       this weekend has sucked so far
14,0,       jb isnt showing in australia any more!
15,0,       ok thats it you win.
16,0,       &lt;-------- This is the way i feel right now...
17,0,"      awhhe man.... I'm completely useless rt now. Funny, all I can do is twitter. http://myloc.me/27HX"
18,1,       Feeling strangely fine. Now I'm gonna go listen to some Semisonic to celebrate
19,0,       HUGE roll of thunder just now...SO scary!!!!
20,0,       I just cut my beard off. It's only been growing for well over a year. I'm gonna start it over. @shaunamanu is happy in the meantime.
21,0,       Very sad about Iran.
22,0,       wompppp wompp
23,1,       You're the only one who can see this cause no one else is following me this is for you because you're pretty awesome
24,0,       &lt;---Sad level is 3. I was writing a massive blog tweet on Myspace and my comp shut down. Now it's all lost *lays in fetal position*
25,0,       ... Headed to Hospitol : Had to pull out of the Golf Tourny in 3rd place!!!!!!!!!!! I Think I Re-Ripped something !!! Yeah THAT !!
26,0,       BoRinG  ): whats wrong with him??    Please tell me........  :-/
27,0,       cant be bothered. i wish i could spend the rest of my life just sat here and going to gigs. seriously.
28,0,"      Feeeling like shit right now. I really want to sleep, but nooo I have 3 hours of dancing and an art assignment to finish. "
29,1,"      goodbye exams, HELLO ALCOHOL TONIGHT "
30,0,       I didn't realize it was THAT deep. Geez give a girl a warning atleast!
31,0,       I hate it when any athlete appears to tear an ACL on live television.
32,0,       i miss you guys too      i think i'm wearing skinny jeans a cute sweater and heels   not really sure   what are you doing today
33,0,   -- Meet your Meat http://bit.ly/15SSCI
```

Το αρχείο τώρα είναι έτοιμο να γίνει import στο εργαλείο του SentiStrength. Για την εξαγωγή του Sentiment χρησιμοποιήθηκε η επιλογή "Analyse All Texts in File (each line separately)". Αυτό επιλέχθηκε γιατί το κάθε tweet στην μορφή .txt είναι ξεχωριστά μόνο του σε κάθε γραμμή.
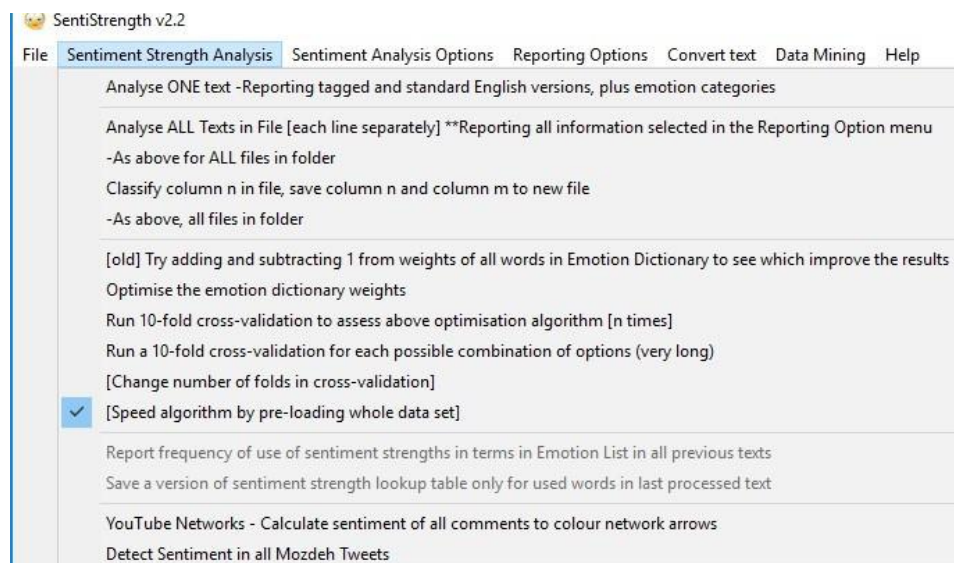
Το SentiStrength αποτελεί ένα πολύ αξιόπιστο εργαλείο καθώς προσφέρει μια πληθώρα επιλογών για την εξαγωγή Sentiment από κείμενο. Επίσης παρέχει την δυνατότητα για εξόρυξη πληροφορίας (Data Mining). Για την παρούσα εργασία προτιμήθηκε να κρατηθούν οι default επιλογές. Κάποιες από αυτές τις επιλογές περιλαμβάνουν τα επαναλαμβανόμενα γράμματα να αυξάνουν το Positive Sentiment , λέξεις όπως το "very" να αυξάνουν το Positive Sentiment ή το "some" να μειώνουν το Negative Sentiment. Επίσης προσφέρει την δυνατότητα να διορθώνει τυχόν επαναλαμβανόμενα λάθη στην ορθογραφία.

Το αρχείο αφού έγινε import και επεξεργάστηκε από το SentiStrength έξαγει τα αποτελέσματα με το ίδιο όνομα + την ονομασία results(train+results) και είναι της μορφής .txt.


```
ItemID,Sentiment,SentimentText  Translation       Positive        Negative        EmotionRationale
1,0,                    is so sad for my APL friend..............        1,0,               is so sad for my APL f
2,0,               I missed the New Moon trailer... 2,0,            I missed the New Moon trailer.  1
3,1,          omg its already 7:30 :O       3,1,          oh my god its already 7:30 :30       2       -2
4,0,          .. Omgaga. Im sooo  im gunna CRy. I've been at this dentist since 11.. I was suposed 2 just get a crown p
5,0,      i think mi bf is cheating on me!!!      T_T       5,0,          i think mi boyfriend is cheating on me!<br
6,0,       or i just worry too much?      6,0,       or i just worry too much?<br>         1       -3
7,1,      Juuuuuuuuuuuuuuuuusssssst Chillin!!   7,1,       Jusst Chillin!      2       -1      7[0] ,1[0] Juuuuuuuuuu
8,0,       Sunny Again       Work Tomorrow :-     TV Tonight       8,0,       Sunny Again       Work Tomorrow  :
9,1,      handed in my uniform today . i miss you already      9,1,      handed in my uniform today .<br> i miss you a
10,1,      hmmmm.... i wonder how she my number @-)      10,1,       hm.<br> i wonder how she my number @-      2
11,0,      I must think about positive..     11,0,       I must think about positive. 3     -1      11[0] ,0[0] I[
12,1,      thanks to all the haters up in my face all day! 112-102      12,1,      thanks to all the haters up in my fa
13,0,      this weekend has sucked so far      13,0,       this weekend has sucked so far      1       -2      13[0]
14,0,      jb isnt showing in australia any more!      14,0,      jb isnt showing in australia any more!      1
15,0,      ok thats it you win. 15,0,       ok thats it you win. 1       -1      15[0] ,0[0] ok[0] thats[0] it[0] you[0
16,0,"     &lt;-------- This is the way i feel right now...      16,0,      &lt; This is the way i feel right now. 1
17,0,"     awhhe man.... I'm completely useless rt now. Funny, all I can do is twitter. http://myloc.me/27HX"      17,0,
18,1,      Feeling strangely fine. Now I'm gonna go listen to some Semisonic to celebrate 18,1,      Feeling strangely fine
19,0,      HUGE roll of thunder just now...SO scary!!!!      19,0,      HUGE roll of thunder just now.<br>SO scary!      1
20,0,      I just cut my beard off. It's only been growing for well over a year. I'm gonna start it over. @shaunamanu is
21,0,      Very sad about Iran.  21,0,       Very sad about Iran.  1       -5      21[0] ,0[0] Very[0] sad[-3][-1 LastWor
22,0,      wompppp wompp 22,0,       womp womp      2       -1      22[0] ,0[0] wompppp/womp[0][+0.6 MultipleLetters] womp
23,1,      You're the only one who can see this cause no one else is following me this is for you because you're pretty a
24,0,      &lt;---Sad level is 3. I was writing a massive blog tweet on Myspace and my comp shut down. Now it's all lost *
25,0,      ... Headed to Hospitol : Had to pull out of the Golf Tourny in 3rd place!!!!!!!!!!!! I Think I Re-Ripped someth
26,0,      BoRinG   ): whats wrong with him??     Please tell me........ :-/      26,0,      BoRinG   BoRinG whats wrong wit
27,0,      can't be bothered. i wish i could spend the rest of my life just sat here and going to gigs. seriously. 27,0,
28,0,"     Feeeling like shit right now. I really want to sleep, but nooo I have 3 hours of dancing and an art assignment
29,1,"     goodbye exams, HELLO ALCOHOL TONIGHT " 29,1,     goodbye exams, HELLO ALCOHOL TONIGHT "  1      -1      29[0]
30,0,      I didn't realize it was THAT deep. Geez give a girl a warning atleast!  30,0,     I didn't realize it was THAT de
31,0,      I hate it when any athlete appears to tear an ACL on live television.      31,0,      I hate it when any athlete appe
32,0,      i miss you guys too      i think i'm wearing skinny jeans a cute sweater and heels    not really sure     what are
33,0,      -- Meet your Meat http://bit.ly/15SSCI  33,0,      - Meet your Meat http://bit.<br>ly/15SSCI       1       -1
```

Στην παρούσα μορφή εμφανίζεται μετά την εξαγωγή συναισθήματος και η οι στήλες Positive , Negative και EmotionRationale. Το αρχείο .txt πρέπει να μετατραπεί πάλι σε csv αρχείο για να μπορέσουμε να διακρίνουμε τα αποτελέσματα του καλύτερα στο Excel.

ItemID,Sentiment,SentimentText	Translation	Positive	Negative	EmotionRatio
1,0, is so sad for my APL friend.............	1,0, is so sad for my APL friend.	2	-4	1[0],0[0] is[0]
2,0, I missed the New Moon trailer...	2,0, I missed the New Moon trailer.	1	-2	2[0],0[0] I[0] m
3,1, omg its already 7:30 :O	3,1, oh my god its already 7:30 :30	2	-2	3[0],1[0] omg[
4,0, .. Omgaga. Im sooo im gunna CRy. I've been at this dentist since 11.. I was suposed 2 ju	4,0, . Omgaga. Im so im gunna CRy. I've been at this dentist since 11. I was	2	-4	4[0],0[0] [[Sen
5,0, i think mi bf is cheating on me!!! T_T	5,0, i think mi boyfriend is cheating on me! T_T	2	-3	5[0],0[0] i[0] t
6,0, or i just worry too much?	6,0, or i just worry too much? 	1	-3	6[0],0[0] or[0]
7,1, Juuuuuuuuuuuuuuuusssst Chillin!!	7,1, Jusst Chillin!	2	-1	7[0],1[0] Juuu
8,0, Sunny Again Work Tomorrow :- TV Tonight	8,0, Sunny Again Work Tomorrow : TV Tonight	2	-1	8[0],0[0] Sunn
9,1, handed in my uniform today . i miss you already	9,1, handed in my uniform today . i miss you already	2	-2	9[0],1[0] hand
10,1, hmmmm.... i wonder how she my number @-)	10,1, hm. i wonder how she my number @-	2	-1	10[0],1[0] hmr
11,0, I must think about positive..	11,0, I must think about positive.	3	-1	11[0],0[0] I[0]
12,1, thanks to all the haters up in my face all day! 112-102	12,1, thanks to all the haters up in my face all day! 112-102	2	-4	12[0],1[0] thai
13,0, this weekend has sucked so far	13,0, this weekend has sucked so far	1	-2	13[0],0[0] this
14,0, jb isnt showing in australia any more!	14,0, jb isnt showing in australia any more!	1	-1	14[0],0[0] jb[0
15,0, ok thats it you win.	15,0, ok thats it you win.	1	-1	15[0],0[0] ok[0
16,0, <------- This is the way i feel right now...	16,0, < This is the way i feel right now.	1	-1	16[0],0[0] &lt[
17,0," awhhe man.... I'm completely useless rt now. Funny, all I can do is twitter. http://myloc	17,0, awhhe man. I'm completely useless rt now. Funny, all I can do is twitter. htt	2	-4	17[0],0[0] awh
18,1, Feeling strangely fine. Now I'm gonna go listen to some Semisonic to celebrate	18,1, Feeling strangely fine. Now I'm gonna go listen to some Semisonic to celebrate	3	-1	18[0],1[0] Fee
19,0, HUGE roll of thunder just now...SO scary!!!!	19,0, HUGE roll of thunder just now. SO scary!	2	-5	19[0],0[0] HUG
20,0, I just cut my beard off. It's only been growing for well over a year. I'm gonna start it over.	20,0, I just cut my beard off. It's only been growing for well over a year. I'm gonna star	2	-1	20[0],0[0] I[0],
21,0, Very sad about Iran.	21,0, Very sad about Iran.	1	-5	21[0],0[0] Very
22,0, wompppp womp	22,0, womp womp	2	-1	22[0],0[0] wor
23,1, You're the only one who can see this cause no one else is following me this is for you bec	23,1, You're the only one who can see this cause no one else is following me this is for you bec	4	-1	23[0],1[0] You
24,0, <--Sad level is 3. I was writing a massive blog tweet on Myspace and my comp shut do	24,0, <Sad level is 3. I was writing a massive blog tweet on Myspace and my comp shut	1	-4	24[0],0[0] &lt[
25,0, ... Headed to Hospitol : Had to pull out of the Golf Tourny in 3rd place!!!!!!!!!!!! I Think I R	25,0, . Headed to Hospitol : Had to pull out of the Golf Tourny in 3rd place! I Think I Re	2	-1	25[0],0[0] [[Se
26,0, BoRinG): whats wrong with him?? Please tell me........ :-/	26,0, BoRinG BoRinG whats wrong with him? Please tell me. .	2	-3	26[0],0[0] BoR
27,0, can't be bothered. i wish i could spend the rest of my life just sat here and going to gigs. s	27,0, can't be bothered. i wish i could spend the rest of my life just sat here and going to g	2	-1	27[0],0[0] can
28,0," Feeeling like shit right now. I really want to sleep, but nooo I have 3 hours of dancing an	28,0, Feeeling like shit right now. I really want to sleep, but no I have 3 hours of dancing an	2	-1	28[0],0[0] Fee
29,1," goodbye exams, HELLO ALCOHOL TONIGHT "	29,1, goodbye exams, HELLO ALCOHOL TONIGHT "	1	-1	29[0],1[0] goo

Υπάρχει επίσης και η δυνατότητα οπτικοποίησης των δεδομένων και έτσι δημιουργήθηκε ένα γράφημα με την χρήση των columns Positive και Negative.

Λόγω του μεγάλου όγκου δεδομένων και της αδυναμίας του excel να οπτικοποιήσει όλα τα δεδομένα, επιλέχθηκαν δύο τυχαία μέρη από τα δεδομένα για να ερευνηθούν τα αποτελέσματα. Όπως είναι εμφανές από τα γραφήματα έπειτα από την επεξεργασία και την ανάλυση των δύο συγκεκριμένων κομματιών το Negative Sentiment υπερισχύει του Positive.

Sentiment