

Βαγγέλης Σαρμαζανίδης: AEM 2302
Ελευθέριος Κουλούμπρης: AEM 2304

Εδω περα παρουσιαζεται γενικα η λειτουργια του προγραμματος.Λεπτομερειες για καθε πακετο και κλαση μπορούν να βρεθουν στα αντιστοιχα αρχεια κωδικα.

Αρχικα γινεται η συλλογη των δεδομενων για την οποια ειναι υπευθυνο το πακετο DataCollection. Αυτο που γινεται ειναι το εξης:

Ο χρονος χωριζεται σε διαστηματα των 5 λεπτων. Κατα την διαρκεια ενος 5 λεπτου διαστηματος γινεται συλλογη των tweets σχετικων με τα ενεργα trends και αυτα αποθηκευονται προσωρινα σε buffers στην κυρια μνημη. Στην συνεχεια στην ληξη καθε διαστηματος συλλεγουμε τα νεα top trends και κανουμε update αναλογα την λιστα με τα ενεργα trends. Ταυτοχρονα παραδιδουμε το μεχρι τωρα συλλεγμενο batch απο tweets σε ενα δευτερο thread το οποιο ξεκινα την προεπεξεργασια τους και συνεχιζεται η συλλογη tweets λαμβανοντας υποψη τα ανανεωμενη λιστα με τα trends. Η προεξεργασια ειναι υλοποιημενη στο πακετο Preprocessing και λειτουργει πανω στο κειμενο των tweets αφαιροντας υπερσυνδεσμους, mentions, stopwords, σημεια στιξης και ταυτοχρονα γινεται το stemming. Αφου επεξεργαστει ολο το batch απο tweets που συλλεχθηκαν το τελευταιο πενταλεπτο. Στην συνεχεια, γινεται χρηση της λειτουργικότητας που παρέχεται απο το πακέτο MongoDB. Τα συλλεγμένα trends και tweets τοποθετούνται στις αντίστοιχες Mongo συλλογές. Για τα trends τοποθετεί εγγραφές με τα terms του trend και τα Dates που χρειάζονται για την ανάλυση (ποτε ήρθε ενα trend για πρώτη φορά, και τότε σταμάτησε να είναι ενεργό). Σε αυτό το στάδιο, οι εγγραφές για τα στην tweets περιέχουν το raw κείμενο όπως αυτο συλλέχθηκε και το κείμενο που παράχθηκε απο την προεπεξεργασία.

Τα παραπανω επαναλαμβάνονται καθε πενταλεπτο.

Τελος μετα την ολοκληρωση της συλλογη των δεδομενων αρχιζει το δευτερο σταδιο του προγραμματος η αναλυσησυναισθηματων.Υπευθυνο για αυτην την λειτουργια ειναι το πακετο SentimentAnalysis. Φορτωνονται τα tweets απο την βαση δεδομενων και γινεται η αναλυση στο προεπεξεργασμενο κειμενο του tweet. Και γινεται update η αντιστοιχη εγγραφη του tweet στην βαση προσθετοντας την βαθμολογια για καθε ενα απο τα 6 συναισθηματα.

Σημείωση: Στο πακέτο StatisticsGathering υλοποιήθηκαν καποιες από τις απαραίτητες συναρτήσεις για την ανάλυση των tweets με βάση το Trend στο οποίο αντιστοιχούν, και την ομαδοποίηση των tweets σε διαστήματα χρόνου.Αλλά,η λειτουργικότητα αυτών των συναρτήσεων δεν έχει ελεγχθεί, ούτε τις έχουμε εκτελέσει για να δημιουργήσουμε τα δεδομένα που χρειάστηκαν για την παραγωγή των γραφημάτων, οποτε δεν υλοποιήθηκε και κάποιο website.