



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

## **Αναζήτηση και Ανάλυση Influential Χρηστών Twitter**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

**ΚΑΡΑΚΟΛΗ ΕΥΑΓΓΕΛΟΥ**

**Επιβλέπων :** Ιωάννης Βασιλείου  
Καθηγητής Ε.Μ.Π.

Αθήνα, Μάρτιος 2015





ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ  
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ  
ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

## Αναζήτηση και Ανάλυση Influential Χρηστών Twitter

### ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

**ΚΑΡΑΚΟΛΗ ΕΥΑΓΓΕΛΟΥ**

**Επιβλέπων :** Ιωάννης Βασιλείου  
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 28<sup>η</sup> Μαρτίου 2015.

(Υπογραφή)

.....  
Ιωάννης Βασιλείου  
Καθηγητής Ε.Μ.Π.

(Υπογραφή)

.....  
Ιωάννης Σταύρακας  
Καθηγητής Ε.Μ.Π.

(Υπογραφή)

.....  
Κώστας Κοντογιάννης  
Καθηγητής Ε.Μ.Π.

Αθήνα, Μάρτιος 2015

(Υπογραφή)

.....

**ΚΑΡΑΚΟΛΗΣ ΕΥΑΓΓΕΛΟΣ**

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © 2015 Καρακόλης Ευάγγελος, 2015. Με επιφύλαξη παντός δικαιώματος – All rights reserved

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα. Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

## Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή Γιάννη Βασιλείου όσο και τον ερευνητή Γιάννη Σταύρακα, για την ευκαιρία που μου έδωσαν να ασχοληθώ με τη συγκεκριμένη διπλωματική εργασία, αλλά και για τη συνεννοησιμότητα τους, όπως επίσης και για την υποστήριξή τους, με τις γνώσεις τους, για την εκπόνηση αυτής της διπλωματικής.

Ακόμη ευχαριστώ τους φίλους μου και τη σύντροφο μου για τη διαρκή στήριξη, αλλά και για τις όμορφες στιγμές που έχουμε περάσει μαζί.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένεια μου για την αμέριστη συμπαράσταση που μου δείχνουν όλα αυτά τα χρόνια και την στήριξη σε όλες μου τις επιλογές.

## Περίληψη

Τα διαδικτυακά κοινωνικά δίκτυα (**OSNs**) έχουν μπει για τα καλά στην καθημερινή μας ζωή, επηρεάζοντας σε μεγάλο βαθμό τόσο την ζωή όσο και την σκέψη των ανθρώπων. Χρησιμοποιούνται για διασκέδαση, ενημέρωση, αλλά και για διαφήμιση, marketing και κάθε λογής επιρροή.

Την επιρροή αυτή, και συγκεκριμένα αυτούς που την ασκούν στο Twitter, επιχειρεί να αναλύσει αυτή η διπλωματική. Ο σκοπός είναι η εύρεση Twitterers που επηρεάζουν την κοινή γνώμη πάνω σε κάποιο συγκεκριμένο θέμα και στη συνέχεια η ταξινόμηση αυτών με βάση την επιρροή (influence) που ασκούν. Για την ταξινόμηση χρησιμοποιούμε τέσσερις διαφορετικούς τρόπους. Πρόκειται για τους indegree (αριθμός followers ενός Twitterer), retweets, mentions (ο αριθμός αναφορών “@username” στα tweets άλλων twitterers με το username του μελετώμενου Twitterer) και TwitterRank (βασίζεται στην ιδέα ότι όσο πιο influential είναι οι followers ενός χρήστη τόσο μεγαλύτερη είναι και η επιρροή που ασκεί). Αξίζει να σημειωθεί ότι οι συγκεκριμένοι τρόποι μέτρησης αντιπροσωπεύουν διαφορετικά είδη επιρροής. Για παράδειγμα, ο αριθμός των followers δείχνει το μέγεθος του κοινού ενός χρήστη, αλλά αυτό δεν σημαίνει ότι έχει τη δυνατότητα να αλληλεπιδρά με το κοινό του, όπως κάποιος που έχει μεγαλύτερο αριθμό replies.

Επιπροσθέτως, στα πλαίσια της εργασίας υλοποιήθηκε μία web εφαρμογή για την οπτικοποίηση των αποτελεσμάτων των αλγορίθμων, καθώς επίσης και για την πλήρη παρουσίαση των τελικών influential χρηστών κατά την περίοδο που έγινε η ανάλυση, αξιοποιώντας και οπτικοποιώντας πληροφορίες που το twitter παρέχει με τρόπο όχι φιλικό για το χρήστη (π.χ. γεωγραφικά δεδομένα). Τέλος, η εφαρμογή δίνει τη δυνατότητα στον τελικό χρήστη να προσπελάσει τους influentials και με ενιαίο τρόπο, ανεξαρτήτως θέματος, με την ταξινόμηση βέβαια να υλοποιείται, αποκλειστικά, με βάση τους followers που έχει ο κάθε influential.

**Λέξεις Κλειδιά:** Twitter, Campaign, ανάλυση, κοινωνικά δίκτυα, επιρροή, Influential, follower, friend, reply, retweet, twitterer



## Abstract

The online social networks (OSNs) have a strong impact on everyday life influencing significantly both actions and thoughts of people. They are used for entertainment, information and also for advertising, marketing and generally persuasion.

The influence aspect and especially the users of twitter who exert this at most (influentials) is the subject that we try to analyze with this thesis. The purpose is to find users of twitter (twitterers) that influence the public opinion on a certain topic and evaluate them depending on the sum of influence they exert. For the evaluation we use four different measures. These are indegree (the number of followers of a twitterer), retweets, mentions (the number of tweets of other twitterers that contain the keyword @twitterer's\_username except retweets) and TwitterRank (an algorithm based on the idea that if a twitterer has influential followers he holds a big amount of influence via his followers). It is worth mentioning that these measures represent different types of influence. For example, indegree (the number of followers of a twitterer) shows the size of the audience a twitterer has; this does not imply that the twitterer is influential in other ways, such as interacting with this audience.

Finally, we developed a web application for the visualization of the results for each measure, and the presentation of influentials, during the campaign, visualizing information that twitter provides not in a user friendly way (such as geoLocation). Furthermore, the application gives the opportunity to access the influentials regardless of the topic ordering them using just the number of followers.

**Keywords:** twitter, campaign, OSNs, influence, Influential, follower, friend, reply, retweet, twitterer





## Πίνακας περιεχομένων

<b>1</b>	<b>Εισαγωγή.....</b>	<b>1</b>
1.1	Social influence στο twitter .....	1
1.2	Αντικείμενο διπλωματικής .....	3
1.2.1	Συνεισφορά .....	3
1.3	Οργάνωση κειμένου .....	4
<b>2</b>	<b>Σχετικές εργασίες.....</b>	<b>5</b>
<b>3</b>	<b>Θεωρητικό υπόβαθρο .....</b>	<b>7</b>
3.1	Influentials και μη Influentials.....	7
3.2	Topic sensitivity στο Twitter .....	8
3.3	Homophily και Reciprocity στο Twitter .....	9
3.4	Μετρικές influence στο Twitter .....	10
3.4.1	Followers στο Twitter (indegree influence).....	10
3.4.2	Retweets στο Twitter (retweet influence) .....	10
3.4.3	Mentions στο Twitter (mention influence) .....	10
3.4.4	Topic-specific Twitter Rank .....	11
3.5	Twitter APIs .....	13
<b>4</b>	<b>Ανάλυση Συστήματος.....</b>	<b>14</b>
4.1	Συλλογή Δεδομένων .....	14
4.2	Υπολογίζοντας το influence .....	15
4.2.1	<i>Indegree, Retweets, Mentions</i> .....	15
4.2.2	<i>Υπολογισμός TwitterRank</i> .....	15
4.3	Αρχιτεκτονική.....	17
4.4	Περιγραφή Λειτουργιών .....	19
4.4.1	<i>Έναρξη νέου campaign</i> .....	19
4.4.2	Επισκόπηση των campaigns που έχουν ολοκληρωθεί.....	20
4.4.3	Αναζήτηση campaign .....	21
4.4.4	Διαγραμματική απεικόνιση αποτελεσμάτων campaign.....	22
4.4.5	Παρουσίαση λίστας Influentials βάσει αποτελεσμάτων campaign.....	23
4.4.6	Απεικόνιση καρτέλας influential χρήστη στο campaign .....	24

4.4.6.1	Παρουσίαση πληροφοριών και επίδοσης στο campaign .....	24
4.4.6.2	Παρουσίαση σημαντικότερων tweets κατά το campaign .....	25
4.4.6.3	Θεματική πλοήγηση στα tweets κατά το campaign .....	26
4.4.6.4	Αναπαράσταση tweets στο χάρτη κατά το campaign .....	27
4.4.6.5	Απεικόνιση tweets στη μονάδα του χρόνου κατά το campaign .....	29
4.4.7	Παρουσίαση λίστας influentials από όλα τα campaigns.....	30
4.4.8	Απεικόνιση καρτέλας influential χρήστη.....	30
4.4.6.1	Παρουσίαση γενικών χαρακτηριστικών .....	30
4.4.6.2	Παρουσίαση σημαντικότερων tweets στα campaigns που συμμετείχε.....	31
4.4.6.3	Θεματική πλοήγηση στα tweets .....	31
4.4.6.4	Αναπαράσταση tweets στο χάρτη .....	31
4.5	Μοντέλο Οντοτήτων Συσχετίσεων .....	32
<b>5</b>	<b>Σχεδίαση Συστήματος .....</b>	<b>36</b>
5.1	Αρχιτεκτονική.....	36
5.2	Περιγραφή Βασικών Κλάσεων.....	36
5.2.1	<i>NewCampaignManager</i> .....	38
5.2.2	<i>DbUpdateManager</i> .....	38
5.2.3	<i>Search</i> .....	38
5.2.4	<i>UserFilter</i> .....	38
5.2.5	<i>UserTimeline</i> .....	39
5.2.6	<i>MentionCollector</i> .....	39
5.2.7	<i>FriendsCollector</i> .....	40
5.2.8	<i>WordAnalysis</i> .....	40
5.2.9	<i>TwitterRank</i> .....	41
5.2.10	<i>Similarity</i> .....	41
5.2.11	<i>Probabilities</i> .....	41
5.2.12	<i>Teleportation</i> .....	41
5.3	Βάση Δεδομένων .....	42
<b>6</b>	<b>Υλοποίηση web-based εφαρμογής.....</b>	<b>44</b>
6.1	Περιγραφή βασικών JSP, servlets και κλάσεων web εφαρμογής .....	44
6.1.1	<i>Home.jsp</i> .....	45
6.1.2	<i>campaign.jsp</i> .....	45
6.1.3	<i>userList.jsp</i> .....	45
6.1.4	<i>user.jsp</i> .....	46

6.1.5	<i>userThem.jsp</i> .....	46
6.1.6	<i>userMap.jsp</i> .....	46
6.1.7	<i>userTime.jsp</i> .....	47
6.1.8	<i>topList.jsp</i> .....	47
6.1.9	<i>top.jsp</i> .....	47
6.1.10	<i>topThem.jsp</i> .....	47
6.1.11	<i>topMap.jsp</i> .....	48
6.1.12	<i>campaignStarted.jsp</i> .....	48
6.1.13	<i>ServletAdmin.java</i> .....	48
6.1.14	<i>MyThread.java</i> .....	48
6.2	Πλατφόρμες και προγραμματιστικά εργαλεία .....	49
6.2.1	<i>HTML, CSS, JavaScript</i> .....	49
6.2.2	<i>JQuery.js</i> .....	50
6.2.3	<i>flot.js, d3.js, Leaflet.js</i> .....	50
6.2.4	<i>JAVA Enterprise Edition</i> .....	50
6.2.5	<i>Servlet, JSP</i> .....	51
6.2.6	<i>Apache Tomcat</i> .....	51
6.2.7	<i>MySql</i> .....	51
<b>7</b>	<b>Παράδειγμα χρήσης web-based εφαρμογής .....</b>	<b>52</b>
<b>8</b>	<b>Επίλογος .....</b>	<b>64</b>
8.1	Σύνοψη και συμπεράσματα .....	64
8.2	Μελλοντικές επεκτάσεις.....	65
<b>9</b>	<b>Βιβλιογραφία .....</b>	<b>67</b>

# 1

## *Εισαγωγή*

### *1.1 Social influence στο Twitter*

Η κοινωνική επιρροή (social influence) ασκείται όταν τα συναισθήματα, οι απόψεις και η συμπεριφορά κάποιου επηρεάζονται από κάποιο άλλο άτομο (influencer). Παραδοσιακά, το ρόλο του influencer αναλάμβαναν τα μέσα μαζικής ενημέρωσης, οι πολιτικοί, και άλλα δημόσια πρόσωπα, στοχεύοντας πολλές φορές στην προώθηση προϊόντων, πολιτικών προσώπων ([3]) κλπ.. Ωστόσο, τα τελευταία χρόνια, με την ανάπτυξη των επικοινωνιών και κυρίως με την εξάπλωση του internet και των κοινωνικών δικτύων, η κατάσταση έχει διαφοροποιηθεί σημαντικά. Πλέον, οι άνθρωποι κατακλύζονται από πληροφορίες από πολλές διαφορετικές πηγές, και μέσω του διαδικτύου, αρκετοί καταφέρνουν να ασκούν επιρροή, χωρίς απαραίτητως να είναι δημόσια πρόσωπα. Εύκολα συμπεραίνει κανείς, λοιπόν, το πόσο σημαντικό ρόλο παίζει στην καθημερινή ζωή η κοινωνική επιρροή.

Ωστόσο, παρότι έχει μελετηθεί εκτενώς σε αρκετά επιστημονικά πεδία, όπως είναι η κοινωνιολογία, οι επικοινωνίες, το marketing και η πολιτική επιστήμη, είναι πολύ δύσκολο να την αναλύσεις. Αυτό, διότι δεν υπάρχει κάποιος κοινά αποδεκτός τρόπος για να ποσοτικοποιηθεί, αλλά και λόγω του γεγονότος ότι πρωταρχικής σημασίας έννοιες, όπως οι ανθρώπινες επιλογές, ή η λειτουργία μιας κοινωνίας δεν μπορούν να μετρηθούν. Επιπλέον, στον προφορικό λόγο και στην καθημερινή ζωή είναι δύσκολο να πούμε ότι κάποιος

επηρέαστηκε από κάποιον άλλον, διότι θα μπορούσε ένα μήνυμα να ανήκει σε περισσότερα από ένα άτομα και δεν μπορούμε να ξέρουμε ποιος επηρέασε ποιον καθώς δεν γνωρίζουμε τις σχέσεις μεταξύ των ατόμων, ούτε πότε είπε κάποιος κάτι, ή αν δύο άτομα είχαν κάποια συζήτηση.

Το κοινωνικό δίκτυο (OSN) twitter μας δίνει τη δυνατότητα να γνωρίζουμε αν δυο χρήστες του έχουν κάποια σχέση καθώς επίσης και ποια μηνύματα λαμβάνει ένας χρήστης και από ποιον. Επίσης μας επιτρέπει να παρακολουθούμε μια συζήτηση μεταξύ δύο ή περισσότερων χρηστών κι επιπλέον αν ένα μήνυμα αναπαράχθηκε κι από ποιον.

Πιο συγκεκριμένα, το twitter επιτρέπει στους χρήστες του να αναρτούν και να διαβάζουν σύντομα μηνύματα, έως 140 χαρακτήρων, τα οποία ονομάζονται tweets. Ένα tweet μπορεί να αποτελεί απάντηση σε κάποιον χρήστη εάν συνοδεύεται από το σύμβολο @ ακολουθούμενο από το username του χρήστη στον οποίο απευθύνεται η απάντηση. Επιπλέον, μπορεί να αποτελεί αναπαραγωγή κάποιου μηνύματος άλλου χρήστη εάν συνοδεύεται στην αρχή του από τα σύμβολα RT @username ή στο τέλος από το via @username, όπου username είναι το username του χρήστη προέλευσης του tweet. Όσον αφορά τις σχέσεις, υπάρχει μόνο ένα είδος σχέσης που λέγεται follow, είναι κατευθυνόμενο, μονομερές και συνδέει δυο χρήστες. Όταν ο χρήστης A κάνει follow τον χρήστη B τότε ο A χαρακτηρίζεται follower του B ενώ ο B friend του A. Όταν ένας χρήστης δημοσιεύει ένα tweet αυτό δημοσιεύεται αυτόματα τόσο στην αρχική του σελίδα όσο και στην αρχική σελίδα των followers του.

Επομένως, το twitter αποτελεί μια εικόνα της κοινωνίας όπου υπάρχουν ξεκάθαρες σχέσεις μεταξύ των χρηστών, σαφής εικόνα για την διάδοση της πληροφορίας (information diffusion), καθώς επίσης και αυτούσιες οι συζητήσεις μεταξύ των χρηστών. Όλα αυτά σε συνδυασμό με το πολύ ολοκληρωμένο API του, αλλά και το γεγονός ότι όλα τα δεδομένα είναι δημόσια, καθιστούν το twitter ένα εξαιρετικό μέσο για την μελέτη της κοινωνικής επιρροής αλλά και την ανίχνευση, τη παρακολούθηση και την ανάλυση πολλών influential χρηστών.

## **1.2 Αντικείμενο διπλωματικής**

Στην διπλωματική αυτή προσπαθούμε να ποσοτικοποιήσουμε το social influence στο twitter και στη συνέχεια να εξάγουμε τους influential twitterers και να βγάλουμε μια τελική κατάταξη για αυτούς, χρησιμοποιώντας τέσσερις διαφορετικούς τρόπους μέτρησης. Η διαδικασία που περιγράφηκε είναι topic- sensitive, διότι πιστεύουμε ότι το influence των twitterers διαφέρει ανάλογα με το topic. Αυτό επιτυγχάνεται με την υλοποίηση μιας εφαρμογής, η οποία, αφ' ενός τρέχει topic- focused campaigns και από την άλλη παρουσιάζει τα αποτελέσματα αλλά και ένα γενικό προφίλ των influential twitterers μέσα στο campaign.

### **1.2.1 Συνεισφορά**

Στη διπλωματική εργασία σχεδιάστηκε και υλοποιήθηκε σύστημα που:

- Χρησιμοποιεί το API του twitter για τη συλλογή μεγάλου αριθμού tweets που έχουν σχέση με ένα συγκεκριμένο topic.
- Εξάγει, από τα tweets, τους twitterers στους οποίους ανήκουν.
- Φιλτράρει τους twitterers και κράταει μόνο έναν αριθμό από αυτούς (τελικοί influentials).
- Μαζεύει τα tweets (μέσω twitter API) που έχουν κάνει οι τελικοί influentials για μια συγκεκριμένη χρονική περίοδο.
- Μαζεύει τα tweets (μέσω twitter API) που αποτελούν απάντηση ή retweets στους influentials για μια συγκεκριμένη χρονική περίοδο.
- Εξάγει (μέσω twitter API) τις σχέσεις μεταξύ των influentials.
- Υπολογίζει την επίδοση των influentials, για κάθε έναν από τους τρόπους μέτρησης indegree, retweets, mentions και TwitterRank.
- Καταχωρεί τα αποτελέσματα σε μία σχεσιακή βάση δεδομένων.

Επιπλέον, για την παρουσίαση των αποτελεσμάτων σχεδιάστηκε και αναπτύχθηκε μια web εφαρμογή που επιτρέπει τις εξής λειτουργίες:

- Επισκόπηση και αναζήτηση για όλα τα campaigns που έχουν τρέξει.

- Έναρξη νέου topic- focused campaign.
- Διαγραμματική απεικόνιση του top 10 των influentials για όλες τις μετρικές που χρησιμοποιήθηκαν.
- Απεικόνιση σε λίστα όλων των influentials ενός campaign και δυνατότητα ταξινόμησης αποτελεσμάτων βάσει των τρόπων υπολογισμού του influence.
- Επισκόπηση των twitterers ενός campaign και της επίδοσης τους σ' αυτό, όπως επίσης και βασικών πληροφοριών για αυτούς μέσα από τη δράση τους στο twitter.
- Ανάγνωση των πιο αγαπημένων (most favorited) tweets ενός twitterer.
- Απεικόνιση των tweets ενός twitterer στο χάρτη και στο χρόνο.
- Θεματική πλοήγηση στα tweets ενός twitterer μέσω word cloud.

### ***1.3 Οργάνωση κειμένου***

Εργασίες σχετικές με το αντικείμενο της διπλωματικής, καθώς επίσης και τα βασικότερα συμπεράσματα τους παρουσιάζονται στο Κεφάλαιο 2. Το Κεφάλαιο 3 ασχολείται με το θεωρητικό υπόβαθρο που πρέπει να κατέχει κάποιος ώστε να κατανοήσει αφ' ενός το influence και τους αλγορίθμους υπολογισμού του και αφ' ετέρου το Twitter API το βασικό εργαλείο για την εξαγωγή δεδομένων από το Twitter. Στο Κεφάλαιο 4 αναλύουμε το σύστημα μας, τόσο ως προς τις λειτουργίες που θα πρέπει να παρέχει, όσο και ως προς το πώς υπολογίζουμε το influence στην εφαρμογή μας. Στο κεφάλαιο 5, παρουσιάζεται η σχεδίαση του υποσυστήματος για τον υπολογισμό νέου campaign, οι βασικές του κλάσεις και η βάση δεδομένων που χρησιμοποιήθηκε. Στο κεφάλαιο 6 παρουσιάζεται η υλοποίηση της web-εφαρμογής με τις βασικότερες jsp σελίδες, τα servlet και τις βασικές κλάσεις, και επίσης τα προγραμματιστικά εργαλεία που χρησιμοποιήθηκαν. Στο κεφάλαιο 7, παρέχεται ένα παράδειγμα χρήσης της εφαρμογής, ενώ στο κεφάλαιο 8 παρέχονται τα συμπεράσματα και κάποιες πιθανές μελλοντικές επεκτάσεις στην εφαρμογή. Τέλος, στο κεφάλαιο 9 παραθέτουμε τη βιβλιογραφία που χρησιμοποιήθηκε.



# 2

## *Σχετικές εργασίες*

Γενικά, έχουν εκπονηθεί αρκετές εργασίες για τη χρήση του Twitter αλλά και για την εξάπλωση του. Ωστόσο, θα επικεντρώσουμε σε εργασίες σχετικές με το influence. Πρωτίστως, η εργασία των Kwak, Lee, Park και Moon ταξινομεί τους χρήστες ως influentials με βάση τον αριθμό των followers και με βάση τον αλγόριθμο PageRank και βρίσκει ότι τα αποτελέσματα είναι παρόμοια. Ακόμη, η εργασία των Weng, Lim, Jiang και He [2] μελέτησε ένα σύνολο 1000 χρηστών από τη Σιγκαπούρη, μαζί με όλους τους Followers αυτών. Συμπέραναν ότι στη σχέση follow μεταξύ δυο χρηστών υπάρχει μεγάλη αμοιβαιότητα. Συγκεκριμένα, παρατήρησαν ότι 72.4% των χρηστών ακολουθεί παραπάνω από 80% από τους followers τους κι επίσης ότι το 80.5% των χρηστών έχει followers το 80% αυτών που ακολουθούν. Επιπλέον, ανέπτυξαν έναν αλγόριθμο που μοιάζει με τον Topic sensitive PageRank για τη μέτρηση του influence, τον TwitterRank, τον οποίο και έχουμε υλοποιήσει. Η βασική ιδέα είναι ότι σε ένα συγκεκριμένο Topic το influence ενός χρήστη είναι το άθροισμα του influence των followers που έχει. Επίσης η εργασία των Cha, Haddadi, Benevenuto, και Gummadi [2] χρησιμοποίησε τρεις μετρικές για το social influence (indegree, retweets, mentions), τις οποίες χρησιμοποιούμε κι εμείς στη παρούσα διπλωματική, και είναι παρόμοια με την εργασία των Ye και Wu [4]. Οι πρώτοι συμπεραίνουν ότι οι χρήστες με μεγάλο αριθμό followers δεν έχουν απαραίτητα και μεγάλο αριθμό mentions ή retweets, ακόμη ότι οι χρήστες μπορεί να είναι influential σε πολλά topics ταυτόχρονα, και τέλος ότι περιορίζοντας τα tweets του ο χρήστης σε ένα συγκεκριμένο topic μπορεί να αυξήσει το influence του σε αυτό. Οι δεύτεροι, συμπέραναν ότι σε σημαντικά

γεγονότα οι απαντήσεις εξαπλώνονται πολύ γρήγορα και σε αρκετούς χρήστες πέρα από τους followers του αρχικού twitterer. Τέλος υπάρχει η εργασία των Bankshy, Hoffmann, Mason, Watts [2], η οποία μελετά την μετάδοση tweets και συμπεραίνει ότι οι χρήστες με τους περισσότερους followers, οι οποίοι ήταν influential στο παρελθόν έχουν περισσότερες πιθανότητες να μεταδώσουν ένα tweet σε περισσότερο κόσμο πέρα από τους followers του.

# 3

## ***Θεωρητικό υπόβαθρο***

### ***3.1 Influentials και μη Influentials***

Υπάρχουν πολλές θεωρίες γύρω από το πώς εξαπλώνεται μία ιδέα ή πώς υιοθετείται μια καινοτομία.

Οι περισσότερες μελέτες διατείνονται ότι μία μειοψηφία από άτομα, σε μια κοινωνία, έχουν κάποια ποιοτικά χαρακτηριστικά που τους κάνουν ιδιαίτερα ικανούς για τη μετάδοση ιδεών. Τα συγκεκριμένα άτομα (influentials) σύμφωνα με την παραδοσιακή έρευνα τείνουν να οδηγούν την υιοθέτηση νέων ιδεών. Συνεπώς, αναγνωρίζοντας και πείθοντας έναν μικρό αριθμό από influentials, μια μικρή εκστρατεία μπορεί να εξαπλωθεί σε έναν πολύ μεγάλο αριθμό ατόμων, με πολύ χαμηλό κόστος. Η συγκεκριμένη θεωρία είναι και η πιο διαδεδομένη και χρησιμοποιείται σε πολλές εταιρείες marketing.

Ωστόσο, πιο σύγχρονες μελέτες, συμπεράναν ότι τα άτομα της εποχής της πληροφορίας τείνουν να επηρεάζονται περισσότερο από τους συνομηλίκους και τους φίλους τους παρά από τους influentials.

Επιπροσθέτως, αρκετοί ερευνητές στέκονται στο γεγονός ότι η θεωρία των *influentials* για τη μετάδοση πληροφορίας και ιδεών δεν λαμβάνει υπόψη τα απλά άτομα. Προκειμένου να ληφθούν υπόψη και τα μη *influential* άτομα, έχουν αναπτυχθεί πολλές εργασίες και προσομοιώσεις. Το συμπέρασμα που βγήκε είναι ότι οι *influentials* τείνουν να ξεκινούν πιο συχνά μια ιδέα που υιοθετείται από πολλούς ανθρώπους, ωστόσο η παρουσία των *influentials* δεν είναι ούτε ικανή ούτε αναγκαία για την εξάπλωση μιας ιδέας. Άρα το αν θα υιοθετηθεί μια ιδέα και από ποιούς δεν έχει να κάνει τόσο με το πρόσωπο που προσπαθεί να την μεταδώσει, όσο με το πόσο έτοιμη είναι μια κοινωνία ανθρώπων να τη δεχτεί.

Επιπλέον, αξίζει να σημειώσουμε ότι το *influence* κάποιου αλλάζει στο χρόνο. Οι *influentials* μπορούν να χάσουν αρκετή από την επιρροή τους ή να κερδίσουν περισσότερη. Ακόμη, κάποιοι μη *influentials* μπορούν να κερδίσουν *influence*, όσο επιμένουν σε ιδέες στις οποίες έχουν αρκετές γνώσεις και εμπειρία.

### ***3.2 Topic sensitivity στο Twitter***

Διαβάζοντας την προηγούμενη ανάλυση, ένα εύστοχο ερώτημα θα μπορούσε να είναι ότι εφ' όσον μη *influentials* μπορούν να κερδίζουν *influence* αν προβάλλουν προτάσεις σε θέματα που είναι γνωσμένα καλοί, συμβαίνει το ίδιο και με τους *influentials*; Και πόσο μετράει, εν τέλει το αντικείμενο της ιδέας, την οποία κάποιος *influential* προτείνει; Όλοι οι *influentials* ασκούν το ίδιο *influence* σε όλα τα θέματα;

Επικεντρώνοντας την ανάλυση μας στο twitter και παραλληλίζοντας τις ιδέες με τα tweets, εύκολα αντιλαμβάνεται κανείς ότι ένα tweet, από κάποιον *influential*, που δεν έχει ενδιαφέρον δεν πρόκειται να αναπαραχθεί πολύ. Ανάλογα, λειτουργεί και το topic. Ένα κοινό χαρακτηριστικό των περισσότερων *influentials* είναι πως είναι γνωστοί στους ανθρώπους που επηρεάζουν, και κυρίως είναι γνωστοί για τις επιδόσεις ή την ενασχόληση τους με συγκεκριμένα θέματα. Στο twitter, αυτός είναι και ο κυριότερος λόγος για τον οποίο ακολουθούν έναν *influential* οι followers του. Συνεπώς όταν αυτός ασχολείται με θέματα διαφορετικά από το αντικείμενο του, όλο και περισσότεροι followers του θα θεωρήσουν αδιάφορες τις ιδέες του (tweets). Βέβαια από την άλλη μεριά, μπορεί να εξακολουθεί να ασκεί μεγαλύτερο *influence* από κάποιους μη *influential* twitterers, μιας και έχει πολύ

μεγαλύτερο κοινό για να επηρεάσει. Ωστόσο, σίγουρα ο ίδιος influential έχει πολύ μικρότερο influence σε θέματα διαφορετικά από το αντικείμενο το οποίο τον κάνει να ξεχωρίζει.

Για τους παραπάνω λόγους, ο αριθμός των followers που είναι το μέτρο που χρησιμοποιείται κατά κύριο λόγο, για να πούμε ότι κάποιος είναι influential στο twitter, κρίνεται ανεπαρκές μέτρο.

### ***3.3 Homophily και reciprocity στο Twitter***

Το αντικείμενο ενασχόλησης και αναγνώρισης, ενός influential, που αναφέρουμε στην προηγούμενη ενότητα, δεν είναι ο μοναδικός λόγος για τον οποίον ο αριθμός των followers είναι ανεπαρκές μέτρο για την αποτίμηση του influence στο Twitter. Υπάρχουν ακόμη οι λόγοι της αμοιβαιότητας (reciprocity) και της ομοφιλίας (homophily).

Reciprocity είναι το φαινόμενο που παρατηρείται όταν ένας twitterer κάνει follow έναν άλλον για λόγους ευγένειας και αμοιβαιότητας. Ένα παράδειγμα είναι όταν κάποιος twitterer κάνει follow κάποιον άλλον, επειδή του έκανε follow και ο δεύτερος. Το συγκεκριμένο φαινόμενο είναι πάρα πολύ συνηθισμένο στο Twitter, ίσως και λόγω της σχέσης friendship στο Facebook η οποία είναι αμοιβαία και απαιτεί επιβεβαίωση, ενδεχομένως, για πολλούς twitterers το follow back να είναι η αντίστοιχη επιβεβαίωση αιτήματος φιλίας. Την θεωρία αυτή επιβεβαιώνει και η εργασία των Weng, Lim, Jiang, He [2] στην οποία παρατηρήθηκε πως το 72.4% των χρηστών του twitter ακολουθούν πάνω από το 80% των followers τους, καθώς επίσης και ότι 80.5% των χρηστών twitter έχει followers το 80% αυτών που ακολουθεί.

Το homophily, από την άλλη, είναι το ακριβώς αντίστροφο φαινόμενο. Δηλαδή, όταν ένας twitterer ακολουθεί κάποιον άλλον, όχι γιατί επηρεάζεται από αυτόν, αλλά γιατί μοιράζονται κοινά ενδιαφέροντα και κοινές ιδέες.

### **3.4 Μετρικές *influence* στο Twitter**

Έχουν επινοηθεί αρκετοί τρόποι μέτρησης του *influence* στο twitter. Σε αυτή την ενότητα παραθέτουμε αυτούς που χρησιμοποιήθηκαν στην συγκεκριμένη εργασία. Κάθε ένας από αυτούς ασχολείται με ένα διαφορετικό είδος *influence*.

#### **3.4.1 *Followers* στο Twitter (*indegree influence*)**

Ο αριθμός των *followers* που έχει κάποιος *twitterer* δείχνει το μέγεθος του κοινού του. Δηλαδή, πόσοι άλλοι *twitterers* διαβάζουν τα μηνύματα (*tweets*) του πρώτου, άμεσα τη στιγμή που τα δημοσιεύει.

#### **3.4.2 *Retweets* στο Twitter (*Retweet influence*)**

Με το συγκεκριμένο μέτρο μετράμε τον αριθμό *retweets* που περιέχουν το όνομα κάποιου *twitterer* σε συγκεκριμένο χρονικό διάστημα, πρόκειται για *tweets* που περιέχουν είτε στην αρχή τη φράση “RT @*twitterer*’s\_username” είτε στο τέλος τη φράση “via @*twitterer*’s\_username”. Το μέτρο αυτό δείχνει την ικανότητα του *twitterer* να δημιουργεί *tweets* με ποιότητα περιεχομένου που αξίζουν να αναδημοσιευτούν.

#### **3.4.3 *Mentions* στο Twitter (*Mention Influence*)**

Πρόκειται για τις αναφορές “@*twitterer*’s\_username” εκτός των *retweets* που παίρνει ένας *twitterer* μέσα από τα *tweets* άλλων *twitterers* σε συγκεκριμένο χρόνο. Το μέτρο αυτό καταδικνύει την ικανότητα του *twitterer* να αλληλεπιδρά με το κοινό του καθώς επίσης και να βάζει και άλλους *twitterers* σε μια συζήτηση.

### 3.4.4 *Topic-specific Twitter Rank*

Διαισθητικά, θα μπορούσε κάποιος να σκεφτεί ότι ένας twitterer θα έχει μεγάλο influence εάν και το άθροισμα του influence των followers του είναι μεγάλο, όπως συμβαίνει στο διαδίκτυο με την αποτίμηση του authority μιας ιστοσελίδας. Ταυτόχρονα, όπως και στις ιστοσελίδες, θα μπορούσε να υποθέσει κάποιος ότι όσο περισσότερα μηνύματα περνούν από έναν twitterer στους followers του τόσο πιο μεγάλο θα είναι το influence που ασκεί. Αυτή η ομοιότητα οδηγεί τους ερευνητές στη χρησιμοποίηση του PageRank και κάποιων παραλλαγών του για τη μέτρηση του influence.

Ωστόσο, παρότι το authority ιστοσελίδων παρουσιάζει πολλές ομοιότητες με το influence ενός twitterer, υπάρχουν και κάποιες θεμελιώδεις διαφορές. Η κυριότερη είναι ότι το influence ενός twitterer δεν εξαρτάται αποκλειστικά από την ποσότητα των tweets, όσο κυρίως από την ποιότητα. Για παράδειγμα, κάποιος twitterer δεν θα δώσει σημασία σε θέματα συζήτησης που βρίσκει βαρετά ακόμη και αν προέρχονται από κάποιον πολύ influential twitterer και ανεξάρτητα της ποσότητας τους. Έτσι, από τη στιγμή που οι twitterers έχουν διαφορετικά ενδιαφέροντα και διαφορετικές εμπειρίες σε αρκετά θέματα, θα έπρεπε και το influence τους να διαφέρει ανάλογα με το θέμα. Ο αλγόριθμος topic-specific TwitterRank συνδυάζει την ιδέα του PageRank με το Topic sensitivity που αναλύουμε σε προηγούμενη ενότητα. Επιπλέον, λαμβάνει υπόψη το homophily και το reciprocity μεταξύ των twitterers. Ακολουθεί ο αλγόριθμος topic-specific TwitterRank.

Πρωτίστως, καταρτίζουμε τον κατευθυνόμενο γράφο  $G(V, E)$  με τους χρήστες και τις follow σχέσεις μεταξύ τους.  $V$  είναι το σύνολο κορυφών (κόμβων) που περιέχει όλους τους twitterers και  $E$  το σύνολο ακμών που περιλαμβάνει τις σχέσεις follow. Υπάρχει μία ακμή μεταξύ δύο χρηστών, όταν υπάρχει σχέση follow ενώ η κατεύθυνση της είναι από τον follower  $A$  στον friend  $B$ .

Διατρέχουμε όλους τους κόμβους (twitterers) του γράφου  $G$ , ακολουθώντας με συγκεκριμένη πιθανότητα τις ακμές  $E$ . Η διαφορά με τον PageRank είναι ότι η μετάβαση από έναν κόμβο σε έναν άλλον είναι Topic-specific. Για το λόγο αυτό κατασκευάζουμε τον Topic-specific πίνακα πιθανοτήτων μετάβασης από έναν κόμβο σε έναν άλλο, ως εξής:

Δοθέντος ενός topic  $t$  κάθε στοιχείο του πίνακα  $P_t$  δηλαδή η πιθανότητα μετάβασης από το follower  $s_i$  στον friend  $s_j$  ορίζεται ως εξής:

$$P_t(i,j) = \frac{|T_j|}{\sum_{a: s_i \text{ follows } s_a} |T_a|} * \text{sim}_t(i,j)$$

$|T_j|$  είναι ο αριθμός tweets που δημοσίευσε ο twitterer  $s_j$  ενώ  $\sum_{a: s_i \text{ follows } s_a} |T_a|$  είναι ο

αριθμός όλων των tweets που δημοσιεύτηκαν από όλους τους φίλους του twitterer  $s_i$ .  
 $\text{sim}_t(i,j)$  είναι η ομοιότητα μεταξύ των χρηστών  $s_i$  και  $s_j$  στο topic  $t$ .

Ο συγκεκριμένος ορισμός λαμβάνει υπόψη, αφ' ενός ότι όσο περισσότερο ένας twitterer δημοσιεύει tweets τόσο περισσότερο επηρεάζει τους followers του, αλλά από την άλλη, μόνο αυτούς που ενδιαφέρονται για το topic των tweets. Όσο πιο πολλές ομοιότητες έχουν 2 twitterers πάνω στο topic  $t$  τόσο μεγαλύτερη είναι και η πιθανότητα μετάβασης από τον έναν στον άλλον στο γράφο.

Επιπλέον, είναι πιθανό κάποιοι twitterers να ακολουθούν ο ένας τον άλλον δημιουργώντας κύκλους στο γράφο χωρίς να ακολουθούν twitterers εκτός κύκλου. Τέτοιοι κύκλοι προσδίδουν μεγάλο influence το οποίο όμως δεν μοιράζεται. Για το λόγο αυτό, στον υπολογισμό του TwitterRank εισάγεται το διάνυσμα μεταφοράς (teleportation vector)  $E_t$  το οποίο προβλέπει την πιθανότητα να μεταβούμε σε κόμβο χωρίς να ακολουθήσουμε κάποια ακμή.

Τελικά, ο Topic-specific TwitterRank υπολογίζεται αναδρομικά από τη σχέση:

$$\overrightarrow{TRt} = \gamma P_t \times \overrightarrow{TRt} + (1 - \gamma) E_t$$

Το  $\gamma$  είναι μια παράμετρος μεταξύ του 0 και του 1 που χρησιμοποιείται για να ελέγξει την πιθανότητα του teleportation.



### 3.5 *Twitter APIs*

Το twitter παρέχει στους χρήστες του μια σειρά από APIs για να διευκολύνει την ανάπτυξη εφαρμογών που χρησιμοποιούν το twitter με διάφορους τρόπους. Κάθε API παρέχει διαφορετικές δυνατότητες και ενδείκνυται για συγκεκριμένες εφαρμογές.

Συγκεκριμένα, υπάρχει το Twitter for Websites, παρέχει ενσωματώσιμα widgets, κουμπιά και client-side εργαλεία για integration του Twitter με websites και Javascript εφαρμογές. Παρέχει λειτουργίες όπως το tweet και το follow button.

Ακόμη, υπάρχει το Twitter Cards που αναπαριστά επιπλέον περιεχόμενο εκτός από το tweet. Χρησιμοποιείται για να δημιουργία μεγαλύτερης κίνησης στα websites που το χρησιμοποιούν.

Επιπροσθέτως, το Ads API και το MoPub χρησιμοποιούνται για τη διαφήμιση, το Fabric παρέχει εργαλεία ανάπτυξης (development tools) που βοηθούν να γίνει μια εφαρμογή πιο stable.

Επιπλέον υπάρχει το OAuth για να συνδέονται οι χρήστες, ώστε να χρησιμοποιούνται τα APIs με ασφάλεια και επωνυμία.

Τέλος, υπάρχουν και τα REST και Streaming APIs, για να αντλεί κάποιος δεδομένα από το twitter. Στην εργασία αυτή χρησιμοποιήθηκε το REST API γιατί παρέχει πρόσβαση σε βασικές λειτουργίες, όπως οι πληροφορίες για το χρήστη και οι ανανεώσεις κατάστασης, καθώς επίσης και πολλές δυνατότητες για χειρισμό όλων αυτών των στοιχείων.

# 4

## *Ανάλυση Απαιτήσεων Συστήματος*

### *4.1 Συλλογή δεδομένων*

Για να ανιχνεύσουμε τους πιο influential twitterers πάνω σε κάποιο topic χρειαζόμαστε κάποιο dataset από tweets σχετικά με το topic.

Αυτό το dataset, το παίρνουμε, κάθε φορά, μέσω του twitter API. Συγκεκριμένα, κάθε φορά που αρχίζει ένα καινούριο campaign κάνουμε αναζήτηση για κάθε keyword που έχει δώσει ο τελικός χρήστης της εφαρμογής, παίρνοντας τα 10.000 τελευταία tweets σχετικά με το συγκεκριμένο keyword. Για παράδειγμα, αν έχουν δοθεί τέσσερα keywords από τον τελικό χρήστη, παίρνουμε μέχρι 40.000 tweets.

Από αυτά τα tweets παίρνουμε τους τελικούς influentials. Αυτό γίνεται, παίρνοντας όλους τους χρήστες από τους οποίους έγιναν τα tweets και κρατώντας τους 50 με τους περισσότερους followers.

Τέλος, μαζεύουμε τα διαθέσιμα tweets του κάθε influential για την περίοδο που μελετάμε. Αυτά μας χρειάζονται τόσο για την παρουσίαση στο τελικό χρήστη της εφαρμογής όσο και για τον υπολογισμό του TwitterRank.

## **4.2 Υπολογίζοντας το influence**

Στις παρακάτω ενότητες παρουσιάζεται ο τρόπος υπολογισμού των μέτρων του influence για την τελική κατάταξη των χρηστών.

### **4.2.1 Indegree, Retweets, Mentions**

Το indegree δηλαδή ο αριθμός των followers είναι διαθέσιμος ήδη από τη στιγμή που έχουμε καταλήξει στους τελικούς influentials.

Όσο για τα retweets και τα mentions τα βρίσκουμε ως εξής: Αρχικά κάνουμε αναζήτηση στο twitter με keyword το @influential's\_username και παίρνουμε όλα τα tweets που περιέχουν το keyword μέχρι την ημερομηνία που έχει δώσει ο τελικός χρήστης της εφαρμογής μέσω του period. Αν έχει δώσει για παράδειγμα 5 μέρες η αναζήτηση μας φτάνει μέχρι 5 μέρες πίσω.

Έχοντας πλέον τα tweets που αναφέρονται στον κάθε influential χρήστη τα ταξινομούμε σε Retweets ή Mentions και βγάζουμε την τελική κατάταξη.

### **4.2.2 Υπολογισμός TwitterRank**

Επειδή ο TwitterRank όπως περιγράφηκε στο προηγούμενο κεφάλαιο, δείχνει το πόσο influence δέχεται ένας twitterer, ενώ εμείς ενδιαφερόμαστε για το πόσο influence ασκεί κάναμε μία αλλαγή στο παραπάνω μοντέλο.

Συγκεκριμένα, αλλάξαμε την έννοια του transition probability ( $P_i$ ). Έτσι, αντί για την πιθανότητα μετάβασης  $P_i$  χρησιμοποιήσαμε την πιθανότητα επιρροής  $P^i$ , από τον friend  $s_i$  στον follower  $s_j$  που ορίζεται ως:

$$P'_t(i,j) = \frac{|Ti|}{\sum_{a:s_j \text{ follows } sa} |Ta|} * \text{sim}_t(i,j)$$

$|Ti|$  είναι ο αριθμός tweets που δημοσίευσε ο twitterer  $s_i$  ενώ  $\sum_{a:s_j \text{ follows } sa} |Ta|$  είναι ο αριθμός όλων των tweets που δημοσιεύτηκαν από όλους τους φίλους του twitterer  $s_j$ ,  $\text{sim}_t(i,j)$  είναι η ομοιότητα μεταξύ των twitterers  $s_i$  και  $s_j$  στο topic  $t$ .

Τελικά το TwitterRank influence  $TR'$  δίνεται από τον τύπο:

$$\overrightarrow{TR't} = \gamma P't \times \overrightarrow{TR't} + (1 - \gamma) E_t$$

Όσον αφορά τα μεγέθη similarity ( $\text{sim}_t$ ),  $\gamma$  και teleportation στο topic ( $E_t$ ), το  $\gamma$  το έχουμε θέσει 0.5 ενώ ακολουθεί η περιγραφή του τρόπου που υπολογίζουμε τα άλλα δυο.

- **Similarity ( $\text{sim}_t(i,j)$ ):** Για να βρούμε το similarity μεταξύ 2 twitterers πάνω στο topic αρχικά βγάζουμε τον κανονικοποιημένο πίνακα  $DT_a$  που περιλαμβάνει τις λέξεις που χρησιμοποιούνται για το topic και τη σχετική συχνότητα τους. Αυτό γίνεται αναλύοντας το αρχικό σώμα tweets στο topic σε πίνακα λέξεων με τις σχετικές τους συχνότητες. Στη συνέχεια, κάνουμε το ίδιο και για τα tweets που έχουμε μαζέψει για τους influentials, κρατώντας μόνο τις λέξεις που χρησιμοποιήθηκαν (με τουλάχιστον 4 γράμματα) στα tweets του topic και τη σχετική τους συχνότητα για όλα τα tweets του συγκεκριμένου influential. Αυτός ο πίνακας για τον influential  $s_i$  στο topic  $t$  είναι ο  $DT'_{it}$ . Τελικά το similarity μεταξύ δύο χρηστών  $s_i$  και  $s_j$  υπολογίζεται ως:

$$\text{sim}_t(i,j) = 1 - |DT'_{it} - DT'_{jt}|$$

- **Teleportation ( $E_t$ ):** Ορίζεται ως εξής:

$$E_t(i) = DT''_{it}$$

Όπου  $DT''_t$  είναι ο κανονικοποιημένος πίνακας  $DT$  στο topic  $t$  για όλους τους influentials. Το στοιχείο  $DT''_t(i)$  του πίνακα δείχνει την πιθανότητα ένα tweet στο topic  $t$  να ανήκει στο twitterer  $i$ .

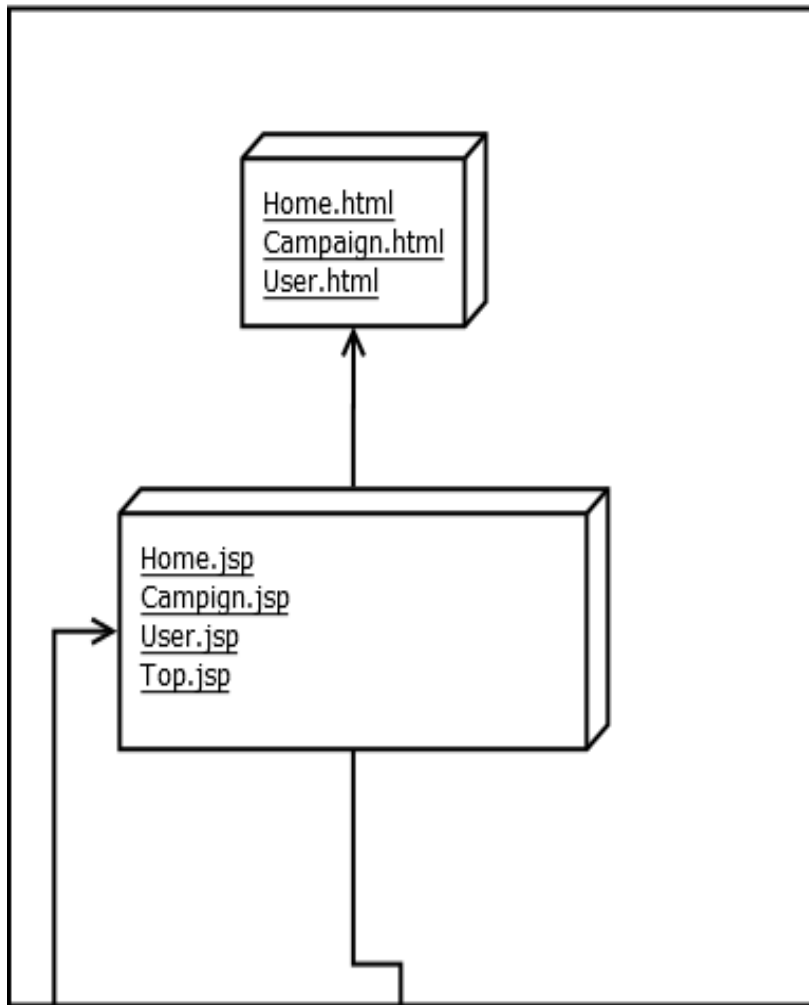
Τέλος, αξίζει να σημειωθεί ότι για την κατάρτιση του γράφου  $G$  για κάθε influencer  $s_i$  παίρνουμε τη λίστα με τους φίλους του, και αν στη λίστα περιλαμβάνεται κάποιος influencer  $s_j$ , προσθέτουμε την ακμή στο γράφο.

### **4.3 Αρχιτεκτονική**

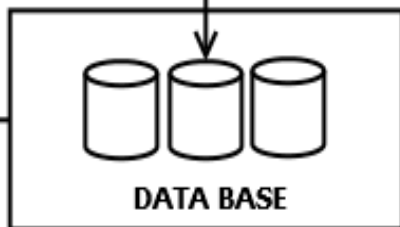
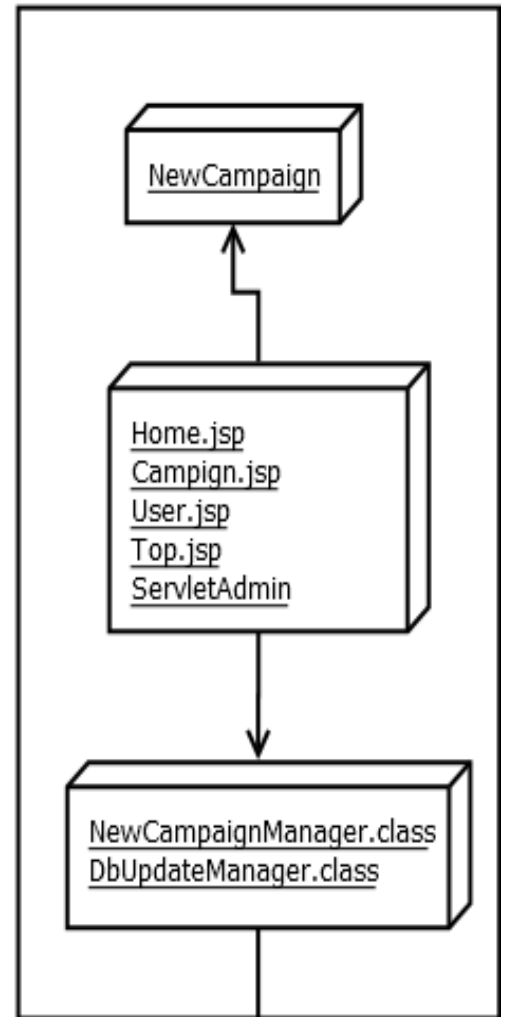
Η αρχιτεκτονική του συστήματος είναι αρκετά απλή. Συγκεκριμένα, το σύστημα μας αποτελείται από 2 υποσυστήματα. Το πρώτο (έστω Υποσύστημα 1) είναι υπεύθυνο για την παροχή δεδομένων και τη σωστή εμφάνιση των σελίδων html που βλέπει ο τελικός χρήστης ενώ το δεύτερο υποσύστημα (έστω Υποσύστημα 2) είναι υπεύθυνο για τα new Campaigns. Τα υποσυστήματα αυτά έχουν κοινά σημεία τη βάση δεδομένων και κάποιες ενδιάμεσες jsp σελίδες και servlet τα οποία συντονίζουν όλη τη λειτουργία του συστήματος. Ο λόγος που τα διαχωρίζουμε μεταξύ τους είναι ότι ενώ για το Υποσύστημα 1 οι jsp σελίδες αλληλεπιδρούν κατ'ευθείαν με τη βάση δεδομένων (κάνοντας μόνο select), στο υποσύστημα 2 απλά δημιουργούν ένα Thread που αναλαμβάνει να δημιουργήσει το νέο campaign καλώντας την κλάση NewCampaignManager.class, η οποία είναι υπεύθυνη για την περάτωση του campaign και για την καταχώρηση των αποτελεσμάτων στη βάση δεδομένων μέσω της κλάσης DbUpdateManager.

Ακολουθεί ένα block diagram που δείχνει την αρχιτεκτονική του συστήματος που περιγράφηκε παραπάνω.

### ΥΠΟΣΥΣΤΗΜΑ 1



### ΥΠΟΣΥΣΤΗΜΑ 2

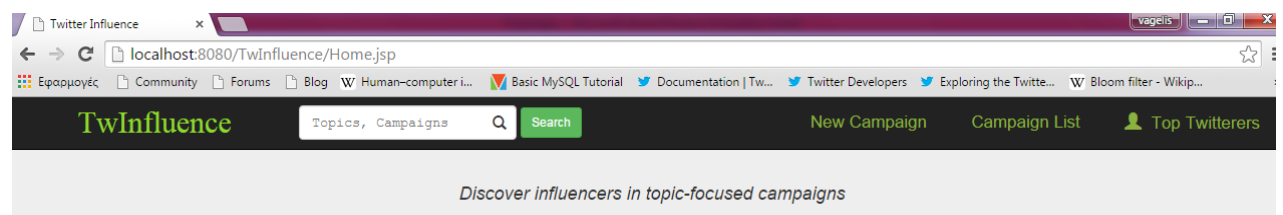


Όσον αφορά την αρχιτεκτονική που χρησιμοποιείται για την ανάλυση και τον υπολογισμό ενός new campaign είναι απλή αρχιτεκτονική ροής δεδομένων, όπου κάθε υπομονάδα του συστήματος περιμένει την προηγούμενη να τελειώσει ώστε να χρησιμοποιήσει τα δεδομένα που εξάγονται. Περισσότερες λεπτομέρειες στο επόμενο κεφάλαιο όπου γίνεται εκτενής ανάφορά και ανάλυση των κυριότερων κλάσεων της εφαρμογής.

## 4.4 Περιγραφή βασικών λειτουργιών του συστήματος

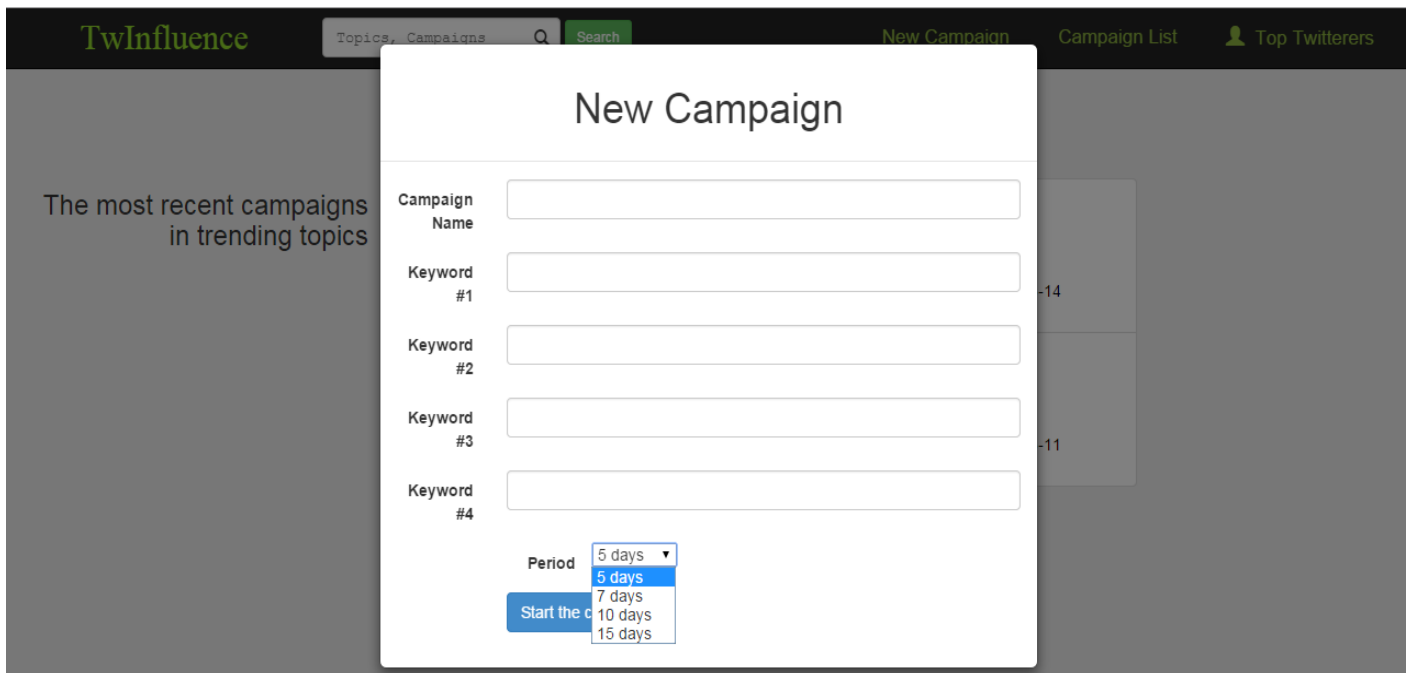
### 4.4.1 Έναρξη νέου campaign

Η εφαρμογή μας, παρέχει στον τελικό χρήστη τη δυνατότητα να ξεκινήσει ένα νέο campaign μέσω του συνδέσμου New Campaign όπως φαίνεται και στη φωτογραφία.



Η επιλογή New Campaign διακρίνεται εύκολα στο πάνω δεξί μέρος της σελίδας.

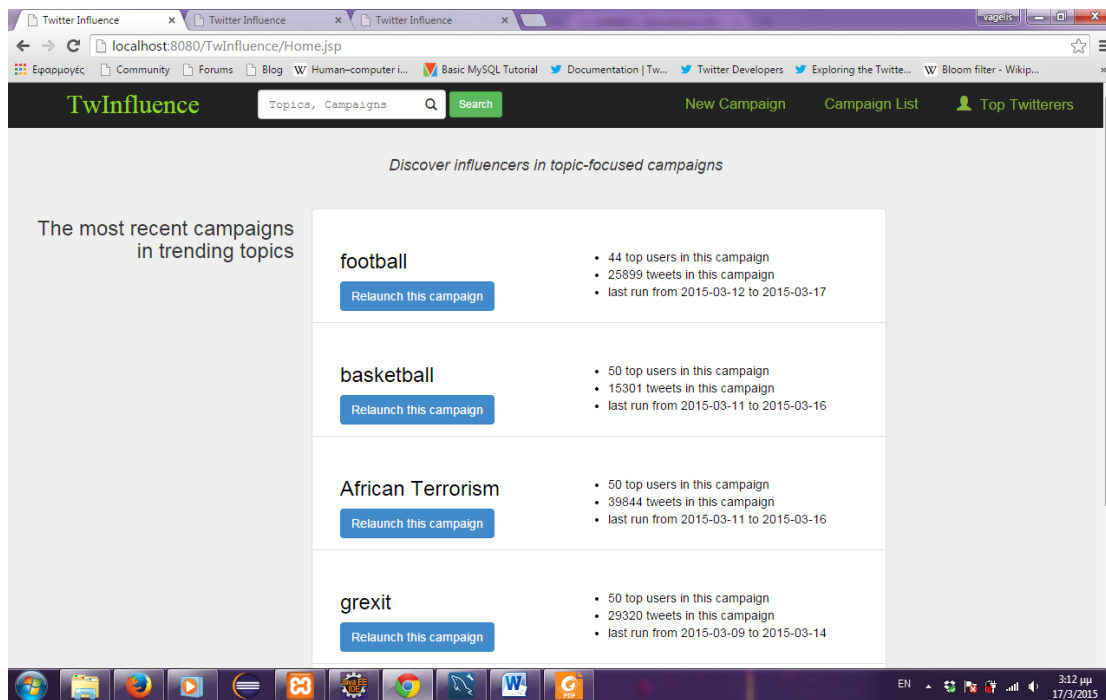
Με το πάτημα του New Campaign εμφανίζεται μια φόρμα στο χρήστη που περιλαμβάνει πεδία (text fields) για το όνομα του campaign, αλλά και για keywords μέσω των οποίων θα γίνει η αναζήτηση. Τέλος, παρέχεται dropdown list όπου ο τελικός χρήστης επιλέγει περίοδο για την ανάλυση ανάμεσα στις επιλογές 5, 7, 10 και 15 μέρες. Στην παρακάτω εικόνα παρουσιάζεται η φόρμα αυτή.



#### 4.4.2 *Επισκόπηση των campaigns που έχουν ολοκληρωθεί*

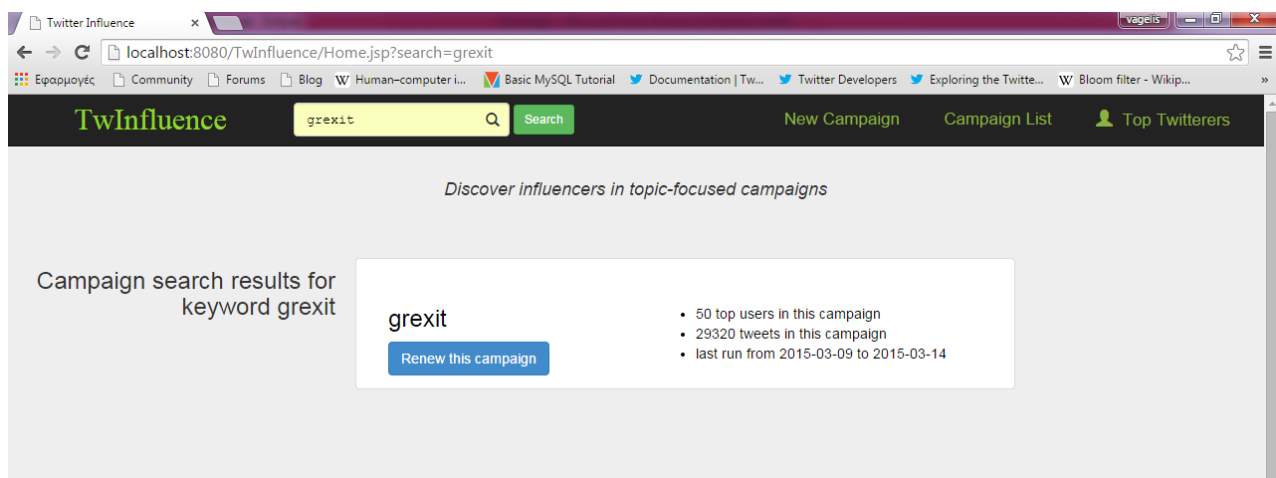
Πρόκειται για την αρχική σελίδα της εφαρμογής, η οποία απεικονίζει τα πιο πρόσφατα ενημερωμένα campaigns, παρέχοντας βασικές πληροφορίες για αυτά. Συγκεκριμένα, αναφέρεται ο αριθμός των influentials του συγκεκριμένου campaign, τα tweets που υπάρχουν από τους twitterers του campaign, αλλά και το χρονικό διάστημα κατά το οποίο έγινε η ανάλυση. Επίσης, σε κάθε campaign παρέχονται οι δυνατότητες για μετάβαση στο campaign, πατώντας κλικ στον σύνδεσμο πάνω στο όνομα του campaign, αλλά και ανανέωσης του campaign μέσω του κουμπιού “Relaunch this campaign”. Ακολουθεί εικόνα που απεικονίζει όσα αναφέρθηκαν σε αυτή την ενότητα.

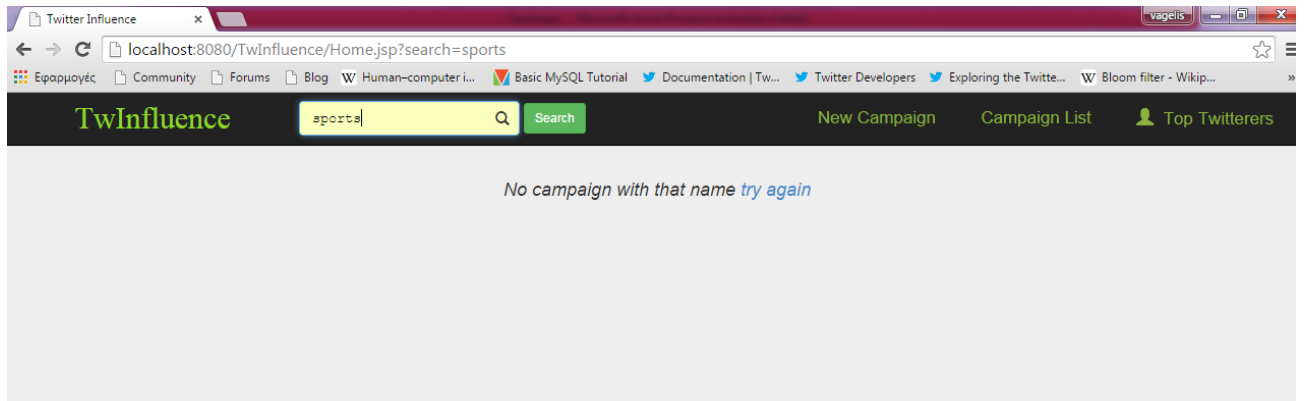




#### 4.4.3 Αναζήτηση campaign

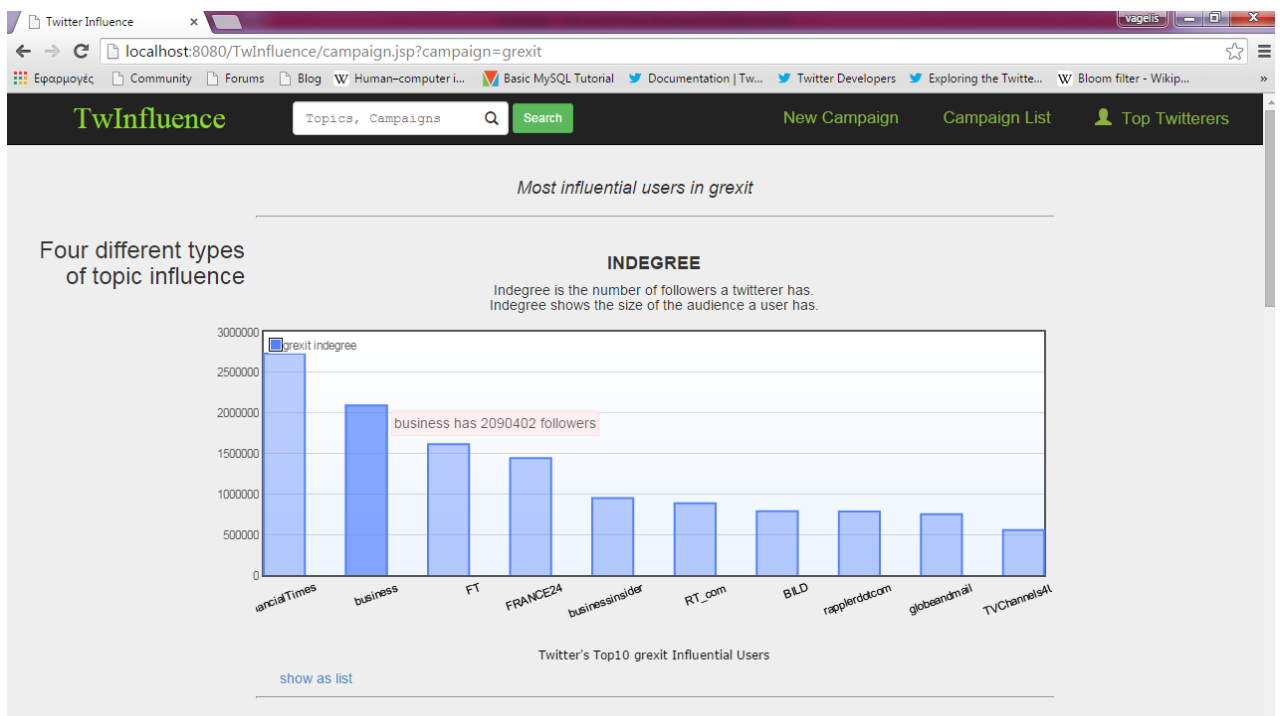
Όπως φαίνεται και στην πρώτη εικόνα της ενότητας 4.4.1 παρέχεται η δυνατότητα αναζήτησης campaign. Συγκεκριμένα ο τελικός χρήστης εισάγει στη μπάρα αναζήτησης το όνομα κάποιου campaign που ψάχνει και αν δεν υπάρχει εμφανίζεται μήνυμα λάθους, ενώ αν υπάρχει εμφανίζεται στην οθόνη. Χαρακτηριστικά παραδείγματα φαίνονται στις εικόνες που ακολουθούν. Η πρώτη δείχνει τη σελίδα αν βρεθεί το campaign που αναζητήθηκε ενώ η δεύτερη το μήνυμα ότι το campaign δεν πέτυχε.

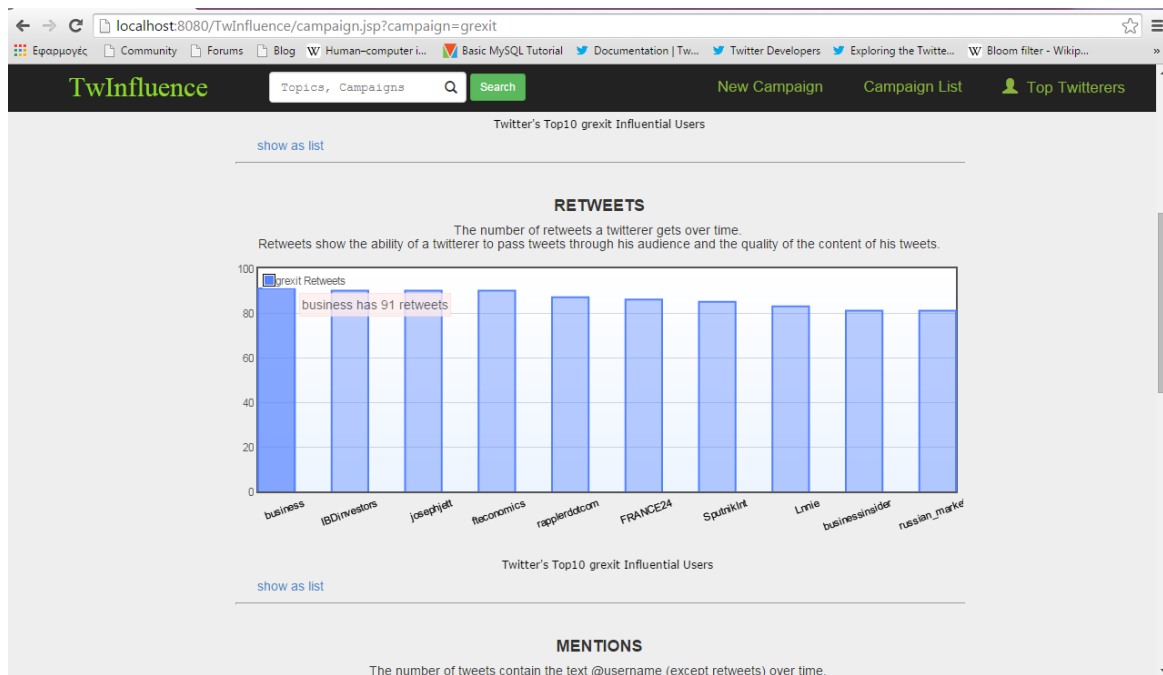




#### 4.4.4 Διαγραμματική απεικόνιση αποτελεσμάτων campaign

Η εφαρμογή μας παρέχει επίσης τη δυνατότητα στον τελικό χρήστη να δει διαγραμματικά τα αποτελέσματα του campaign, για κάθε μία από τις τέσσερις μετρικές που έχουμε χρησιμοποιήσει. Επίσης, αξίζει να σημειωθεί ότι όταν ο χρήστης πάει το ποντίκι πάνω από κάποια μπάρα του διαγράμματος εμφανίζεται tooltip με τα αποτελέσματα, ακόμη, κάνοντας κλικ σε κάποια μπάρα ανακατευθύνεται σε νέα σελίδα με πληροφορίες για τον συγκεκριμένο influential. Ακολουθούν εικόνες που δείχνουν αυτή τη λειτουργικότητα για το campaign “grexit”.





#### 4.4.5 Παρουσίαση λίστας Influentials βάσει αποτελεσμάτων campaign

Όπως φαίνεται κι από τις δύο προηγούμενες εικόνες, κάτω από τα διαγράμματα υπάρχει η επιλογή “show as list”. Αν ο χρήστης πατήσει το συγκεκριμένο σύνδεσμο έχει τη δυνατότητα να δει τους influencers σε μορφή λίστας, η οποία είναι μεν ταξινομημένη με βάση το μέτρο που κοιτούσαμε, ωστόσο δίνει πιο γενικές πληροφορίες για τον κάθε influential, όπως περιγραφή, τοποθεσία κλπ. Επίσης στην συγκεκριμένη σελίδα δίνεται και η δυνατότητα ταξινόμησης με βάση άλλους αλγορίθμους (δεξί μέρος της σελίδας). Τέλος, αξίζει να σημειωθεί, ότι κάνοντας κλικ είτε στο όνομα του twitterer είτε στην φωτογραφία του γίνεται ανακατεύθυνση σε σελίδα παρουσίασης του συγκεκριμένου twitterer, που περιγράφεται στην επόμενη ενότητα. Ακολουθεί εικόνα αυτής της λειτουργίας για το campaign “gredit”.

localhost:8080/Twinfluence/userList.jsp?campaign=grexit&run=1&algo=mnt

TwInfluence Topics, Campaigns Search New Campaign Campaign List Top Twitterers

Most influential users in grexit.

Users with the best score in mentions

	<p><b>BILD (@BILD)</b> Berlin GREXIT influential</p> <p>Hier twittern die BILD-Redaktion in Berlin und die Reporter aus aller Welt · Impressum: <a href="http://t.co/YZZZQPyBRS">http://t.co/YZZZQPyBRS</a></p> <ul style="list-style-type: none"> <li>followers: 786292</li> <li>friends: 522</li> <li>tweets: 109914</li> </ul>	<p>Check out the results with other algorithms</p> <p><a href="#">indegree</a></p> <p><a href="#">retweets</a></p> <p><a href="#">twitterRank</a></p>
	<p><b>Report on Business (@globebusiness)</b> Canada GREXIT influential</p> <p>The latest headlines from The Globe and Mail's Report on Business. Share info anonymously with us through SecureDrop: <a href="https://t.co/wqZUZhWagJ">https://t.co/wqZUZhWagJ</a></p> <ul style="list-style-type: none"> <li>followers: 70047</li> <li>friends: 416</li> <li>tweets: 96590</li> </ul>	
	<p><b>John osalvo (@jojokejohn)</b> Nevada GREXIT influential</p> <p>For you to peruse&amp;enjoy- toys&amp;collectibles-<a href="http://t.co/a9qQrh7r69">http://t.co/a9qQrh7r69</a> &amp; @lovetheytoyz all are important, o.k. to be funny but not ill-spirited</p> <ul style="list-style-type: none"> <li>followers: 85274</li> <li>friends: 86382</li> <li>tweets: 575526</li> </ul>	
	<p><b>José Carlos Díez (@josediez)</b> España GREXIT influential</p> <ul style="list-style-type: none"> <li>followers: 78823</li> <li>friends: 825</li> <li>tweets: 65936</li> </ul>	

#### 4.4.6 Απεικόνιση καρτέλας influential χρήστη στο campaign

Στην ενότητα αυτή παρέχουμε αρκετές πληροφορίες για έναν συγκεκριμένο χρήστη και με διαφορετικούς τρόπους, οπότε κρίνεται σκόπιμο να παρουσιάσουμε ξεχωριστά όλες τις επιμέρους λειτουργίες.

##### 4.4.6.1 Παρουσίαση πληροφοριών και επίδοσης στο campaign

Στο συγκεκριμένο κομμάτι παρουσιάζεται καρτέλα με πληροφορίες για τον twitterer στο twitter από τη μία και από την άλλη η επίδοση του σε κάθε αλγόριθμο στο campaign. Η επίδοση σε κάθε αλγόριθμο δίνεται σε progress bars, όπου φαίνεται η θέση που κατέλαβε, αλλά και η απόλυτη απόδοση σε σχέση με άλλους twitterers. Ακολουθεί εικόνα για τη συγκεκριμένη λειτουργία για τον twitterer Telegraph News στο campaign “grexit”.

localhost:8080/TwInfluence/user.jsp?campaign=grexit&run=1&username=TelegraphNews

Εφαρμογές Community Forums Blog W Human-computer i... Basic MySQL Tutorial Documentation | Tw... Twitter Developers Exploring the Twitte... Bloom filter - Wikip...

**TwInfluence** Topics, Campaigns Search New Campaign Campaign List Top Twitterers

**Telegraph News (@TelegraphNews)**, London, UK GREXIT influential

Breaking news, analysis and more from The Daily and Sunday Telegraph. Facebook: <http://t.co/Yav0GnTnzi> Got a tip? [mynews@telegraph.co.uk](mailto:mynews@telegraph.co.uk)  
[Telegraph News on twitter.](#)

tweets: 121422 followers: 515181 friends: 491

Indegree: 11th  
 Retweets: 20th  
 Mentions: 23th  
 TwitterRank: 25th

Favorite Tweets Thematic Navigation Tweets Map Tweets over time

#### 4.4.6.2 Παρουσίαση σημαντικότερων tweets κατά το campaign

Πρόκειται για τη λειτουργικότητα της καρτέλας “Favorite Tweets” που φαίνεται στο παραπάνω σχήμα. Εδώ παρουσιάζονται τα 20 tweets που έγιναν πιο πολύ favorite (most favorited tweets) κατά την διάρκεια του campaign. Επίσης φαίνεται το πόσες φορές επισημάνθηκε ως favorite πόσες φορές έγινε retweet, αλλά και πόσο καιρό πριν έγινε το tweet. Στην παρακάτω εικόνα φαίνεται αυτή η λειτουργία για το twitterer Telegraph News στο campaign “grexit”.

The Solar Impulse 2 takes off, starting its attempt to fly around the world without fuel <http://t.co/5mbQDKAwzk> <http://t.co/H8bOC7hZ9p> Favorites: 168 Retweets: 292 6 days ago

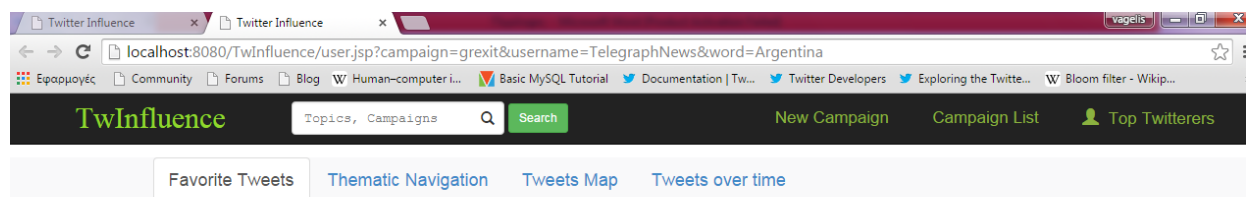
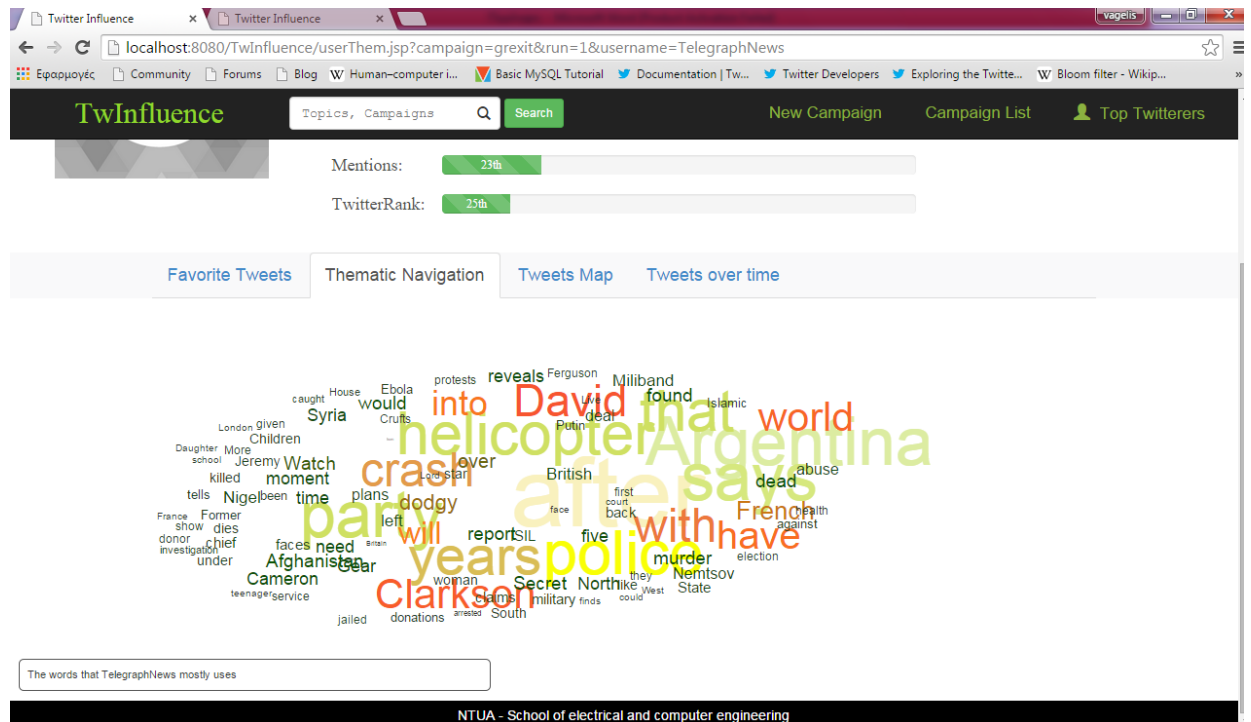
How a farewell cuddle revived a dying baby <http://t.co/HXL2VQUYGT> <http://t.co/oL9WyWf32i> Favorites: 104 Retweets: 89 2 days ago

Two mummified bodies found of climbers who disappeared on Mexico s highest peak 55 years ago <http://t.co/KsxxMyWSPH> <http://t.co/bxBq9yAhz3> Favorites: 88 Retweets: 160 8 days ago

Solar eclipse will plunge Britain into morning twilight <http://t.co/Svae6mFH8L> <http://t.co/D0KksSStXK> Favorites: 83 Retweets: 126 7 days ago

#### 4.4.6.3 Θεματική πλοήγηση στα tweets κατά το campaign

Το tab “Thematic Navigation” παρέχει στον τελικό χρήστη ένα Word Cloud, ώστε να μπορεί να προσπελάσει τα tweets του συγκεκριμένου twitterer με βάση το θέμα τους. Όταν κάνει κλικ σε κάποια από τις λέξεις ανοίγει ένα νέο tab στον browser που δείχνει τα tweets που περιέχουν τη συγκεκριμένη λέξη. Ακολουθούν εικόνες από ένα παράδειγμα για το twitterer Telegraph News στο campaign “grexit”. Στο παράδειγμα πατάμε πάνω στη λέξη “Argentina” και ανοίγει μια νέα καρτέλα με τα tweets με τη λέξη αυτή.



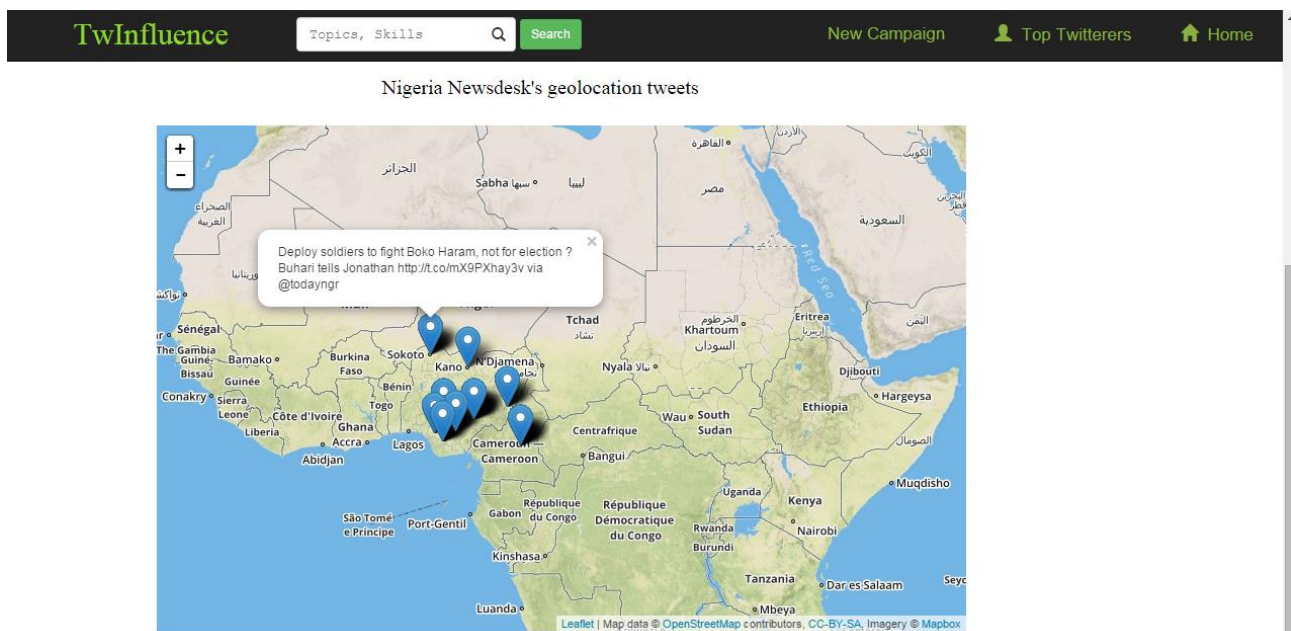
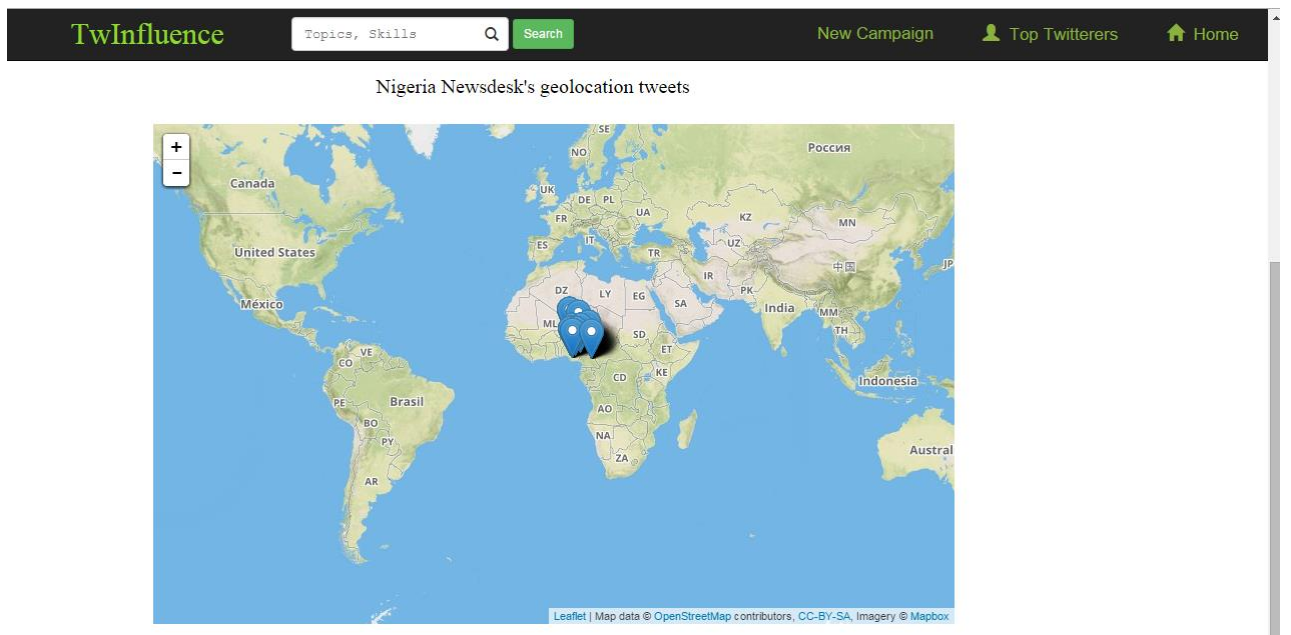
Telegraph News's tweets during the campaign with the word Argentina

French Olympians filming reality TV survival show killed in helicopter crash in Argentina <a href="http://t.co/LJCCFbauFs">http://t.co/LJCCFbauFs</a> <a href="http://t.co/RsNO3vtj">http://t.co/RsNO3vtj</a>	Favorites: 27 Retweets: 79 5 days ago
French Olympians killed in Argentina helicopter crash <a href="http://t.co/P87zufDzmm">http://t.co/P87zufDzmm</a> <a href="http://t.co/nX1QbmT8DH">http://t.co/nX1QbmT8DH</a>	Favorites: 19 Retweets: 67 5 days ago
France mourns sporting stars killed in Argentina helicopter crash <a href="http://t.co/ByKDEvDP2x">http://t.co/ByKDEvDP2x</a> <a href="http://t.co/WdKTX4dsKw">http://t.co/WdKTX4dsKw</a>	Favorites: 13 Retweets: 27 5 days ago

#### 4.4.6.4 Αναπαράσταση tweets στο χάρτη κατά το campaign

Η αλήθεια είναι ότι τη λειτουργία του GeoLocation tag στα tweets την χρησιμοποιούν πολύ λίγοι twitterers. Ωστόσο, κατά την άποψη μου, είναι μία πολύ χρήσιμη λειτουργία, ειδικά για ειδησεογραφικά sites, καθώς θα ήταν πολύ χρήσιμο για τους πελάτες τους να μπορούν να βλέπουν πού αναφέρονται οι ειδήσεις στο χάρτη, ώστε να κρίνουν αν τους ενδιαφέρουν ή όχι.

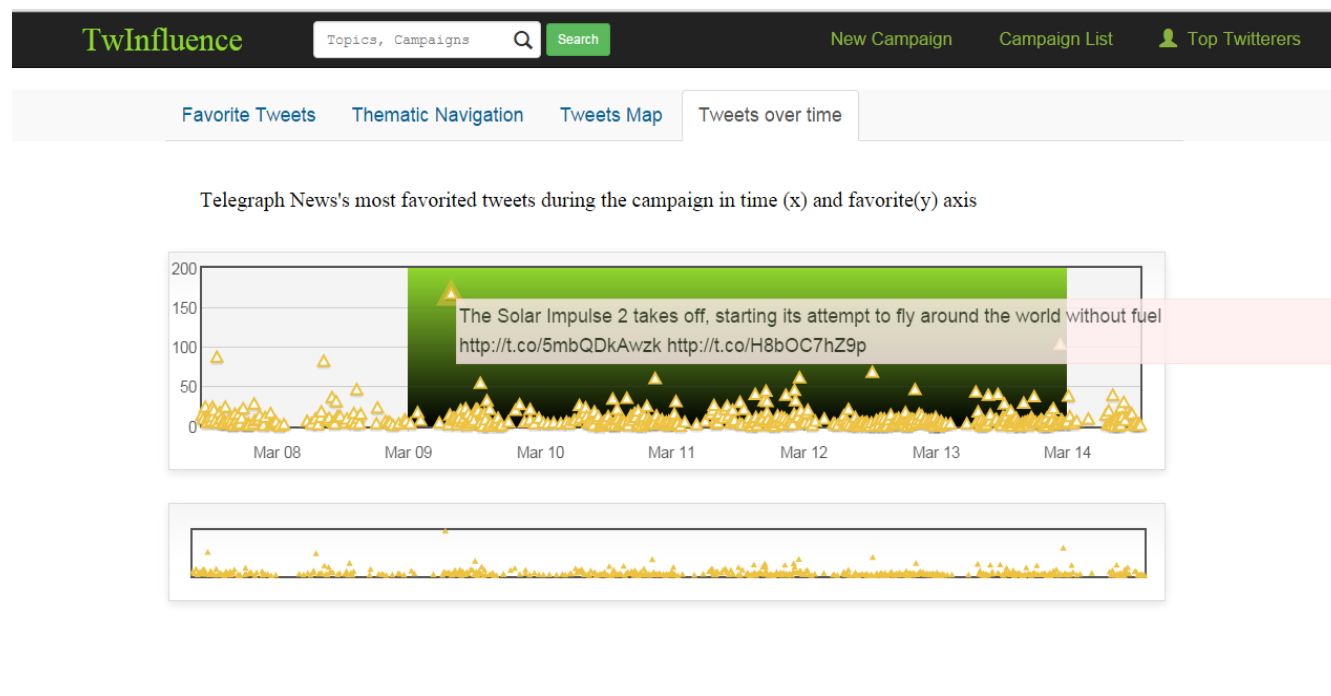
Αυτός είναι και ο λόγος που υλοποιήθηκε η λειτουργία Tweets Map ώστε να μπορούν να γνωρίζουν οι τελικοί χρήστες της εφαρμογής αν υπάρχουν γεωχωρικά δεδομένα, και να μπορούν να διαβάζουν tweets ανάλογα με την τοποθεσία που τους ενδιαφέρει. Παρακάτω παρατίθεται ο χάρτης tweets του twitterer NigeriaNewsdesk, που είναι ένας από τους ελάχιστους που χρησιμοποιούν το GeoLocation tag αρκετά συχνά. Η συγκεκριμένη λειτουργικότητα παρέχεται από το tab “Tweets Map”





#### 4.4.6.5 Απεικόνιση tweets στη μονάδα του χρόνου κατά το campaign

Πρόκειται για το tab “Tweets over time” και δείχνει όλα τα tweets που επισημάνθηκαν ως favorite αλλά και το πόσες φορές όπως και τον χρόνο στον οποίο δημοσιεύτηκαν. Συγκεκριμένα, αυτά φαίνονται σε μία γραφική παράσταση, όπου στον x-άξονα απεικονίζεται το χρόνο και στον y το favoritesCount (δηλαδή πόσες φορές επισημάνθηκε ως favorite ένα tweet). Τα σημεία πάνω στη γραφική παράσταση αναπαριστούν τα tweets και όταν περάσουμε το ποντίκι από πάνω τους εμφανίζεται tooltip με το tweet. Στην γραφική μας παράσταση, τα σαββατοκύριακα έχουν σημειωθεί με χρώμα ανοιχτό γκρι, ενώ οι καθημερινές με πράσινο. Τέλος παρέχεται η δυνατότητα zoom και unzoom μέσω του overview που βρίσκεται κάτω από τη γραφική παράσταση. Ακολουθεί εικόνα που παρουσιάζει τη λειτουργία αυτή για το twitterer Telegraph News στο campaign “grexit”.





#### 4.4.7 Παρουσίαση λίστας *influential* από όλα τα campaigns

Πρόκειται για τον σύνδεσμο “Top Twitterers”, στο πανθαρ στο πάνω μέρος των σελίδων της εφαρμογής. Είναι παρόμοια λειτουργία με την ενότητα 4.4.5. Οι διαφορές είναι ότι η ταξινόμηση βγαίνει μόνο από τους followers και είναι γενική για όλα τα campaigns, επίσης σε κάθε twitterer της λίστας αναγράφονται όλα τα campaign που είναι influential.

#### 4.4.8 Απεικόνιση καρτέλας *influential* χρήστη

Παρόμοια λειτουργικότητα με την ενότητα 4.4.6, με την διαφορά ότι εδώ δεν παρέχεται η δυνατότητα αναπαράστασης tweets στο χρόνο και ούτε η επίδοση του twitterer σε κάποιο συγκεκριμένο campaign

##### 4.4.8.1 Παρουσίαση γενικών χαρακτηριστικών

Εδώ απουσιάζουν τα progress bars σε σχέση με το 4.4.6.1. Ακολουθεί η εικόνα για τη λειτουργία αυτή για το twitterer Telegraph News.

TwInfluence

Topics, Campaigns Search

New Campaign Campaign List Top Twitterers

Telegraph News (@TelegraphNews), London, UK **GREXIT** influential

Breaking news, analysis and more from The Daily and Sunday Telegraph. Facebook: <http://t.co/Yav0GnTnzi> Got a tip? [mynews@telegraph.co.uk](mailto:mynews@telegraph.co.uk)  
[Telegraph News on twitter.](#)

tweets: 121422 followers: 515181 friends: 491

Favorite Tweets Thematic Navigation Tweets Map

Telegraph News's tweets during the campaign

The Solar Impulse 2 takes off, starting its attempt to fly around the world without fuel <http://t.co/5mbQDkAwzk> <http://t.co/H8bOC7hZ9p> Favorites: 168 Retweets: 292 6 days ago

#### 4.4.8.2 Παρουσίαση σημαντικότερων tweets στα campaigns που συμμετείχε

Η ίδια ακριβώς λειτουργία με την 4.4.6.2 η μόνη διαφορά είναι ότι περιλαμβάνει τα tweets από όλα τα campaigns.

#### 4.4.8.3 Θεματική πλοήγηση στα tweets

Η ίδια ακριβώς λειτουργία με την 4.4.6.3 η μόνη διαφορά είναι ότι περιλαμβάνει τα tweets από όλα τα campaigns.

#### 4.4.8.4 Αναπαράσταση tweets στο χάρτη

Η ίδια ακριβώς λειτουργία με την 4.4.6.4 η μόνη διαφορά, ότι περιλαμβάνει τα tweets από όλα τα campaigns.

## 4.5 Μοντέλο Οντοτήτων συσχετίσεων

Για την υλοποίηση των λειτουργιών που περιγράφονται στην προηγούμενη ενότητα κρίθηκε αναγκαία η ανάπτυξη μιας βάσης δεδομένων που αποθηκεύει κάποια δεδομένα για κάθε campaign που ολοκληρώνεται.

Ακολουθούν τα σύνολα οντοτήτων που θα συμπεριλάβουμε στη βάση:

- **Campaign:** Πρόκειται για τα topics στα οποία γίνονται τα campaigns. Κρατάμε ως attributes το ID, που είναι και πρωτεύον κλειδί, το όνομα του campaign και το runs, δηλαδή το πόσες φορές έχει τρέξει campaign στο συγκεκριμένο topic. Το πεδίο runs ενημερώνεται κάθε φορά που ολοκληρώνεται ένα νέο campaign στο topic.
- **Keyword:** λέξεις κλειδιά οι οποίες χρησιμοποιήθηκαν για την αναζήτηση πάνω σε συγκεκριμένο campaign. Κρατάμε τη λέξη και το ID της (πρωτεύον κλειδί).
- **Run:** Εξαρτάται από το campaign και κρατάει κάποιες πληροφορίες για την κάθε φορά που έτρεξε ένα campaign. Συγκεκριμένα κρατάει την ημερομηνία έναρξης και λήξης του συγκεκριμένου run, επίσης κρατάει τον σειριακό αριθμό του run, που μαζί με το ID του campaign αποτελούν πρωτεύοντα κλειδιά.
- **User:** Πρόκειται για πίνακα με τους twitterers που έχει πιάσει η εφαρμογή και περιέχει γενικές πληροφορίες για αυτούς. Συγκεκριμένα, κρατάει το ID (πρωτεύον κλειδί) του χρήστη, το όνομα και το username του, ακόμα το location και το description του στο twitter, τον αριθμό followers και friends του, τον αριθμό tweets, καθώς επίσης και τη φωτογραφία του στο twitter. Όταν ένα νέο run πιάνει κάποιον user που είναι ήδη στη βάση ενημερώνει ξανά κάποια από τα πεδία του run (π.χ. τον αριθμό followers, tweets, friends).
- **Campaigned\_User:** Αδύναμη οντότητα που εξαρτάται από τον user και το run. Περιλαμβάνει συμπληρωματικές πληροφορίες για τον user μέσα σε ένα συγκεκριμένο run, που κυρίως έχουν να κάνουν με την απόδοση του στο run. Συγκεκριμένα, κρατάει τον αριθμό retweets, mentions στο run αλλά και το score του με τον αλγόριθμο TwitterRank στο συγκεκριμένο run.

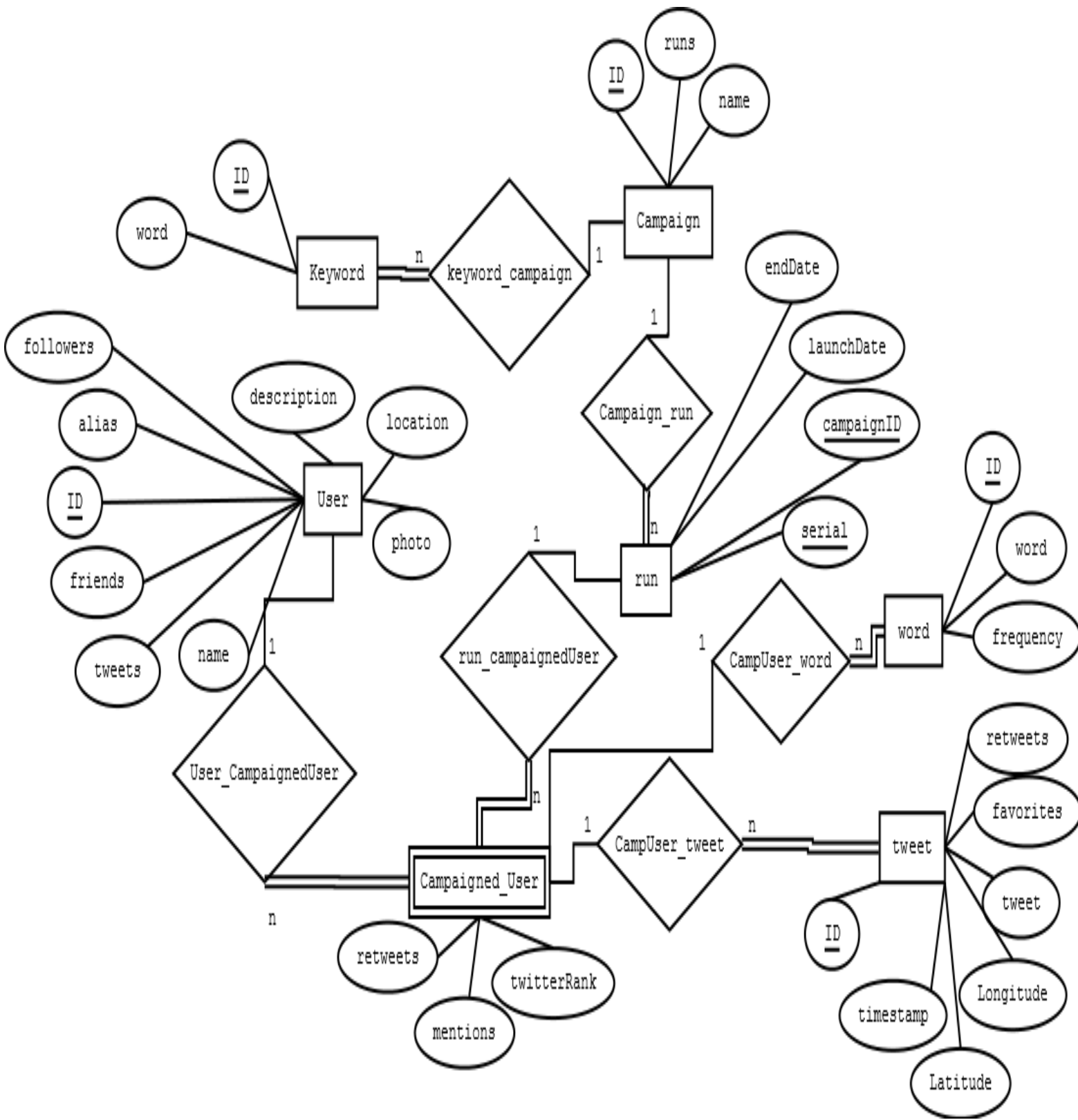
- **Tweet:** Πίνακας με τα tweets των twitterers. Κάθε tweet ανήκει σε συγκεκριμένο twitterer, σε συγκεκριμένο campaign και run. Περιλαμβάνει πληροφορίες για κάθε tweet. Πιο συγκεκριμένα, έχουμε ID το πρωτεύον κλειδί, timestamp δηλαδή την ώρα και ημερομηνία που ανέβηκε το tweet, Latitude και Longitude δηλαδή τα γεωχωρικά δεδομένα για την απεικόνιση του στο χάρτη. Ακόμη περιλαμβάνει το κείμενο του tweet, το πόσες φορές έγινε retweet και favorite.
- **Word:** Οι λέξεις που χρησιμοποιεί ο twitterer σε ένα run ενός campaign. Περιλαμβάνει το ID της λέξης (πρωτεύον κλειδί), τη λέξη και τη συχνότητα της. Ο συγκεκριμένος πίνακας χρησιμοποιείται για τη λειτουργία Thematic navigation στα tweets που περιγράφεται σε προηγούμενη ενότητα.

Παρακάτω αναλύουμε τις σχέσεις μεταξύ των οντοτήτων για το E-R diagram.

- **Keyword\_Campaign:** Σχέση μεταξύ keyword και campaign. Η σχέση είναι 1 προς N από την πλευρά του campaign. Αυτό, διότι ένα campaign περιγράφεται από πολλά keywords. Επιπλέον η σχέση είναι καθολική από την πλευρά του keyword, καθώς κάθε keyword αναφέρεται σε ένα campaign.
- **Campaign\_Run:** Σχέση μεταξύ Campaign και Run. Η σχέση είναι 1 προς N από την πλευρά του campaign. Αυτό διότι σε ένα campaign μπορούν να έχουν γίνει αρκετά runs. Επιπλέον η σχέση είναι καθολική από την πλευρά του run καθώς κάθε run αναφέρεται σε ένα campaign.
- **Run\_CampaignedUser:** Σχέση μεταξύ Run και Campaigned\_user. Η σχέση είναι 1 προς N από την πλευρά του run. Αυτό, διότι σε ένα run συμμετέχουν πολλοί campaigned\_users. Επιπλέον η σχέση είναι καθολική από την πλευρά του campaigned\_user, καθώς κάθε campaigned\_user έχει βρεθεί σε κάποιο run κάποιου campaign.

- **User\_CampaignedUser:** Σχέση μεταξύ User και Campaigned\_user. Η σχέση είναι 1 προς N από την πλευρά του user. Αυτό, διότι ένας user μπορεί να είναι ένας ή παραπάνω campaigned\_users. Για παράδειγμα, θα μπορούσε να είναι influential σε παραπάνω από ένα campaign. Επιπλέον η σχέση είναι καθολική από την πλευρά του campaigned\_user καθώς κάθε campaigned\_user έχει να κάνει με κάποιο συγκεκριμένο user.
- **CampUser\_word:** Σχέση μεταξύ Campaigned\_user και word. Η σχέση είναι 1 προς N από την πλευρά του Campaigned\_user. Αυτό διότι ένας campaigned\_User μπορεί να χρησιμοποιεί πολλές λέξεις (words). Επιπλέον η σχέση είναι καθολική από την πλευρά του word καθώς κάθε word ανήκει σε κάποιον συγκεκριμένο campaigned\_user.
- **CampUser\_tweet:** Σχέση μεταξύ Campaigned\_user και tweet. Η σχέση είναι 1 προς N από την πλευρά του campaigned\_user. Αυτό, διότι ένας campaigned\_user μπορεί δημοσιεύει ένα ή παραπάνω tweets. Επιπλέον η σχέση είναι καθολική από την πλευρά του tweet καθώς κάθε tweet έχει γραφτεί από κάποιον συγκεκριμένο campaigned\_user.

Ακολουθεί το E – R διάγραμμα που περιγράφηκε σε αυτή την ενότητα.



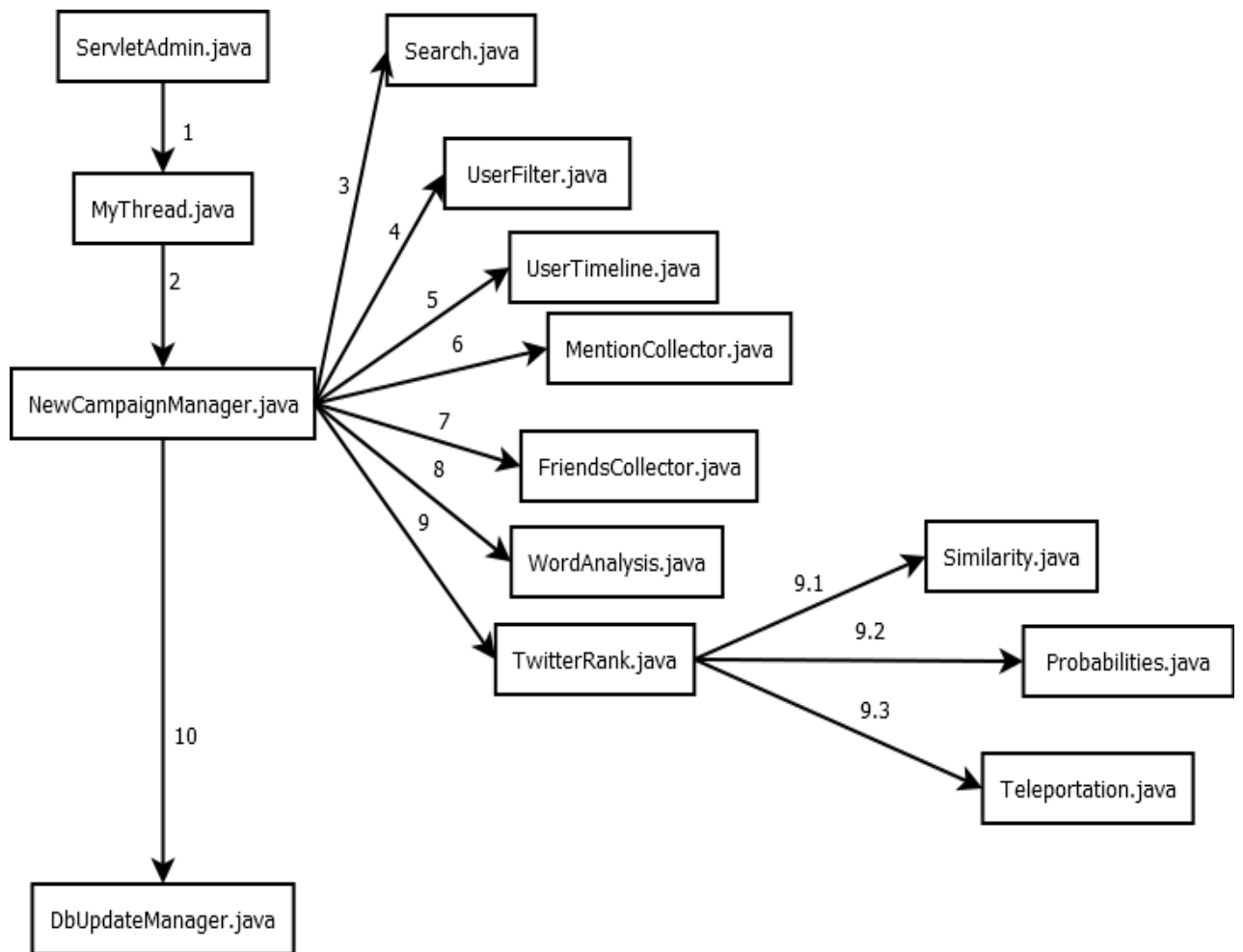
# 5

## *Σχεδίαση Συστήματος*

### *5.1 Αρχιτεκτονική*

Η αρχιτεκτονική του συστήματος για τον υπολογισμό ενός νέου campaign φαίνεται στο παρακάτω σχήμα. Συγκεκριμένα, η κλάση ServletAdmin (Servlet) καλεί ένα Thread της κλάσης MyThread ενώ το Thread δημιουργεί μέσω της μεθόδου run() ένα αντικείμενο της κλάσης NewCampaignManager. Αυτό το αντικείμενο καλεί κατά σειρά τους constructors των κλάσεων Search, UserFilter, UserTimeline, MentionCollector, FriendsCollector, WordAnalysis και TwitterRank. Η TwitterRank με τη σειρά της καλεί τους constructors των κλάσεων Similarity, Probabilities, και Teleportation. Μόλις το αντικείμενο της TwitterRank είναι διαθέσιμο ο NewCampaignManager καλεί την κλάση DbUpdateManager. Αξίζει να σημειωθεί ότι τα δεδομένα της Search χρειάζονται για τη δημιουργία ενός αντικειμένου της UserFilter, ενώ αυτά της UserFilter χρειάζονται μετέπειτα για όλες τις κλάσεις των οποίων καλείται ο constructor. Αντίστοιχα τα δεδομένα των WordAnalysis και FriendsCollector χρειάζονται στον υπολογισμό της TwitterRank. Τέλος τα δεδομένα των UserFilter, UserTimeline, MentionCollector, και TwitterRank χρειάζονται για την αρχικοποίηση του constructor της DbUpdateManager. Ενώ παρόμοια εξάρτηση συναντούμε και μεταξύ των Similarity και Probabilities.





## 5.2 Περιγραφή Βασικών Κλάσεων

Παρακάτω περιγράφονται οι βασικότερες κλάσεις της εφαρμογής για τον υπολογισμό του influence όπως έχει περιγραφεί στα προηγούμενα κεφάλαια. Δεν περιγράφονται κλάσεις που αποτελούν απλά ορισμούς και ούτε κλάσεις που εκτελούν κάποια μικρή λειτουργικότητα δευτερεύουσας σημασίας. Ακόμη οι κλάσεις που έχουν να κάνουν με τη λειτουργία της web-εφαρμογής παρουσιάζονται στο κεφάλαιο 6 (όπως οι κλάσεις ServletAdmin και MyThread).

### **5.2.1 *NewCampaignManager***

Η κλάση αυτή είναι υπεύθυνη για τη λειτουργία και το συντονισμό ενός νέου campaign. Αναλαμβάνει τη δημιουργία αντικειμένων των κλάσεων που χρειάζονται καθώς επίσης και τον έλεγχο, ότι τα αντικείμενα αυτά αρχικοποιήθηκαν σωστά. Αν δεν έχουν αρχικοποιηθεί σωστά ξανακαλεί τους constructors τους.

### **5.2.2 *DbUpdateManager***

Η συγκεκριμένη κλάση, αναλαμβάνει να καταχωρήσει στη βάση δεδομένων όλα τα δεδομένα που αντλήθηκαν ή υπολογίστηκαν για το συγκεκριμένο campaign. Για την ομαλή καταχώρηση στη βάση επίσης αναλαμβάνει να καθαρίσει τα δεδομένα τύπου String (όπως τα tweets, και τα words) από ειδικούς χαρακτήρες που θα μπορούσαν να προκαλέσουν exception με την προσπάθεια εισαγωγής στη βάση. Τέλος, εφ'όσον τα exceptions είναι απίθανο να αποφευχθούν σε τόσο πολλά δεδομένα, η κλάση αυτή αναλαμβάνει και το exception handling, ώστε η λειτουργία του προγράμματος να συνεχίζεται ομαλά.

### **5.2.3 *Search***

Η κλάση Search χρησιμοποιεί το Twitter API για να ψάξει το αρχικό σώμα από tweets με τα keywords που δίνονται. Περιλαμβάνει τα πεδία tweets και tweetsCount και getters για αυτά, καθώς είναι private, ενώ τα tweets αντλούνται στον constructor της κλάσης και δεν θέλουμε setters. Τέλος, επειδή το Twitter API επιτρέπει συγκεκριμένο αριθμό requests στο τέταρτο και αν τον υπερβούμε δημιουργεί exceptions, η Search αναλαμβάνει το time και το exception handling. Για αυτό, δημιουργήθηκαν και οι μέθοδοι makeSearchRequest και makeLimitRequest, οι οποίες χρησιμοποιούνται για το exception handling και ξανακαλούνται όσο δεν επιστρέφουν σωστά αποτελέσματα.

### **5.2.4 *UserFilter***

Η κλάση αυτή παίρνει τα αποτελέσματα του search και εξάγει τους twitterers στους οποίους ανήκουν τα Statuses του search, τους ταξινομεί με βάση τους followers και τελικά επιλέγει

τους τελικούς influentials. Περιλαμβάνει τα στοιχεία FUserList και τη σταθερά FINAL\_USERS, καθώς επίσης και τις μεθόδους findFinalUsers, getUsers, removeDoubleStatuses και επίσης περιλαμβάνει και τους getters getFUserList και getFINAL\_USERS. Η FUserList είναι η λίστα με τους τελικούς influentials. Η σταθερά FINAL\_USERS είναι ο αριθμός των τελικών influentials. Αυτή τη στιγμή έχει οριστεί ίσος με 50. Ωστόσο σε μεγάλα υπολογιστικά συστήματα θα μπορούσε να είναι και 100 ή 200. Όσο περισσότεροι είναι οι τελικοί influentials τόσο πιο αξιόπιστα αποτελέσματα θα παρέχει ο αλγόριθμος TwitterRank. Η μέθοδος removeDoubleStatuses προσπελαύνει τα tweets του search και επιστρέφει μία λίστα από tweets που όλα ανήκουν σε διαφορετικούς twitterers. Η μέθοδος getUsers εξάγει τους twitterers από τα tweets του search, ενώ τελικά η findFinalUsers υπολογίζει τους τελικούς influentials και τους επιστρέφει ταξινομημένους με βάση τον αριθμό των followers τους.

### **5.2.5 UserTimeline**

Η κλάση UserTimeline χρησιμοποιεί το Twitter API για να πάρει τα Timelines από όλους τους τελικούς influentials μέχρι την ημερομηνία που έχει δώσει ο χρήστης της εφαρμογής. Περιλαμβάνει το πεδίο userTweetList και τον αντίστοιχο getter για αυτό, καθώς είναι private. Επίσης η κλάση αυτή αναλαμβάνει το exception handling και το time management. Για αυτό, δημιουργήθηκαν και οι μέθοδοι makeTimelineRequest και makeLimitRequest, οι οποίες χρησιμοποιούνται για το exception handling και ξανακαλούνται όσο δεν επιστρέφουν σωστά αποτελέσματα.

### **5.2.6 MentionCollector**

Η κλάση MentionCollector χρησιμοποιεί το Twitter API για να πάρει τα mentions και τα retweets που κερδίζουν όλοι οι τελικοί influentials για το χρονικό διάστημα που έχει δώσει ο χρήστης της εφαρμογής, ακόμη είναι υπεύθυνη για την κατάταξη αυτών σε mentions και retweets. Περιλαμβάνει την ιδιότητα mentionedUserList και τον αντίστοιχο getter για αυτή, καθώς είναι private. Επίσης περιλαμβάνει τις μεθόδους distributeTweets, makeRateLimitsRequest, makeSearchRequest. Η distributeTweets αναλαμβάνει να κατατάξει τα mentions ενός twitterer σε retweets ή mentions. Τέλος οι μέθοδοι makeRateLimitsRequest, makeSearchRequest χρησιμοποιούνται για το exception handling και ξανακαλούνται όσο δεν

επιστρέφουν σωστά αποτελέσματα (συμβαίνει exception). Το time handling γίνεται στον constructor της κλάσης.

### 5.2.7 *FriendsCollector*

Η κλάση FriendsCollector χρησιμοποιεί το Twitter API για να πάρει όλους τους friends όλων των τελικών influentials. Περιλαμβάνει το πεδίο userWithFollowingIDs και τον αντίστοιχο getter για αυτό, καθώς είναι private. Επίσης, περιλαμβάνει τις μεθόδους followingUsers, makeRateLimitsRequest, makeFriendRequest. Η followingUsers παίρνει ως όρισμα τους φίλους κάποιου influential και υπολογίζει ποιοι από αυτούς είναι στους τελικούς influentials. Τέλος οι μέθοδοι makeRateLimitsRequest, makeFriendsRequest χρησιμοποιούνται για το exception handling και ξανακαλούνται όσο δεν επιστρέφουν σωστά αποτελέσματα (συμβαίνει exception). Το time handling γίνεται στον constructor της κλάσης.

### 5.2.8 *WordAnalysis*

Στην κλάση WordAnalysis γίνεται η ανάλυση σε λέξεις των αρχικών tweets του campaign (που προέκυψαν από το search), καθώς επίσης και η ανάλυση σε λέξεις των tweets των χρηστών (πάνω στο Topic και γενικότερα) και απορρίπτονται όλες οι λέξεις με λιγότερα των 4 γραμμάτων, όπως και αυτές που περιλαμβάνουν χαρακτήρες που δεν ανήκουν στο αγγλικό αλφάβητο. Περιλαμβάνει τις ιδιότητες DTsearch, DTuser, DTtopicUser και getters για αυτές. Το DTsearch είναι η λίστα με τις λέξεις των tweets του search και τις συχνότητες τους αντίστοιχα. Παρόμοια, το DTuser είναι η λίστα με τις λέξεις των tweets του twitterer και τις συχνότητες τους. Το DTtopicUser είναι η λίστα με τις λέξεις πάνω στο topic του campaign των tweets του twitterer και τις συχνότητες τους. Περιλαμβάνει τις μεθόδους isAlpha που επιστρέφει true αν μία λέξη περιέχει μόνο γράμματα της αγγλικής αλφαβήτου, αλλιώς false, ακόμη τις countAppearedWords και countAppearedTopicWords που μετράνε τις λέξεις ενός twitterer γενικά και πάνω στο topic αντίστοιχα.

### 5.2.9 *TwitterRank*

Η κλάση `TwitterRank` είναι υπεύθυνη για την εκτέλεση του αλγορίθμου `TwitterRank`. Περιλαμβάνει τα πεδία `twitterRank`, `twitterRankNorm` και τη σταθερά `gamma`, επίσης περιλαμβάνει getters για αυτά. Ο πίνακας `twitterRank` κρατάει το score του κάθε influential με το συγκεκριμένο αλγόριθμο ενώ ο `twitterRankNorm` είναι ο πίνακας που προκύπτει κανονικοποιώντας τον `twitterRank`. Η σταθερά `gamma` είναι το  $\gamma$  του τύπου υπολογισμού του `TwitterRank` που δίνεται στην ενότητα 4.2.2 και το ορίσαμε ίσο με 0.5.

### 5.2.10 *Similarity*

Η κλάση `Similarity` βρίσκει το `similarity` μεταξύ όλων των `twitterers` οι οποίοι συνδέονται με μια σχέση `follow`, όπως περιγράφεται στην ενότητα 4.2.2. Στους υπόλοιπους θεωρείται ίσο με μηδέν. Έχει ως ιδιότητα τον δισδιάστατο πίνακα `similarity` και τον αντίστοιχο `getter`. Επίσης, υπάρχει η μέθοδος `findSimilarity` που υπολογίζει το `similarity` μεταξύ δυο `twitterers`.

### 5.2.11 *Probabilities*

Υπολογίζει την πιθανότητα  $P^i(i, j)$ , όπως αυτή περιγράφηκε στην ενότητα 4.2.2. Έχει ως ιδιότητα τον δισδιάστατο πίνακα `probability` που είναι ο  $P^i(i, j)$  και τον αντίστοιχο `getter`. Επίσης περιλαμβάνει τη μέθοδο `computePo` που υπολογίζει την πιθανότητα ένας `twitterer` να επηρεάζεται από κάποιον άλλο. Δηλαδή υπολογίζει ένα στοιχείο του πίνακα `probability` κάθε φορά που καλείται.

### 5.2.12 *Teleportation*

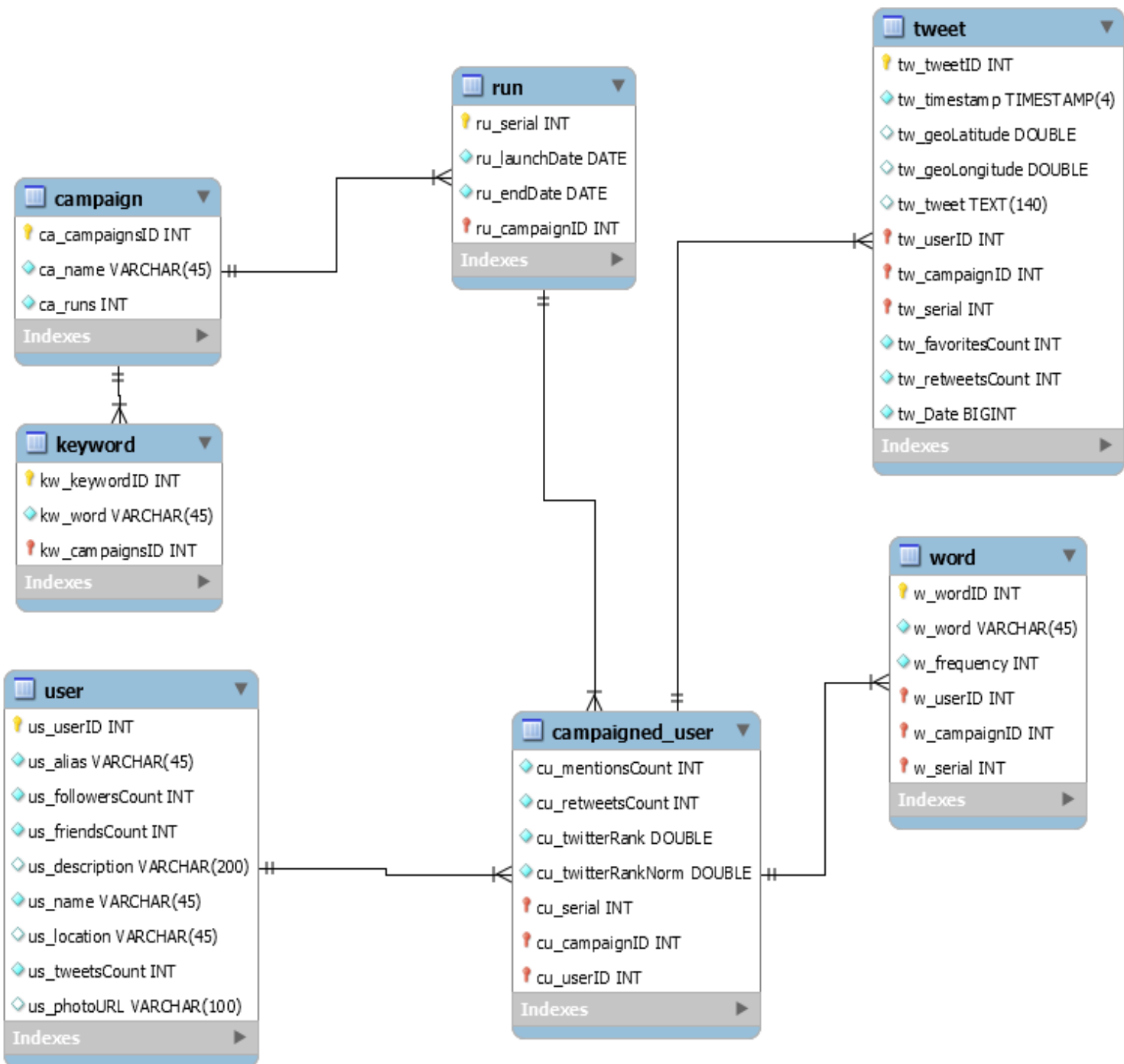
Υπολογίζει το `Teleportation vector`  $E_t$  όπως αυτό περιγράφηκε στην ενότητα 4.2.2. Περιλαμβάνει ως ιδιότητα τον μονοδιάστατο πίνακα `teleportation` που είναι ο πίνακας  $E_t$  και `getter` για αυτόν. Επίσης περιλαμβάνει τη μέθοδο `findTeleportation` που κάθε φορά που καλείται υπολογίζει ένα στοιχείο του πίνακα `teleportation`.

### 5.3 Βάση Δεδομένων

Εδώ δίνεται το σχεσιακό σχήμα της βάσης δεδομένων μας, όπως αυτό προκύπτει από το μοντέλο οντοτήτων-συσχετίσεων που δόθηκε παραπάνω. Να σημειωθεί ότι οι συσχετίσεις έχουν παραληφτεί καθώς αποτελούν συσχετίσεις ένα προς πολλά. Θα γίνονται εμφανείς μέσω ξένων κλειδιών.

Οι πίνακες της βάσης είναι:

- **Keyword** (kw\_keywordID, kw\_word, kw\_campaignsID )
- **Campaign** (ca\_campaignsID, ca\_name, ca\_runs)
- **User** (us\_userID, us\_alias, us\_followersCount, us\_friendsCount, us\_description, us\_name, us\_location, us\_tweetsCount, us\_photoURL)
- **Run** (ru\_serial, ru\_campaignID, ru\_launchDate, ru\_endDate)
- **Campaigned\_user** (cu\_serial, cu\_campaignID, cu\_userID, cu\_mentionsCount, cu\_retweetCount, cu\_twitterRank, cu\_twitterRankNorm)
- **Tweet** (tw\_tweetID, tw\_serial, tw\_campaignID, tw\_userID, tw\_timestamp, tw\_geoLatitude, tw\_geoLongitude, tw\_tweet, tw\_favoritesCount, tw\_retweetsCount, tw\_Date)
- **Word** (w\_wordID, w\_serial, w\_campaignID, w\_userID, w\_word, w\_frequency)



# 6

## *Ανάπτυξη web εφαρμογής*

Πολύ σημαντικό μέρος αυτής της εργασίας ήταν η υλοποίηση μιας web- εφαρμογής που να οπτικοποιεί τα αποτελέσματα ενός campaign, να παρουσιάζει συνοπτικά τα campaigns αλλά και να δίνει τη δυνατότητα στον τελικό χρήστη της να έχει μια πλήρη εικόνα για τους influential twitterers και όχι απλά την επίδοσή τους. Επίσης, σημαντικό για τον τελικό χρήστη της εφαρμογής είναι και η έναρξη ενός νέου campaign, όπως και η ανανέωση κάποιου παλαιότερου.

### *6.1 Περιγραφή βασικών jsp, servlets και κλάσεων web-εφαρμογής*

Ακολουθεί περιγραφή των βασικότερων σελίδων και κλάσεων της web-εφαρμογής. Αξίζει να σημειωθεί ότι οι λειτουργίες New Campaign, σύνδεσμος για μετάβαση στην αρχική σελίδα (σύνδεσμοι Campaign List και TwiInfluence), σύνδεσμος για μετάβαση στους κορυφαίους twitterers (Top Twitterers) από όλα τα campaigns και μπάρα με κουμπί Search για αναζήτηση campaign παρέχονται από όλες τις σελίδες.



### **6.1.1 Home.jsp**

Είναι η αρχική σελίδα της εφαρμογής. Εδώ ο χρήστης έχει τη δυνατότητα να δει συνοπτικά τα πιο πρόσφατα campaigns (μπορεί να πάει και σε πιο παλιά μέσω του συνδέσμου "more campaigns"), με κάποιες σημαντικές πληροφορίες για αυτά, όπως πότε ήταν η τελευταία φορά που ανανεώθηκε, πόσους twitterers και πόσα tweets έχει. Επίσης, ο χρήστης έχει τη δυνατότητα να μεταβεί σε ένα campaign (κάνοντας κλικ στο όνομα του campaign) που τον ενδιαφέρει ή και να το ανανεώσει μέσω του κουμπιού "Relaunch this campaign".

### **6.1.2 campaign.jsp**

Είναι η σελίδα στην οποία μεταβαίνει ο χρήστης όταν επιλέγει ένα campaign. Η σελίδα αυτή δείχνει γραφικά τα Top 10 των influentials για κάθε μια από τις τέσσερις μετρικές που αναπτύχθηκαν. Επίσης, υπάρχει η δυνατότητα για αναπαράσταση σε λίστα του αποτελέσματος για τους influentials (κάτω από κάθε γραφική παράσταση υπάρχει ο σύνδεσμος "show as list") ταξινομημένων βάσει της επίδοσης στο μέτρο του οποίου τον σύνδεσμο πάτησε ο χρήστης. Τέλος, παρέχεται tooltip που λέει στον χρήστη το αποτέλεσμα για τον twitterer από του οποίου την μπάρα πέρασε το ποντίκι, ενώ κάνοντας κλικ πάνω στη μπάρα κάποιου twitterer ανοίγει νέο tab παρουσιάζοντας τον συγκεκριμένο twitterer.

### **6.1.3 userList.jsp**

Είναι η σελίδα στην οποία μεταβαίνει ο χρήστης όταν επιλέγει "show as list" στη σελίδα campaign.jsp. Η σελίδα αυτή να μεν παρουσιάζει τα αποτελέσματα του campaign.jsp, αλλά η συγκεκριμένη προσανατολίζεται περισσότερο στην παρουσίαση twitterers ώστε ο χρήστης να επιλέξει ποιοι τον ενδιαφέρουν και λιγότερο στην επίδοσή τους στο campaign (μόνο η ταξινόμηση προδίδει την επίδοσή τους). Ο χρήστης πατώντας κλικ στο όνομα ή στη φωτογραφία κάποιου twitterer της λίστας μεταβαίνει σε σελίδα παρουσίασης του συγκεκριμένου twitterer. Ακόμη, ο χρήστης της εφαρμογής μπορεί να δει περισσότερους twitterers στη λίστα μέσω του συνδέσμου "more users" στο κάτω μέρος της σελίδας. Τέλος παρέχεται μέσω των συνδέσμων στο δεξί μέρος της σελίδας η δυνατότητα ταξινόμησης βάσει επίδοσης σε άλλους αλγορίθμους του campaign.

#### **6.1.4 *user.jsp***

Η συγκεκριμένη σελίδα είναι η σελίδα μετάβασης όταν κάνει ο χρήστης κλικ είτε στο όνομα ή στη φωτογραφία του twitterer από τη σελίδα `userList.jsp` είτε πατώντας κλικ στη μπάρα κάποιου twitterer στη σελίδα `campaign.jsp`. Εδώ παρουσιάζονται χαρακτηριστικά του twitterer τόσο μέσα στο campaign (επίδοση και ranking για κάθε έναν αλγόριθμο) όσο και γενικότερα χαρακτηριστικά για την παρουσία του στο twitter, όπως αριθμός followers, friends και tweets, περιγραφή τοποθεσία, φωτογραφία κλπ. Επίσης, παρέχεται η δυνατότητα μετάβασης στο campaign, όπως και μετάβασης στο προφιλ του στο twitter. Στο δεύτερο μισό της σελίδας έχουμε τα tabs “Favorite Tweets”, “Thematic Navigation”, “Tweets Map” και “Tweets over Time”. Το “Favorite Tweets” είναι το ενεργό tab αυτής της σελίδας και παρουσιάζει τα πιο αγαπημένα tweets του twitterer (most favorited tweets) για τα οποία παρέχεται ο αριθμός που έχουν επισημανθεί ως favorite, ο αριθμός που έχουν γίνει retweets αλλά και πριν πόσο καιρό έγινε το tweet.

#### **6.1.5 *userThem.jsp***

Το πρώτο μισό της σελίδας είναι ίδιο με αυτό της σελίδας `user.jsp`. Στο δεύτερο μισό υλοποιεί το tab “Thematic Navigation”, που είναι και το ενεργό tab. Συγκεκριμένα επιτρέπει τη θεματική πλοήγηση στα tweets του συγκεκριμένου twitterer. Παρέχει ένα word cloud με τις λέξεις που κυρίως χρησιμοποιεί ο Twitterer. Για κάθε κλικ πάνω σε κάποια λέξη ανοίγει ένα νέο tab που παρουσιάζει τα tweets που περιέχουν το συγκεκριμένο keyword ταξινομημένα με βάση το πόσες φορές επισημάνθηκαν ως favorite.

#### **6.1.6 *userMap.jsp***

Ομοίως, το πρώτο μισό της σελίδας είναι ίδιο με αυτό της σελίδας `user.jsp`. Η σελίδα αυτή υλοποιεί το tab “Tweets Map”. Παρέχει ένα χάρτη και απεικονίζει τα tweets που περιέχουν GeoLocation tag πάνω στο χάρτη. Ο χρήστης κάνοντας κλικ σε κάποιο από τα σημεία με το μπλε marker μπορεί να δει το tweet που βρίσκεται σε εκείνο το σημείο. Επίσης, ο χάρτης παρέχει δυνατότητες zoom και unzoom.

### **6.1.7 *userTime.jsp***

Ομοίως, το πρώτο μισό της σελίδας είναι ίδιο με αυτό της σελίδας `user.jsp`. Η σελίδα αυτή υλοποιεί το tab “Tweets over Time”. Παρέχει μια γραφική παράσταση με σημεία, όπου στον x-άξονα απεικονίζεται λεπτομερώς ο χρόνος, ενώ στον y-άξονα το πόσες φορές επισημάνθηκε ένα tweet ως favorite. Τα σημεία της γραφικής παράστασης είναι τα tweets του `Twitterer`. Ο χρήστης περνώντας το ποντίκι πάνω από κάποιο σημείο μπορεί να δει το συγκεκριμένο tweet. Επίσης, παρέχονται δυνατότητες zoom σε συγκεκριμένο διάστημα χρόνου επιλέγοντας στη γραφική παράσταση το επιθυμητό διάστημα, ενώ unzoom μπορεί να κάνει ο χρήστης από το overview διάγραμμα που βρίσκεται από κάτω. Τέλος, αξίζει να σημειωθεί ότι στη γραφική παράσταση τα σαββατοκύριακα σημειώνονται με γκρι χρώμα ενώ οι καθημερινές με πράσινο.

### **6.1.8 *topList.jsp***

Έχει την ίδια σχεδόν λειτουργικότητα και εμφάνιση με τη σελίδα `userList.jsp`. Η μόνη διαφορά είναι ότι δίπλα στο όνομα και την τοπθεσία κάθε `twitterer` εμφανίζονται και όλα τα campaigns στα οποία είναι influential και κάνοντας κλικ σε κάποιο από αυτά η εφαρμογή ανακατευθύνει το χρήστη στο αντίστοιχο campaign.

### **6.1.9 *top.jsp***

Παρουσιάζει την ίδια λειτουργικότητα με τη σελίδα `user.jsp`, μόνο που εδώ δεν αναφερόμαστε σε συγκεκριμένο campaign, συνεπώς το κομμάτι της επίδοσης μέσα στο campaign απουσιάζει σε σχέση με τη σελίδα `user.jsp`. Η άλλη διαφορά είναι ότι στη σελίδα αυτή παρουσιάζονται όλα τα campaigns που έχει συμμετάσχει ο `twitterer` και σύνδεσμοι σε αυτά. Τέλος στη συγκεκριμένη σελίδα απουσιάζει το tab “Tweets over Time” που υπάρχει στην σελίδα `user.jsp`.

### **6.1.10 *topThem.jsp***

Το πρώτο μισό της σελίδας είναι ίδιο με αυτό της σελίδας `top.jsp`, ενώ το δεύτερο μισό δεν διαφέρει σε τίποτε από το δεύτερο μισό της σελίδας `userThem.jsp`.

### **6.1.11 topMap.jsp**

Το πρώτο μισό της σελίδας είναι ίδιο με αυτό της σελίδας top.jsp, ενώ το δεύτερο μισό δεν διαφέρει σε τίποτε από το δεύτερο μισό της σελίδας userMap.jsp.

### **6.1.12 campaignStarted.jsp**

Πρόκειται για τη σελίδα που εμφανίζεται όταν ο χρήστης επιχειρεί να ξεκινήσει ένα καινούριο campaign. Αν τα δεδομένα που έχουν συμπληρωθεί στη φόρμα είναι σωστά εμφανίζεται μήνυμα ότι το campaign ξεκίνησε και σύντομα θα υπάρξουν αποτελέσματα. Επίσης παρέχεται σύνδεσμος επιστροφής στην αρχική σελίδα για να συνεχίσει ο χρήστης την πλοήγηση του στο web site. Αν τα δεδομένα δεν είναι σωστά (π.χ. δεν έχει συμπληρωθεί το όνομα του campaign) τότε εμφανίζεται μήνυμα λάθους στη σελίδα όπως επίσης και σύνδεσμος μετάβασης στην αρχική σελίδα.

### **6.1.13 ServletAdmin.java**

Το servlet αυτό καλείται όταν ο χρήστης της εφαρμογής κάνει submit την φόρμα για το New Campaign και αναλαμβάνει τον έλεγχο αν είναι εντάξει τα στοιχεία που δόθηκαν στη φόρμα ώστε να ξεκινήσει το καινούριο campaign. Αν τα στοιχεία δεν είναι σωστά κάνει αναμετάβαση στη σελίδα campaignStarted.jsp παραμετροποιώντας την κατάλληλα, ώστε να εμφανίσει μήνυμα για λάθος δεδομένα στη φόρμα. Αν πάλι τα στοιχεία είναι σωστά τότε υπολογίζει τις ημερομηνίες στις οποίες θα τρέξει το νέο campaign και στη συνέχεια καλεί την κλάση MyThread.java, παραμετροποιώντας την κατάλληλα. Τέλος κάνει ανακατεύθυνση στη σελίδα campaignStarted.jsp, με κατάλληλη παραμετροποίηση ώστε να εμφανίζει το μήνυμα ότι ξεκίνησε κανονικά το καινούριο campaign.

### **6.1.14 MyThread.java**

Η κλάση MyThread κάνει extend την κλάση Thread. Πρόκειται για thread που τρέχει παράλληλα με την εφαρμογή και είναι υπεύθυνο για τη δημιουργία του νέου campaign. Συγκεκριμένα, καλεί τον constructor της κλάσης NewCampaignManager για την οποία

έχουμε γράψει σε προηγούμενο κεφάλαιο. Όταν τελειώσει ο NewCampaignManager, τα αποτελέσματα του νέου campaign εμφανίζονται κι αυτά στη σελίδα με τα άλλα campaigns.

## **6.2 Πλατφόρμες και προγραμματιστικά εργαλεία**

Οι τεχνολογίες που χρησιμοποιήθηκαν για την υλοποίηση της εφαρμογής μας στο client side κομμάτι είναι οι γλώσσες HTML, CSS και JavaScript στην οποία χρησιμοποιήθηκαν οι βιβλιοθήκες JQuery.js, flot.js, d3.js και Leaflet.js, ενώ για το server side κομμάτι χρησιμοποιήθηκε η γλώσσα Java και συγκεκριμένα οι τεχνολογίες Servlet και Java Server Pages (JSP), και ο web server Apache Tomcat. Ενώ, η εφαρμογή χρησιμοποιεί τη σχεσιακή βάση δεδομένων MySQL.

### **6.2.1 HTML, CSS, JavaScript**

Οι τρεις αυτές γλώσσες παρουσιάζονται μαζί καθότι είναι αναπόσπαστα κομμάτια. Πρόκειται για τις γλώσσες που χρησιμοποιεί ο κάθε web browser και κάθεμια από αυτές επιτελεί ένα συγκεκριμένο είδος λειτουργιών στις ιστοσελίδες. Η γλώσσα HTML (HyperText Markup Language) είναι η κύρια γλώσσα σήμανσης στις ιστοσελίδες. Χρησιμοποιείται για την αναπαράσταση περιεχομένου και έχει αρκετές δυνατότητες όπως αναπαράσταση ήχου και βίντεο. Αντίθετα, η γλώσσα CSS (Cascading Style Sheets) ανήκει στην κατηγορία των γλωσσών φύλλων στυλ που χρησιμοποιείται για τον έλεγχο της εμφάνισης ενός εγγράφου που έχει γραφτεί σε γλώσσα HTML (και σε άλλες όπως XML), μέσω της CSS μπορούμε να επηρεάζουμε το χρώμα, το στυλ, το μέγεθος των στοιχείων μιας ιστοσελίδας. Τέλος, η JavaScript είναι μια light weight γλώσσα σεναρίων (scripting language) που υποστηρίζεται από όλους τους browsers και χρησιμοποιείται για να ελέγξει τη συμπεριφορά των ιστοσελίδων. Για παράδειγμα, μέσω της JavaScript μπορούμε να αλλάζουμε δυναμικά το περιεχόμενο μιας ιστοσελίδας ή την εμφάνιση της χωρίς να επιβαρύνουμε τον Server.

### 6.2.2 *JQuery.js*

Η JQuery είναι μία βιβλιοθήκη της JavaScript, η οποία απλοποιεί σημαντικά τον client-side προγραμματισμό. Υποστηρίζει όλους τους σύγχρονους web-browsers όπως και το HTML DOM (Document Object Model) που βοηθάει ιδιαίτερα στο δυναμικό περιεχόμενο ιστοσελίδων.

### 6.2.3 *flot.js, d3.js, Leaflet.js*

Η jquery.flot είναι μια βιβλιοθήκη της JavaScript για τη δημιουργία γραφικών παραστάσεων. Προσφέρει αρκετές επιλογές σε διαγράμματα γραφικά και άλλες λειτουργίες. Χρησιμοποιήθηκε για τις γραφικές παραστάσεις της σελίδας campaign.jsp καθώς επίσης και για τη γραφική παράσταση στη σελίδα userTime.jsp. Όσον αφορά την d3.js, είναι μια βιβλιοθήκη της JavaScript για το χειρισμό HTML αρχείων που βασίζονται σε δεδομένα. Στόχος της συγκεκριμένης βιβλιοθήκης είναι να οπτικοποιεί τα δεδομένα και να τα κάνει να φαίνονται πιο ζωντανά. Συνδυάζεται άψογα με το DOM. Αυτή η βιβλιοθήκη χρησιμοποιήθηκε για την υλοποίηση των word clouds στις σελίδες userThem.jsp και topThem.jsp. Τέλος, η Leaflet.js είναι μία ανοιχτού κώδικα βιβλιοθήκη της JavaScript που χρησιμοποιείται για εφαρμογές με διαδραστικούς χάρτες. Στην εφαρμογή μας, χρησιμοποιήθηκε στις σελίδες userMap.jsp και topMap.jsp.

### 6.2.4 *JAVA Enterprise Edition*

Πρόκειται για μια από τις πιο διαδεδομένες γλώσσες προγραμματισμού γενικού σκοπού. Είναι αντικειμενοστρεφής, υψηλού επιπέδου και με αυστηρό σύστημα τύπων, πράγμα που την κάνει ασφαλή. Επίσης είναι ανεξάρτητη πλατφόρμας, που σημαίνει ότι μπορεί να εκτελείται σε πολλές συσκευές χωρίς να χρειάζονται αλλαγές στον κώδικα. Τέλος, υπάρχουν πάρα πολλές βιβλιοθήκες για έναν τεράστιο αριθμό λειτουργιών, πράγμα που την κάνει εξαιρετική επιλογή για κάποιον προγραμματιστή, μιας και αρκετές λειτουργίες παρέχονται έτοιμες μέσα από βιβλιοθήκες. Στην εφαρμογή μας όλος ο κώδικας για το NewCampaign και τον υπολογισμό του influence γράφτηκε σε JAVA.

### **6.2.5 Servlet, JSP**

Τα servlets είναι κλάσεις της JAVA που τρέχουν σε WWW servers και δημιουργούν δυναμικό περιεχόμενο (συνήθως html). Χρησιμοποιούνται κυρίως σε εφαρμογές που απαιτούν προεπεξεργασία πριν την αποστολή απόκρισης στο χρήστη. Ως JAVA classes, κληρονομούν πλήρως όλα τα πλεονεκτήματα που προσφέρει η JAVA. Πιο συγκεκριμένα, είναι platform independent, χρησιμοποιούν τις βιβλιοθήκες της JAVA, προσφέρουν ασφάλεια κώδικα κλπ.

Όσον αφορά τις JSP (Java Server Pages), είναι σελίδες HTML που περιέχουν ειδικές ετικέτες με κώδικα JAVA. Χρησιμοποιούνται για τον ίδιο λόγο που χρησιμοποιούνται και τα servlets και παρέχουν σχεδόν την ίδια λειτουργικότητα καθώς ένα JSP αρχείο κατά τη μεταγλώτιση του μετατρέπεται σε servlet.

### **6.2.6 Apache Tomcat**

Ο Apache Tomcat είναι ένας web server ανοιχτού κώδικα. Πρόκειται για έναν servlet container που ακούει στο port 8080. Υποστηρίζει αρκετή λειτουργικότητα της JAVA EE συμπεριλαμβανομένων των servlets, των JSP σελίδων, των JAVA EL και των Web Sockets. Τέλος, υποστηρίζει και απλές στατικές σελίδες.

### **6.2.7 MySql**

Η MySQL είναι ένα σύστημα διαχείρισης σχεσιακών βάσεων δεδομένων που μετρά περισσότερες από 11 εκατομμύρια εγκαταστάσεις. Το πρόγραμμα τρέχει έναν εξυπηρετητή (server) παρέχοντας πρόσβαση πολλών χρηστών σε ένα σύνολο από βάσεις δεδομένων. Είναι ανοιχτού κώδικα και είναι ίσως το πιο διαδεδομένο σύστημα διαχείρισης σχεσιακών βάσεων δεδομένων στον κόσμο.

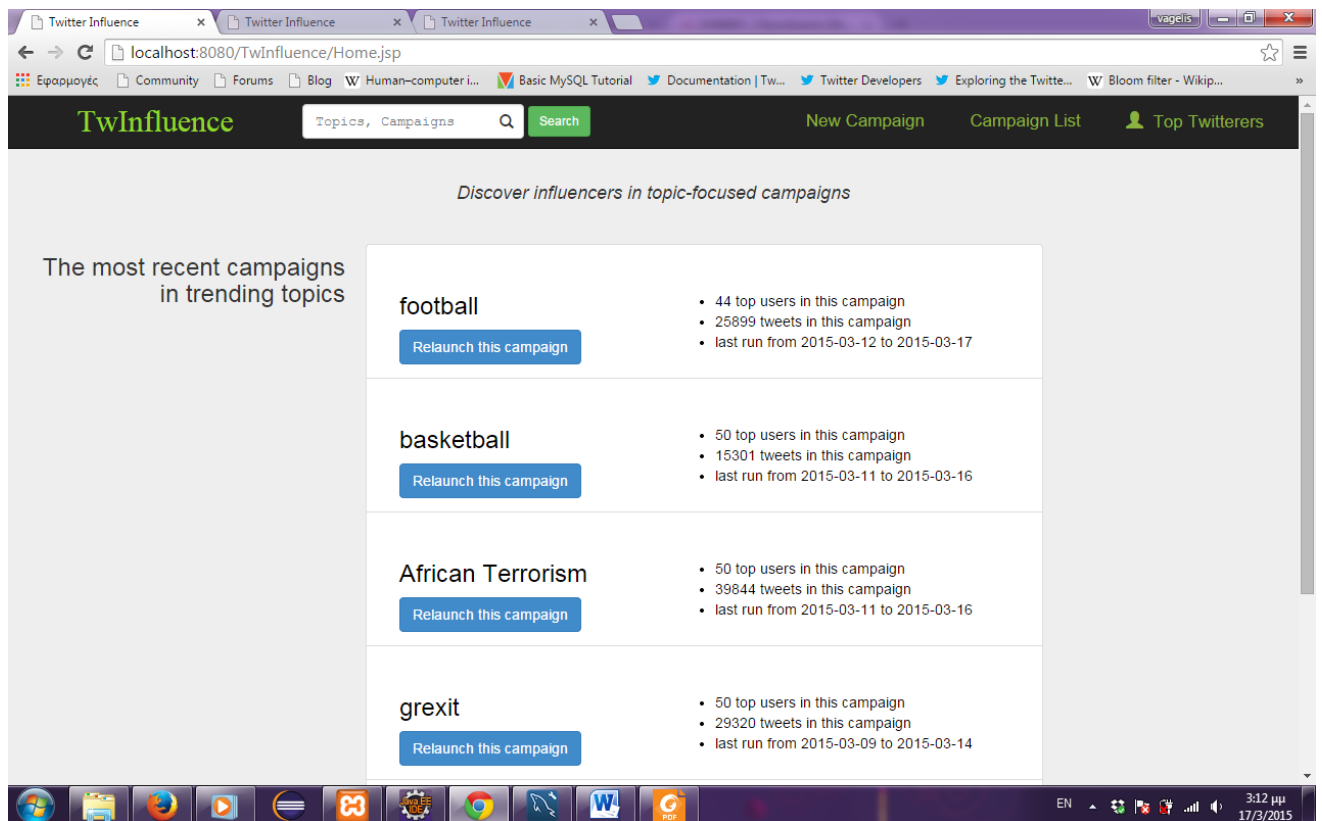
# 7

## *Παράδειγμα χρήσης web-based εφαρμογής*

Ας υποθέσουμε ότι ένας χρήστης του Twitter ενδιαφέρεται να βρει τους πιο influential twitterers πάνω στο θέμα του “grexit” δηλαδή της εξόδου της Ελλάδας από το ευρώ και ότι θα χρησιμοποιήσει την εφαρμογή μας για το σκοπό αυτό. Η διαδικασία έχει ως εξής:

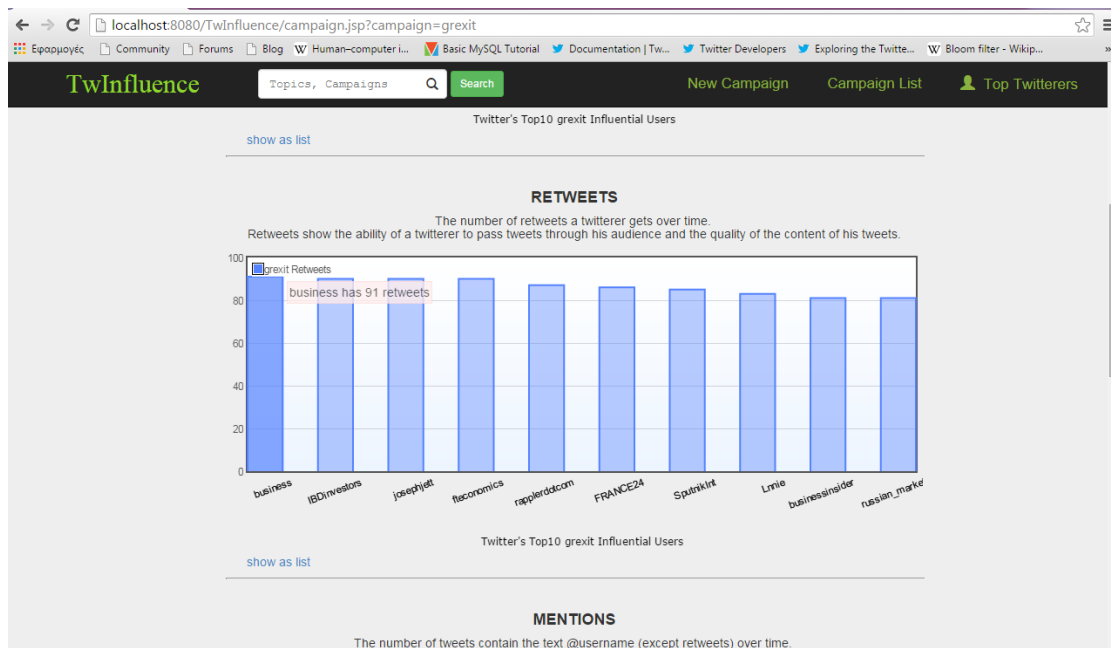
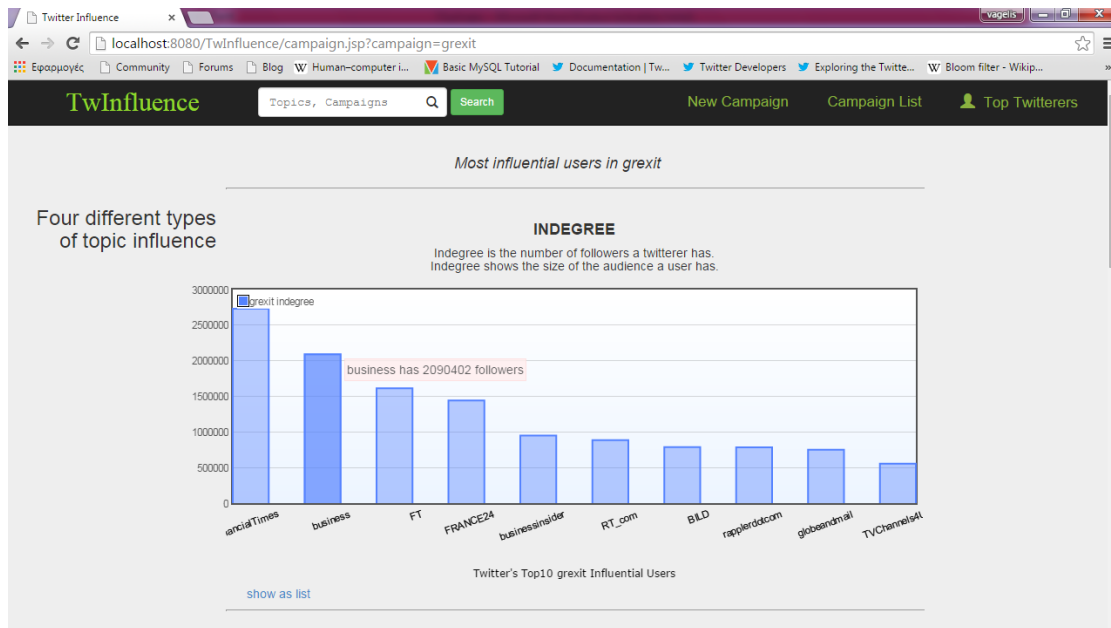
Πρώτον, εισέρχεται στην αρχική σελίδα, όπου έχει τη δυνατότητα να περιηγηθεί σε διάφορα campaigns ή να κάνει αναζήτηση κατευθείαν σε αυτό που ψάχνει. Στην επόμενη εικόνα φαίνεται η αρχική σελίδα της εφαρμογής μας (σελίδα Home.jsp).

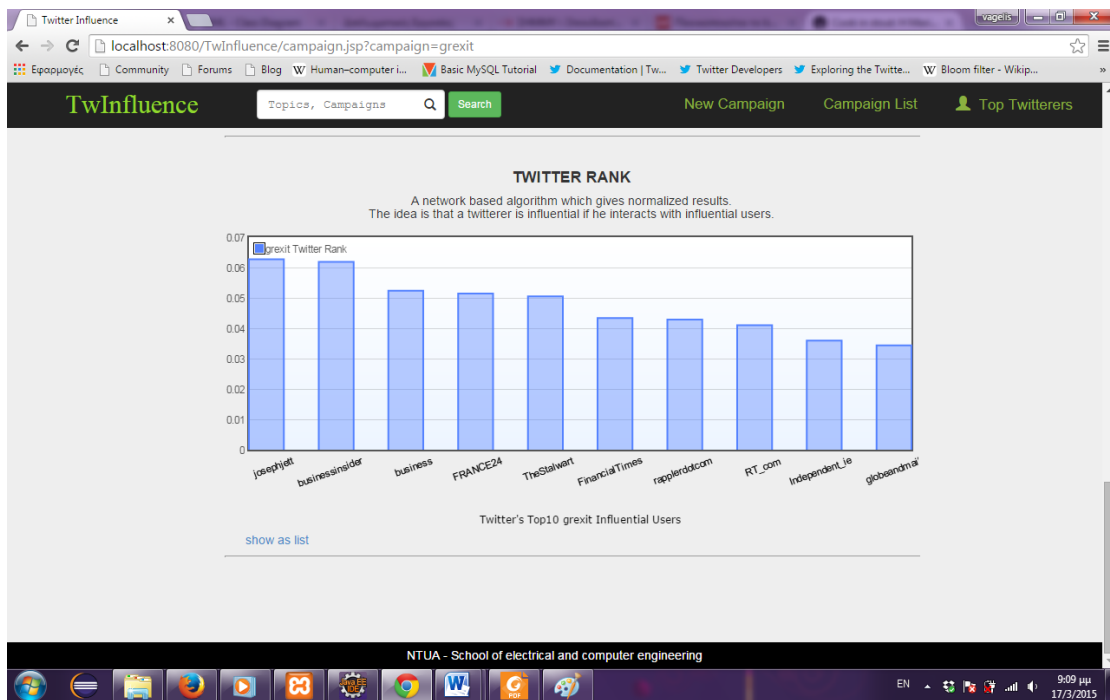
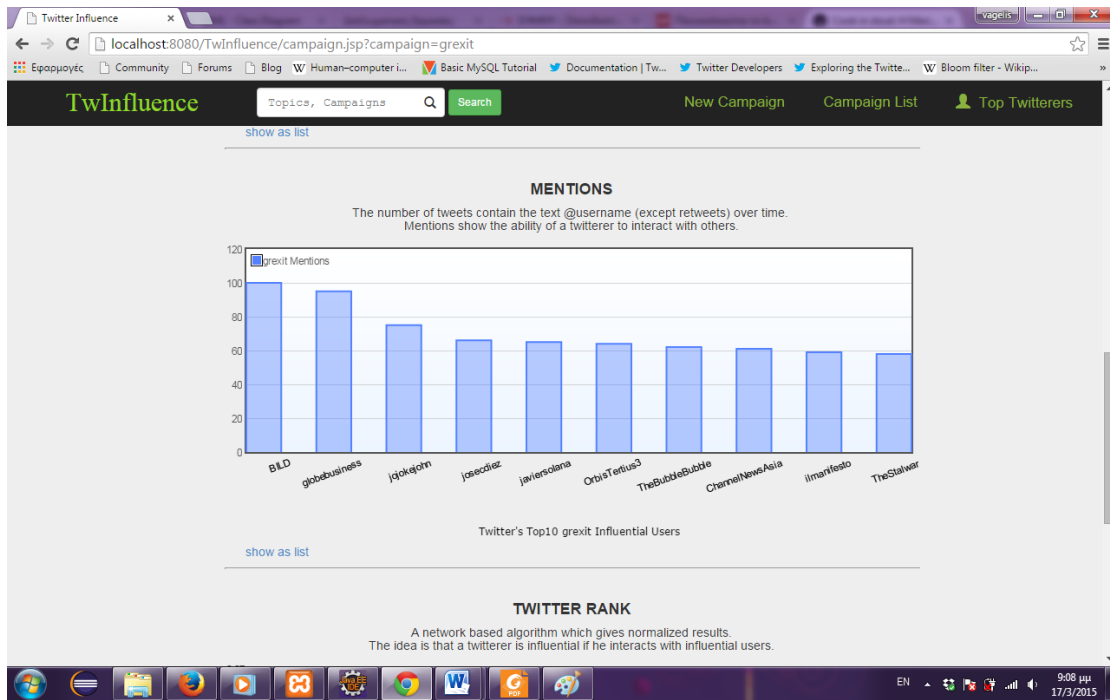




Το campaign “grexit” είναι ευδιάκριτο (το τέταρτο στην παραπάνω εικόνα).

Ο χρήστης βλέπει τα χαρακτηριστικά του campaign παρατηρεί ότι είναι σχετικά πρόσφατο και περιέχει αρκετά tweets και 50 twitterers, οπότε το επιλέγει κάνοντας κλικ στο σύνδεσμο με το όνομα “grexit”, οπότε ανακατευθύνεται στη σελίδα campaign.jsp για να δει τα αποτελέσματα. Ακολουθεί η σελίδα αυτή σε τέσσερις εικόνες που δείχνουν όλα τα είδη influence της εφαρμογής.





Στη συνέχεια, ο χρήστης διάβασε την περιγραφή για κάθε τύπο influence και μελετάει τα αποτελέσματα. Επίσης, ενδιαφέρεται να δει περισσότερους από 10 twitterers. Συνεπώς, κρίνει ότι μάλλον ο αλγόριθμος TwitterRank του ταιριάζει περισσότερο και επιλέγει να δει τους twitterers σε μορφή λίστας (σύνδεσμος “show as list”). Γίνεται ανακατεύθυνση στη σελίδα user.jsp. Ακολουθεί εικόνα της σελίδας αυτής.

Twitter Influence

localhost:8080/TwInfluence/userList.jsp?campaign=grexit&run=1&algo=tr

Topics, Campaigns Search New Campaign Campaign List Top Twitterers

Most influential users in grexit.

Users with the best score in twitterRank

Check out the results with other algorithms

- Joseph Jett (@josephjett)** New York, NY GREXIT influential  
Father, CEO, Author, Speaker, Harvard MBA, MIT MS BS <http://t.co/QRm7LIHHS> Private Equity, Intl. Corporate Finance, Expert Witness, Reputation Management  
• followers: 121194 • friends: 26368 • tweets: 168469
- Business Insider (@businessinsider)** New York, NY GREXIT influential  
What you need to know.  
• followers: 949514 • friends: 160 • tweets: 196123
- Bloomberg Business (@business)** New York and the World GREXIT influential  
The first word in business news.  
• followers: 2090402 • friends: 641 • tweets: 91806
- FRANCE 24 (@FRANCE24)** Paris, France GREXIT influential  
International News 24/7 in your language: @France24\_fr - @France24\_en - @France24\_ar Contact us: <http://t.co/ZaaXsh8d4l> Mobile: <http://t.co/gqoPU'BO4z5>  
• followers: 1442646 • friends: 392 • tweets: 155787

indegree retweets mentions

Εμφανίζονται μόνο πέντε twitterers και ο χρήστης επιλέγει μερικές φορές το “more users” ώστε να βλέπει περισσότερους twitterers.

Twitter Influence

localhost:8080/TwInfluence/userList.jsp?campaign=grexit&run=1&algo=tr&cursor=1

Topics, Campaigns Search New Campaign Campaign List Top Twitterers

Most influential users in grexit.

Users with the best score in twitterRank

Check out the results with other algorithms

- FRANCE 24 (@FRANCE24)** Paris, France GREXIT influential  
International News 24/7 in your language: @France24\_fr - @France24\_en - @France24\_ar Contact us: <http://t.co/ZaaXsh8d4l> Mobile: <http://t.co/gqoPU'BO4z5>  
• followers: 1442646 • friends: 392 • tweets: 155787
- Joseph Weisenthal (@TheStalwart)** New York City GREXIT influential  
Doing something new at Bloomberg.  
• followers: 88123 • friends: 3547 • tweets: 208737
- Financial Times (@FinancialTimes)** London, UK GREXIT influential  
Headlines from the FT. Follow us on @FT for news stories, features and updates. Our customer service team is @FTcare.  
• followers: 2729593 • friends: 649 • tweets: 97353
- Rappler (@rapplerdotcom)** Philippines GREXIT influential  
The Social News Network. Uncompromised journalism that inspires smart conversation and a thirst for change. Download the Rappler app for iOS, Android & Windows8  
• followers: 782148 • friends: 385 • tweets: 229632
- RT (@RT\_com)** GREXIT influential  
RT, the global news network, broadcasts from Moscow, London and Washington studios to over 100 countries.  
• followers: 885821 • friends: 479 • tweets: 112263

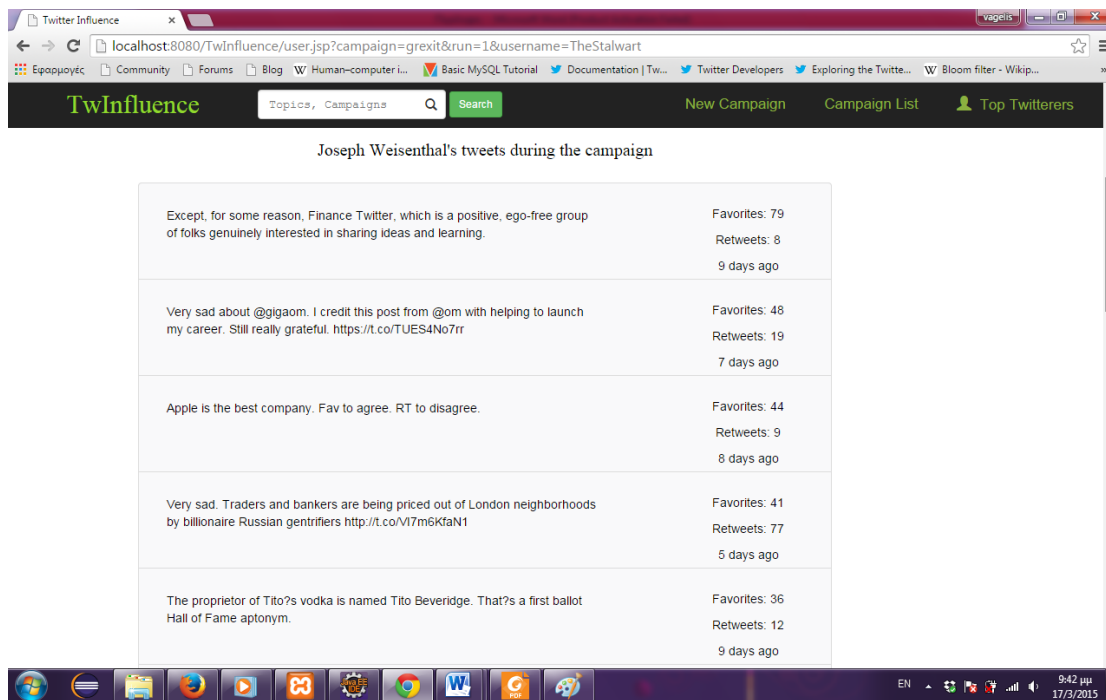
indegree retweets mentions

Ο χρήστης της εφαρμογής πιστεύει ότι ο Twitterer Joseph W. παρουσιάζει πολύ μεγάλο ενδιαφέρον. Οπότε, κάνει κλικ στη φωτογραφία του ώστε να μεταβεί στη σελίδα παρουσίασης του (σελίδα user.jsp). Ακολουθεί η εικόνα για τη σελίδα αυτή.

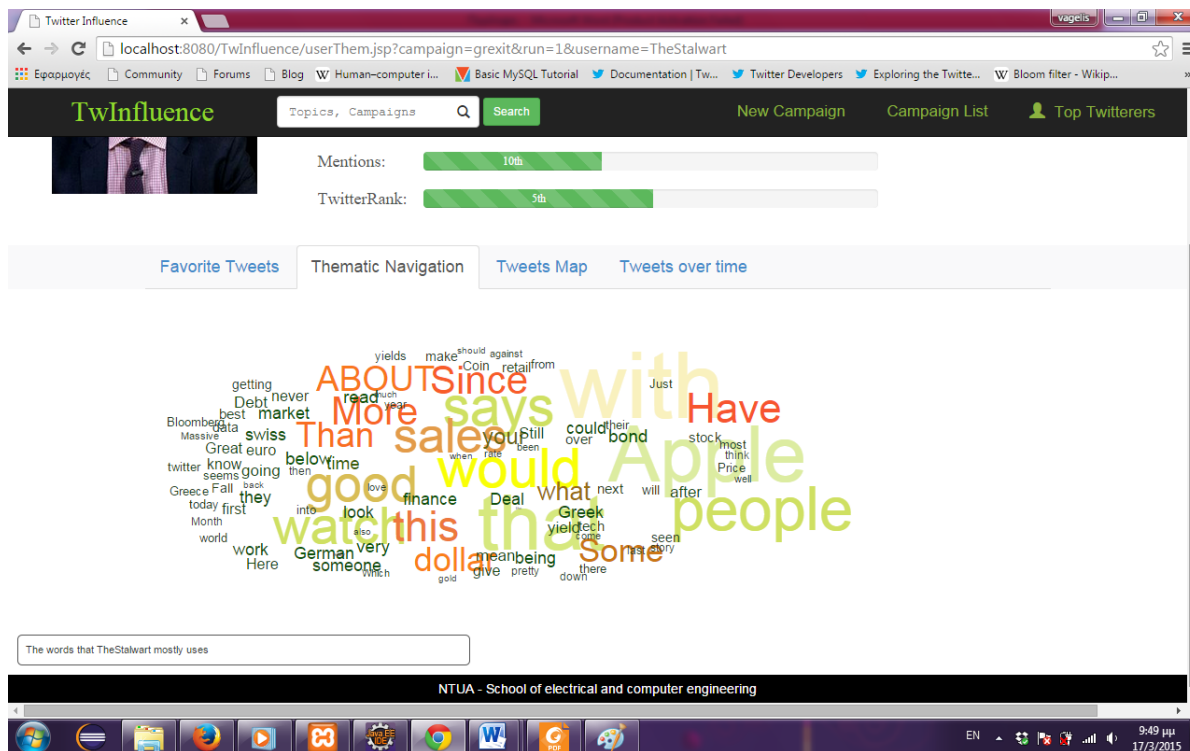
The screenshot shows a web browser window displaying the TwInfluence application. The browser's address bar shows the URL: localhost:8080/TwInfluence/user.jsp?campaign=grexit&run=1&username=TheStalwart. The application header includes the TwInfluence logo, a search bar, and navigation links for 'New Campaign', 'Campaign List', and 'Top Twitterers'. The main content area displays the profile of Joseph Weisenthal (@TheStalwart), identified as a GREXIT influential user from New York City. The profile includes a profile picture, a bio snippet 'Doing something new at Bloomberg. Joseph Weisenthal on twitter.', and statistics: 208737 tweets, 88123 followers, and 3547 friends. Below the statistics are four progress bars representing Indegree (3rd), Retweets (32th), Mentions (10th), and TwitterRank (5th). Navigation tabs at the bottom of the profile section include 'Favorite Tweets', 'Thematic Navigation', 'Tweets Map', and 'Tweets over time'. A tweet is displayed with the text: 'Except, for some reason, Finance Twitter, which is a positive, ego-free group of folks genuinely interested in sharing ideas and learning.' and 'Very sad about @gigaom. I credit this post from @om with helping to launch'. The tweet shows 79 favorites and 8 retweets, posted 9 days ago. The Windows taskbar at the bottom shows the system tray with the time 9:27 AM and date 17/3/2015.

Παρατηρεί ότι ο συγκεκριμένος twitterer είναι από τη Νέα Υόρκη, δουλεύει στο Bloomberg και έχει πάνω από 200.000 tweets. Επίσης βλέπει ότι στο campaign αυτό είναι στη 34<sup>η</sup> θέση στο indegree, στην 32<sup>η</sup> στα retweets, στη 10<sup>η</sup> στα Mentions και στην 5<sup>η</sup> στο TwitterRank. Αυτό σημαίνει πως παρότι δεν έχει πολύ μεγάλο κοινό (followers), πηγαίνει ικανοποιητικά στα retweets, δηλαδή τα tweets του έχουν ωραίο περιεχόμενο, ωστόσο, τα πηγαίνει εξαιρετικά στα mentions και στο TwitterRank. Αυτό σημαίνει ότι είναι πολύ καλός στο να συνομιλεί και να κρατάει επαφή με το κοινό του (mentions), καθώς επίσης και ότι είναι πολύ καλά δικτυωμένος, έχοντας followers, πολλούς, πολύ influential twitterers από το συγκεκριμένο campaign.

Τώρα ο χρήστης της εφαρμογής θέλει να δει τα καλύτερα tweets του συγκεκριμένου twitterer. Κάνει scroll down στη σελίδα και τα διαβάζει, βλέποντας επίσης πότε γράφτηκαν και πόσα retweets πήραν.



Ωστόσο, ο χρήστης της εφαρμογής ενδιαφέρεται να δει περισσότερα tweets χωρίς απαραίτητα να έχουν επισημανθεί ως favorite πολλές φορές, ενδιαφέρεται κυρίως για το θέμα των tweets. Οπότε επιλέγει το tab “Thematic Navigation” και βλέπει το word cloud με τις λέξεις που έχει κυρίως χρησιμοποιήσει ο twitterer. Ακολουθεί η αντίστοιχη εικόνα.

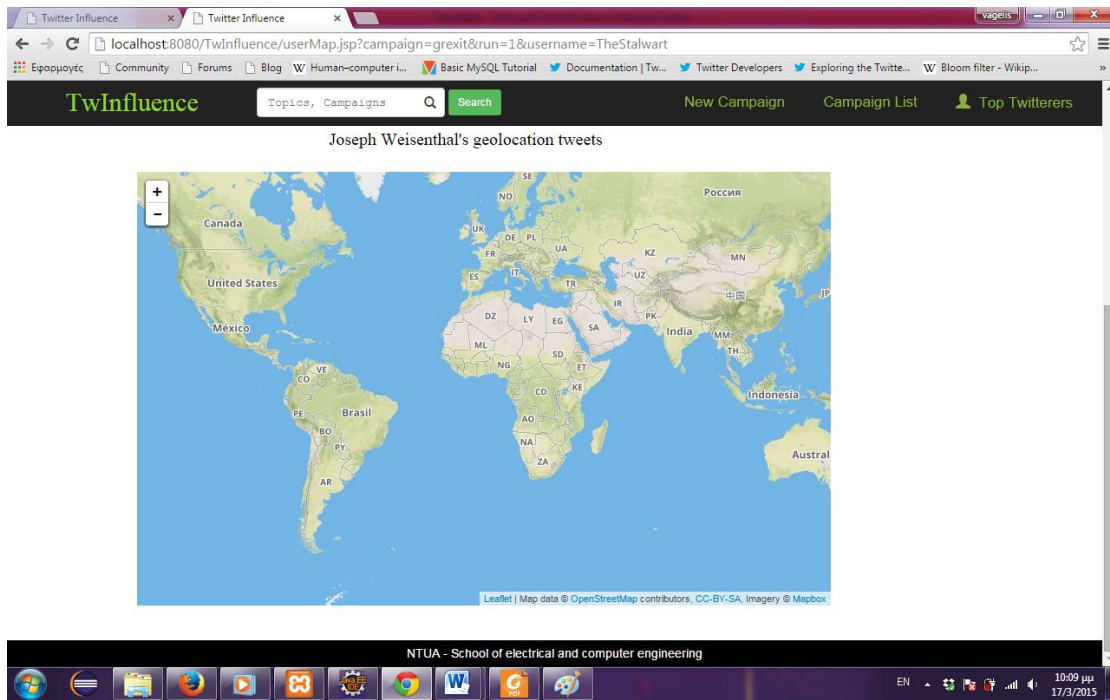


Ο χρήστης της εφαρμογής, στο αριστερό μέρος του word cloud, βλέπει με μικρά σκούρα γράμματα τη λέξη “Greece”. Είναι ακριβώς στο θέμα που τον ενδιαφέρει οπότε κάνει κλικ πάνω στη λέξη. Ανοίγει ένα νέο tab με όλα τα tweets με τη λέξη αυτή (24 tweets). Στην επόμενη εικόνα φαίνονται μερικά.

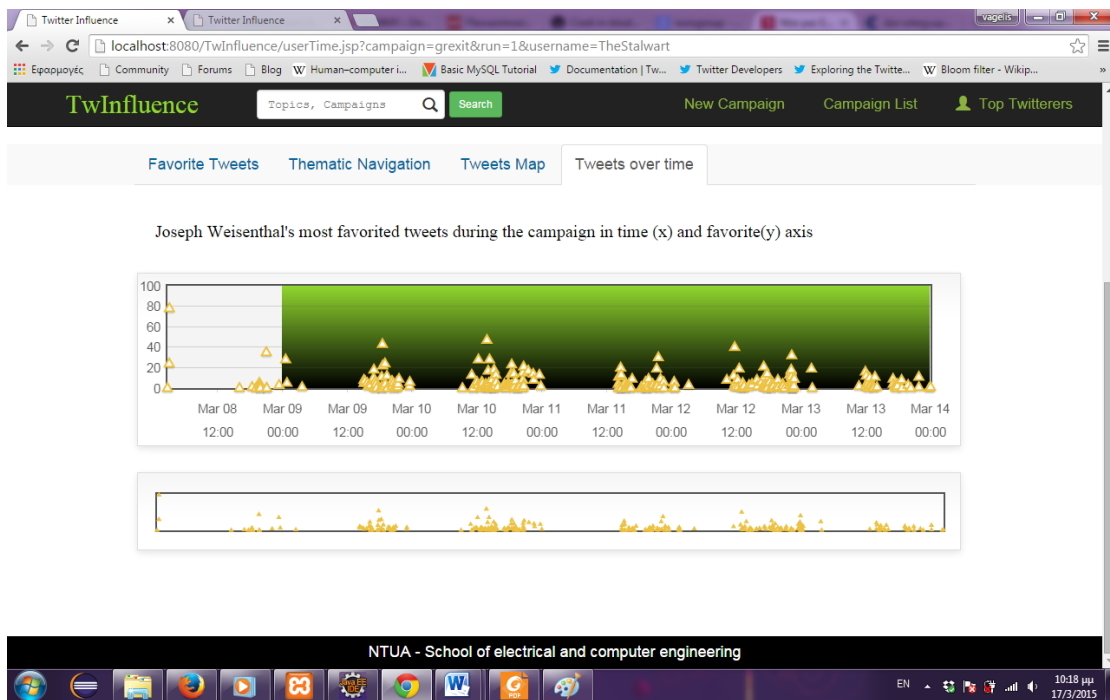
The screenshot shows a web browser window displaying the TwInfluence application. The URL is localhost:8080/TwInfluence/user.jsp?campaign=grexit&username=TheStalwart&word=Greece. The page title is "Joseph Weisenthal's tweets during the campaign with the word Greece". The page displays a list of tweets with their respective engagement metrics (Favorites and Retweets) and timestamps.

Tweet Text	Favorites	Retweets	Time Ago
Not good developments in Greece <a href="http://t.co/ztiZmLhZs">http://t.co/ztiZmLhZs</a> From @nchrysoloras	29	52	8 days ago
This is pretty wild. Greece filing a complaint against Wolfgang Schauble <a href="http://t.co/Ub6vibT9IK">http://t.co/Ub6vibT9IK</a>	14	8	5 days ago
89% of people in Luxembourg feel like they re a "citizen of the EU." Just 45% in Greece feel the same. Via Goldman: <a href="http://t.co/PzcfJ9GrKH">http://t.co/PzcfJ9GrKH</a>	13	26	4 days ago
RT @business: Euro-area finance ministers to Greece: ?We have to stop wasting time? <a href="http://t.co/mxwAgNSVU">http://t.co/mxwAgNSVU</a> <a href="http://t.co/F0DEyv69Fo">http://t.co/F0DEyv69Fo</a>	0	74	8 days ago
RT @Stefan_Riecher: But wait, wasn t there a deal and all good just a couple weeks ago? #Greece #Grexit	0	4	8 days ago

Βέβαια, ο χρήστης της εφαρμογής, εύλογα, πιστεύει ότι θα μπορούσαν να υπάρχουν και άλλα tweets για την Ελλάδα, που όμως απλά δεν περιέχουν τη λέξη. Κάνει κλικ στο tab “Tweets Map” μήπως υπάρχει κάποιο tweet για την Ελλάδα πάνω στο χάρτη, ωστόσο ο χρήστης μας απογοητεύεται, καθώς ο κύριος W. δεν χρησιμοποιεί καθόλου το GeoLocation στα tweets του, όπως φαίνεται και στην εικόνα.

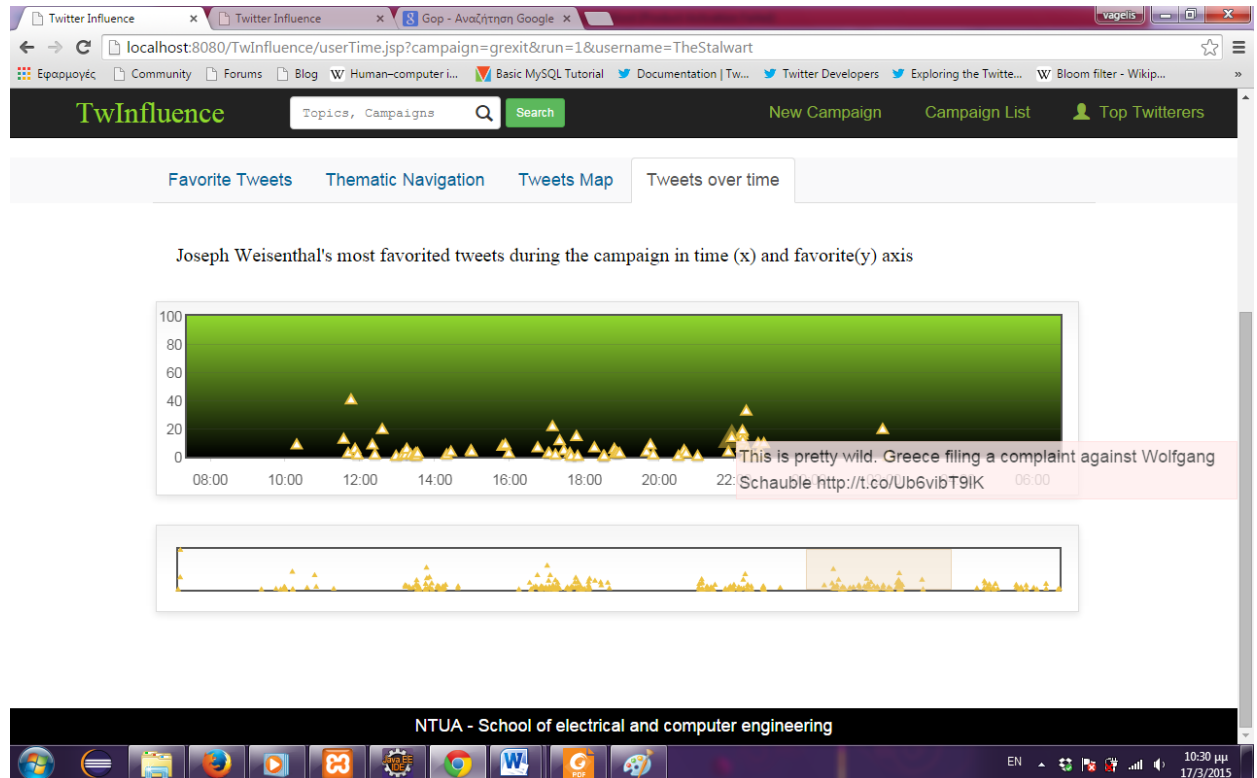


Τέλος, ως προς την Ελλάδα, ο χρήστης μας θυμάται ότι την Παρασκευή 13 Μαρτίου υπήρξε ένα, κρίσιμο για την πορεία της Ελλάδας στην Ευρώπη, Eurogroup. Στην Ελλάδα το αποτέλεσμα σχολιάστηκε θετικά. Πώς όμως σχολιάζεται στην Αμερική από influential twitterers και τι λέει ο κύριος W.; Ο χρήστης της εφαρμογής επιλέγει το tab “Tweets over Time”.



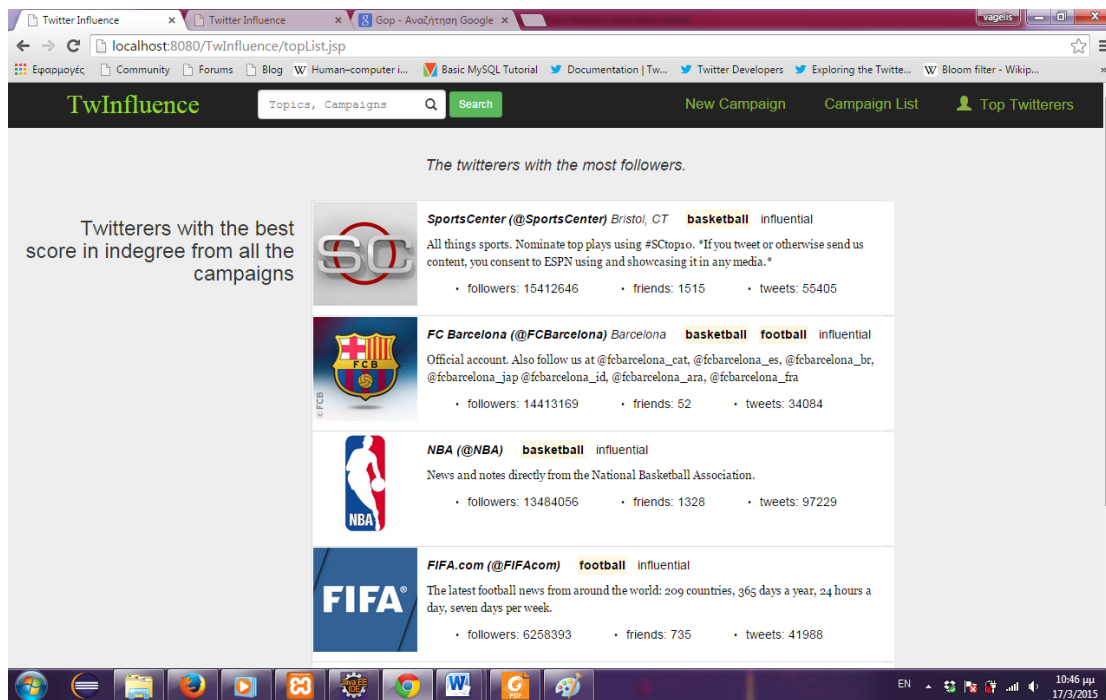


Ο χρήστης κάνει zoom στο διάστημα Πέμπτη έως Σάββατο και βλέπει μερικά ενδιαφέροντα tweets, και τελικά κάνει zoom την Παρασκευή από τις 8.00 μέχρι τις 24.00 όπως φαίνεται και στην εικόνα που ακολουθεί.



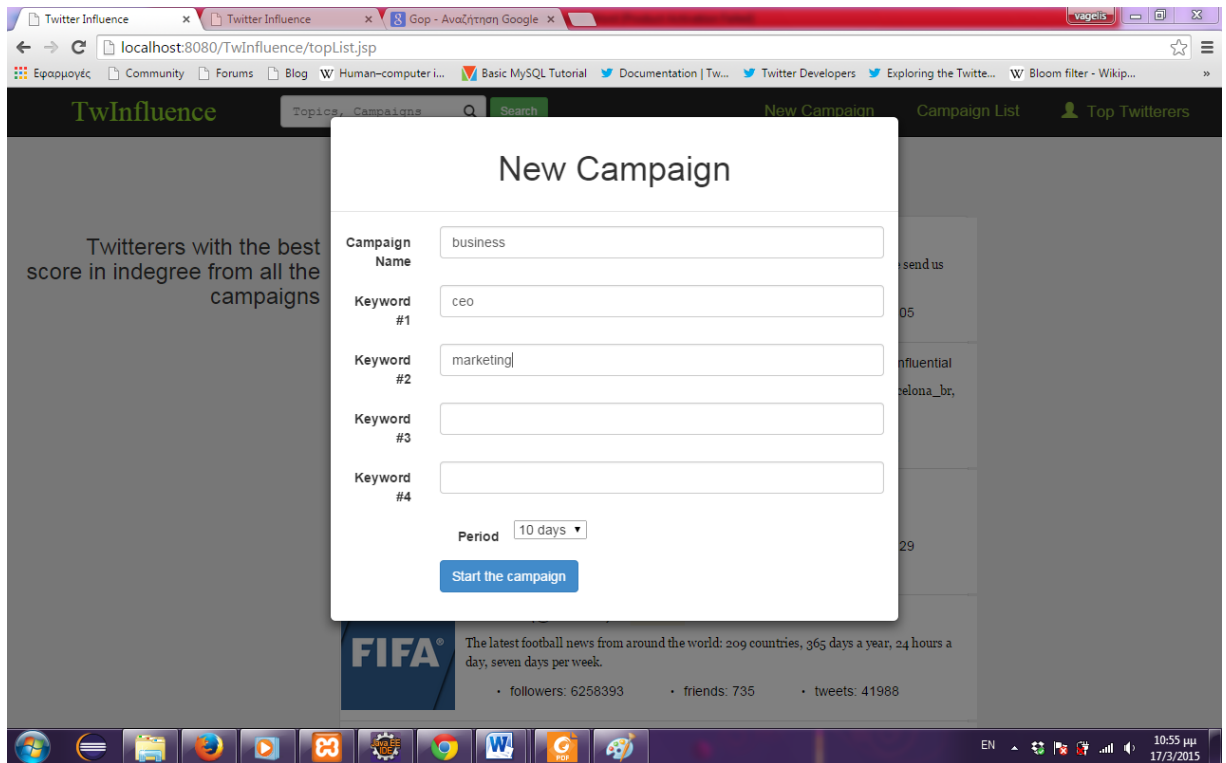
Τώρα, ο χρήστης μας εντυπωσιασμένος από τα tweets του κυρίου W. θέλει να τον βρει στο Twitter. Διαλέγει την επιλογή “Mr W. on Twitter” δίπλα στην περιγραφή του twitterer στο πάνω μέρος της σελίδας και ανακατευθύνεται στο προφίλ του twitterer στο Twitter.

Αφού βρήκε αυτό που έψαχνε ο χρήστης της εφαρμογής ενδιαφέρεται να δει γενικότερα τους influentials ανεξαρτήτως θέματος. Οπότε επιλέγει το σύνδεσμο “Top Twitterers” πάνω δεξιά στο πανθαρ της σελίδας. Βλέπει twitterers που οι περισσότεροι δεν έχουν σχέση με το “grexit”, ωστόσο κάποιιοι του είναι γνωστοί. Όπως φαίνεται στην παρακάτω εικόνα.

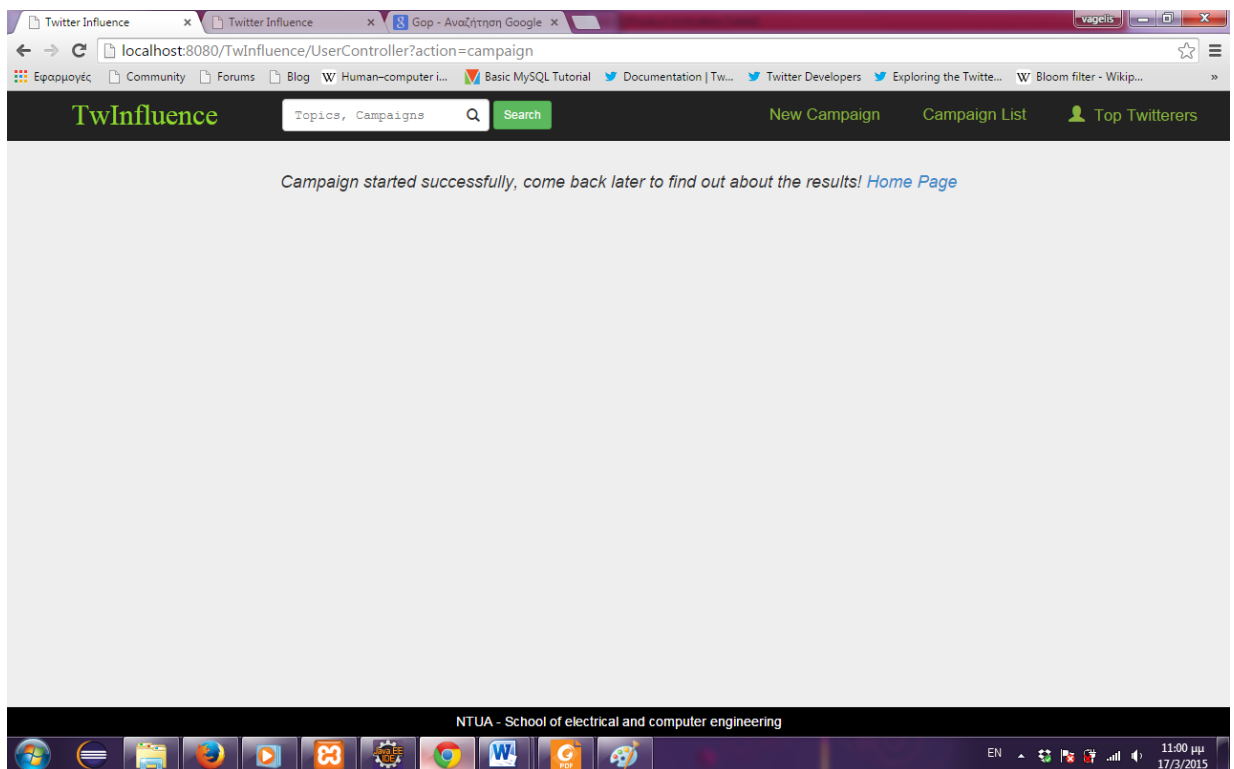


Ο χρήστης παρατηρεί ότι στη δεύτερη θέση δεσπόζει η Barcelona, η αγαπημένη του ομάδα. Οπότε, μπαίνει στον πειρασμό να δει τα tweets της, κάνοντας κλικ στο σύνδεσμο της. Η διαδικασία είναι παρόμοια με πριν.

Τέλος, αφού τελείωσε και με τη Barcelona ο χρήστης αποφασίζει να τρέξει ένα καινούριο campaign με θέμα τις επιχειρήσεις. Κάνει κλικ στην επιλογή New Campaign στο πάνω μέρος της σελίδας και συμπληρώνει τη φόρμα που του δίνεται, όπως φαίνεται στην εικόνα.



Πατάει το κουμπί “Start the campaign” και του εμφανίζεται ένα μήνυμα ότι το campaign ξεκίνησε, όπως φαίνεται και στην εικόνα.



# 8

## *Επίλογος*

### *8.1 Σύνοψη και συμπεράσματα*

Συνοψίζοντας, αναπτύξαμε μια web- εφαρμογή που επιτρέπει σε κάθε χρήστη να τρέχει campaigns και να έχει πρόσβαση σε influential twitterers χωρίς απαραίτητα να έχει θεωρητικές ή τεχνικές γνώσεις, είτε γύρω από το influence, είτε περί προγραμματισμού και ανάπτυξης εφαρμογών. Επιπλέον, το influence παρουσιάζεται στο χρήστη με τρόπο ξεκάθαρο, και μόνο για πράγματα που τον ενδιαφέρουν, δηλαδή τελικούς influentials με την επίδοσή τους αποκρύπτοντας θέματα για τα οποία ο τελικός χρήστης δεν ενδιαφέρεται ιδιαίτερα, όπως είναι η υλοποίηση.

Όσον αφορά το θεωρητικό κομμάτι, τα αποτελέσματα που εξάγονται από την εφαρμογή μας, παρατηρούμε ότι είναι συμβατά με τη θεωρία. Πιο συγκεκριμένα, παρατηρούμε, πρωτίστως, ότι για τους περισσότερους influential twitterers, η κατάταξή τους στο influence είναι πολύ διαφορετική από μέτρο σε μέτρο. Το γεγονός αυτό καταδεικνύει την ύπαρξη διαφορετικών τύπων influence, όπως παρουσιάζουμε και στο θεωρητικό υπόβαθρο (κεφάλαιο 3). Επιπλέον, παρατηρούμε ότι υπάρχουν αρκετοί influential twitterers, οι οποίοι είναι influential σε παραπάνω από ένα campaign και πολλές φορές τα campaigns που συμμετέχουν είναι τελείως διαφορετικά μεταξύ τους, πράγμα που συμβαδίζει με την παραδοσιακή θεωρία που λέει ότι οι influentials μπορεί να ασκούν επιρροή σε αρκετά topics.

## 8.2 Μελλοντικές επεκτάσεις

Το σύστημα μας θα μπορούσε να εμπλουτιστεί στο μέλλον τόσο ως προς την απόδοση (κυρίως ταχύτητα υπολογισμού του campaign) όσο και με επιπλέον υπηρεσίες που θα μπορούσε να παρέχει.

Ως προς τον χρόνο, για τον υπολογισμό του campaign, εφ' όσον η κυριότερη πηγή καθυστέρησης είναι τα όρια σε request στο χρόνο που επιτρέπει το Twitter API, θα μπορούσε να χρησιμοποιηθεί παράλληλη εκτέλεση για κάποιες χρονοβόρες μεθόδους. Συγκεκριμένα, οι constructors των κλάσεων UserTimeline, MentionCollector και FriendsCollector θα μπορούσαν να τρέχουν παράλληλα, καθότι δεν υπάρχει καμία εξάρτηση μεταξύ τους, ενώ κάνουν διαφορετικά είδη request στο twitter API. Επιπλέον, στον υπολογισμό του campaign τα αποτελέσματα των μετρικών Indegree, Retweets και Mentions είναι διαθέσιμα αρκετά νωρίτερα από τα αποτελέσματα του TwitterRank, πράγμα το οποίο θα μπορούσαμε να το εκμεταλλευτούμε, παρέχοντας στον τελικό χρήστη τα αποτελέσματα αυτά και ένα μήνυμα ότι εκκρεμεί το τελευταίο αποτέλεσμα.

Όσον αφορά τις επιπλέον λειτουργίες, πρώτον, μέχρι στιγμής, στο σύστημα μας ένα campaign ανανεώνεται χειροκίνητα, όταν ένας χρήστης πατήσει το κουμπί “Relaunch this campaign”, αυτή η λειτουργία θα μπορούσε να γίνεται αυτόματα, ανά τακτά χρονικά διαστήματα ώστε να είναι πάντα ενημερωμένα τα περισσότερα campaigns καθώς επίσης και να μη δημιουργείται μεγάλος φόρτος στο σύστημα ή μεγάλες καθυστερήσεις, αν για παράδειγμα πολλοί χρήστες πατήσουν το “Relaunch this campaign” σε πολλά campaigns.

Επιπροσθέτως, καλό θα ήταν να αναπτυχθεί κάποιος μηχανισμός για καλύτερη αναζήτηση, με στόχο την απόρριψη χρηστών οι οποίοι δραστηριοποιούνται σε topics που δεν έχουν άμεση σχέση με το συγκεκριμένο campaign, αλλά χρησιμοποιούν παρόμοιες λέξεις. Ένα τέτοιο παράδειγμα αποτελεί ένα tweet που περιέχει τον όρο “sports fashion”. Μια αναζήτηση με το keyword “sports” θα επιστρέψει και tweets που έχουν τον προηγούμενο όρο. Πιθανό αποτέλεσμα είναι στο campaign “sports” να υπάρχουν και twitterers που είναι “fashion” influentials.

Μια ακόμη πιθανή επέκταση θα μπορούσε να είναι η χρήση οντολογίας για τις σχέσεις μεταξύ των campaigns. Με αυτόν τον τρόπο, το σύστημα μας θα μπορούσε να καταλάβει ότι για παράδειγμα ένας influential twitterer στο campaign “football” θα ήταν και influential στο campaign “sports”.

Τέλος, κάποιες ακόμη ενδιαφέρουσες λειτουργίες θα ήταν, πρωτίστως, η δυνατότητα αναζήτησης twitterer (μέχρι τώρα παρέχεται μόνο η αναζήτηση campaign), αλλά και πιο προχωρημένη αναζήτηση π.χ. με βάση τοποθεσία και campaign. Δεύτερον, αφού παρέχουμε τη δυνατότητα στο χρήστη να βλέπει τα tweets ενός influential twitterer, θα είχε ενδιαφέρον να μπορούσε να δει μόνο τα tweets που σχετίζονται με το συγκεκριμένο campaign π.χ. σε ένα νέο tab στην σελίδα user.jsp, με το όνομα “campaigned tweets”.

# 9

## *Βιβλιογραφία*

- [1] Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, Krishna P. Gummadi. Measuring User Influence in Twitter: The Million Follower Fallacy. In proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, 2010
  
- [2] Jianshu Weng, Ee-Peng Lim, Jing Jiang, Qi He. TwitterRank: Finding Topic-sensitive Influential Twitterers. WSDM'10, February 4–6, 2010, New York City, New York, USA
  
- [3] Eytan Bakshy, Jake M. Hofman, Winter A. Mason, Duncan J. Watts. Everyone's an Influencer: Quantifying Influence on Twitter. WSDM'11, February 9–12, 2011, Hong Kong, China.
  
- [4] Shaozhi Ye, Felix Wu. Measuring Message Propagation and Social Influence on Twitter.com. 2010
  
- [5] Cha, M.; Mislove, A.; and Gummadi. A Measurement Driven Analysis of Information Propagation in the Flickr Social Network. In WWW. 2009

- [6] Berry, J., and Keller, E. *The Influentials: One American in ten tells the other nine how to vote, where to eat, and what to buy.* 2003.
- [7] Kwak, H., Lee, C., Park, H., Moon, S.: What is Twitter, a social network or a news media? In: *WWW '10: Proceedings of the 19th International Conference on World Wide Web.* pp. 591–600. 2010.