

## ΦΑΣΗ Ι. Άσκησης

Συνδεθείτε με το Twitter μέσω του Twitter Streaming/REST API

-<https://dev.twitter.com/streaming/overview>

-<https://dev.twitter.com/rest/public>

και συγκεντρώστε για μία εβδομάδα:

(α) τα tweets που περιέχουν το hashtag #ND ή #SYRIZA ή ισοδύναμα π.χ. #NewDemocracy ή #Syrizanel

Προσοχή θα συγκεντρώσετε τα tweets που περιέχουν ένα μόνο από τα δύο hashtags.

(b) Όμοια συγκεντρώστε τα tweets που περιέχουν @username το όνομα ενός πολιτικού αρχηγού.

(γ) Όμοια φτιάξτε δύο κατηγορίες από τα παραπάνω tweets, μια με αυτά που έχουν θετικό emoticon και μια άλλη με αυτά που έχουν αρνητικό emoticon.

- Με τη χρήση της βιβλιοθήκης Python, Tweepy ή Java Twitter4J, συνδεθείτε με το Twitter API
- Το Twitter παρέχει ένα REST API για πρόσβαση στα δεδομένα του, με περιορισμό 180 ερωτήματα ανά διάστημα 15 λεπτών. Για τη πρόσβαση στο API ο χρήστης της εφαρμογής πρέπει να αυθεντικοποιηθεί. Αυτό γίνεται είτε μέσω OAuth2.0 είτε με τη δήλωση της εφαρμογής στο Twitter και χρήση των κλειδιών που παρέχει το ίδιο το twitter. (<https://apps.twitter.com/>)  
Για την λήψη των tweets ενός χρήστη μπορεί να χρησιμοποιηθεί η μέθοδος statuses/user\_timeline του API η οποία δίνει πρόσβαση στο ιστορικό (timeline) του χρήστη. Σε κάθε ερώτημα επιστρέφει 100 tweets με όριο τα 3200 tweets ανά χρήστη.  
Για κάθε tweet για το οποίο γίνεται αναζήτηση, το API, απαντάει με ένα πλήθος δεδομένων όπως:
  - *id*
  - *Ημερομηνία δημιουργίας*
  - *Κείμενο*
    - *Χρήστης*
    - *id*
    - *twitter name*
    - *Πραγματικό Όνομα*
    - *Φωτογραφία profile*
    - *Χρώματα σελίδας*
    - ...
  - *Συντεταγμένες*
  - *Δηλωμένο ως Favorite*
  - *URL*
    - *Short*
    - *Expanded*
    - *Display*
  - *Στοιχεία απάντησης*
  - *Αναφορές σε άλλα άτομα*
  - ... ..Για ευκολότερη χρήση του API αλλά και του Parsing των δεδομένων από JSON σε πραγματικά αντικείμενα, χρησιμοποιείτε κάποια από τις βιβλιοθήκες Tweepy ή Twitter4J.
- Εναλλακτικά μπορείτε να συγκεντρώσετε τα tweets σε πραγματικό χρόνο μέσω του streaming API το οποίο δίνει παρόμοια λειτουργία με ελαστικότερους περιορισμούς.

- Χρησιμοποιήστε ένα χώρο αποθήκευσης/αναζήτησης για τα tweets που συγκεντρώσατε, π.χ. Elasticsearch, SOLR, MySQL, PostgreSQL.

Το πρόγραμμα στο τέλος θα τυπώνει το πλήθος των tweets που συγκεντρώθηκαν σε καθεμιά κατηγορία.

**Παραδοτέο**: Αναφορά (report) της μεθοδολογίας και διαδικασιών που ακολουθήσατε για την συγκέντρωση των δεδομένων και τα αποτελέσματα για δύο συνεχόμενες βδομάδες.

**Ημερομηνία Παράδοσης** : 1 Δεκεμβρίου 2016

**Μονάδες**: 15%