



ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

**ΣΧΟΛΗ ΨΗΦΙΑΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΜΑΤΙΚΗΣ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΠΛΗΡΟΦΟΡΙΚΗ ΚΑΙ ΤΗΛΕΜΑΤΙΚΗ»**

ΚΑΤΕΥΘΥΝΣΗ «ΤΕΧΝΟΛΟΓΙΕΣ ΚΑΙ ΕΦΑΡΜΟΓΕΣ ΙΣΤΟΥ»

Οπτική απεικόνιση διάδοσης και περιεχομένου αναρτήσεων στο Twitter
Διπλωματική Εργασία

Δημήτριος Παππάς

Αθήνα, 2017



ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

**ΣΧΟΛΗ ΨΗΦΙΑΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΜΑΤΙΚΗΣ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΠΛΗΡΟΦΟΡΙΚΗ ΚΑΙ ΤΗΛΕΜΑΤΙΚΗ»**

ΚΑΤΕΥΘΥΝΣΗ «ΤΕΧΝΟΛΟΓΙΕΣ ΚΑΙ ΕΦΑΡΜΟΓΕΣ ΙΣΤΟΥ»

Τριμελής Εξεταστική Επιτροπή

Καραγιώργου Σοφία (Επιβλέπουσα)

**Επισκέπτης Λέκτορας, Τμήμα Πληροφορικής και Τηλεματικής, Χαροκόπειο
Πανεπιστήμιο**

Βαρλάμης Ηρακλής

**Επίκουρος Καθηγητής, Τμήμα Πληροφορικής και Τηλεματικής, Χαροκόπειο
Πανεπιστήμιο**

Τσερπές Κωνσταντίνος

**Επίκουρος Καθηγητής, Τμήμα Πληροφορικής και Τηλεματικής, Χαροκόπειο
Πανεπιστήμιο**

Ο Δημήτριος Παππάς

δηλώνω υπεύθυνα ότι:

- 1)** Είμαι ο κάτοχος των πνευματικών δικαιωμάτων της πρωτότυπης αυτής εργασίας και από όσο γνωρίζω η εργασία μου δε συκοφαντεί πρόσωπα, ούτε προσβάλλει τα πνευματικά δικαιώματα τρίτων.
- 2)** Αποδέχομαι ότι η ΒΚΠ μπορεί, χωρίς να αλλάξει το περιεχόμενο της εργασίας μου, να τη διαθέσει σε ηλεκτρονική μορφή μέσα από τη ψηφιακή Βιβλιοθήκη της, να την αντιγράψει σε οποιοδήποτε μέσο ή/και σε οποιοδήποτε μορφότυπο καθώς και να κρατά περισσότερα από ένα αντίγραφα για λόγους συντήρησης και ασφάλειας.

ΕΥΧΑΡΙΣΤΙΕΣ

Ολοκληρώνοντας τη μεταπτυχιακή μου εργασία στο πλαίσιο των μεταπτυχιακών μου σπουδών στο Χαροκόπειο Πανεπιστήμιο (Τμήμα Πληροφορικής και Τηλεματικής), θα ήθελα να ευχαριστήσω όλους όσους συνέβαλαν άμεσα στην επιτυχή υλοποίησή της.

Καταρχάς, οφείλω ένα μεγάλο ευχαριστώ στην επιβλέπουσα της εργασίας μου, Καραγιώργου Σοφία, που στήριξε τις προσπάθειές μου, βοήθησε αρχικά στην επιλογή του σχετικού θέματος και κατεύθυνε αποτελεσματικά την αρχική μου διερεύνηση επί του αντικειμένου της εργασίας. Επίσης, θα ήθελα να εκφράσω τις ευχαριστίες μου στον Βαρλάμη Ηρακλή, που με καθοδήγησε αποτελεσματικά κατά τη διάρκεια υλοποίησης της έρευνας.

Τέλος, οφείλω να ευχαριστήσω την οικογένειά μου για την στήριξή και συμπαράσταση που επέδειξε κατά τη διάρκεια των σπουδών μου και ειδικότερα κατά τη περίοδο υλοποίησης της παρούσας έρευνας.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ	5
Περίληψη.....	7
Abstract	9
ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ	10
ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ.....	13
ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ	14
ΚΕΦ.1: Εισαγωγή.....	15
1.1 Αντικείμενο έρευνας	15
1.2 Σκοπός έρευνας.....	16
1.3 Ερευνητικό κενό – Προσφορά έρευνας	17
1.5 Δομή έρευνας.....	17
ΚΕΦ.2: Κοινωνικά δίκτυα και Twitter	19
2.1 Web 2.0	19
2.2 Κοινωνικά δίκτυα	19
2.2.1 Τι είναι.....	19
2.2.2 Κίνητρα ανάπτυξης	22
2.2.3 Πλεονεκτήματα χρήσης.....	22
2.2.4 Κίνδυνοι από τη χρήση τους	23
2.2.5 Στοιχεία χρήσης.....	24
2.3 Το φαινόμενο Twitter.....	25
2.3.1 Τι είναι.....	25
2.3.2 Το tweet	27
2.3.3. Το retweet.....	28
2.3.4 Το follow	28
2.3.5 Το hashtag	29
ΚΕΦ.3: Πρότερες σχετικές έρευνες οπτικής απεικόνισης αναρτήσεων στο Twitter	30
3.1 Διαθέσιμοι Ιστότοποι	30
3.1.1 TrendsMap	30
3.1.2 OneMillionTweetMap	33
3.1.3 TweetMap	35
3.1.4 KeyHole.....	36
3.2 Σχετικές έρευνες.....	37
3.2.1 Οπτικοποίηση ειδήσεων σε χάρτη.....	37
3.2.2 Οπτικοποίηση γεωγραφικού ενδιαφέροντος χρηστών.....	40
3.2.3 Οπτικοποίηση συνοδευτικών εικόνων αναρτήσεων tweets.....	41
3.2.4 Οπτικοποίηση χαρακτηριστικών αναρτήσεων tweets	43
3.2.5 Οπτικοποίηση τοποθεσίας ανάρτησης και τοποθεσία αναφοράς tweets.....	44
3.3 Σύνοψη - Σύγκριση	45
ΚΕΦ.4: Μεθοδολογία έρευνας – Σχεδίαση έρευνας – Εργαλεία που χρησιμοποιήθηκαν	48
4.1 Δείγμα έρευνας	48
4.2 Εργαλεία συλλογής δεδομένων	48
4.2.1 Twitter Streaming API	49
4.2.2 REST API.....	51
4.2.3 Python	52

4.2.4 PostgreSQL	52
4.3 Εργαλεία επεξεργασίας δεδομένων	53
4.3.1 MS SQL Server	53
4.4 Εργαλεία υλοποίησης ιστοτόπου	55
4.4.1 Google Maps API	56
4.4.2 Μοντέλο προγραμματισμού .NET	58
4.4.3 Η γλώσσα προγραμματισμού C#	59
4.4.4 Η γλώσσα σεναρίων jQuery	59
ΚΕΦ.5: Υλοποίηση	61
5.1 Συλλογή δεδομένων	61
5.2 Επεξεργασία δεδομένων	63
5.3 Παρουσίαση δεδομένων	69
5.4 Υλοποίηση ιστοσελίδας απεικόνισης	71
ΚΕΦ.6: Οπτική απεικόνιση	74
6.1 Απεικόνιση όλων των tweets που συλλέχθηκαν	74
6.1.1 Συστοιχίες σήμανσης (marker clusters)	74
6.1.2 Οπτικός χάρτης εστίασης (heatmap)	77
6.2 Απεικόνιση διάδοσης hashtags	78
6.3 Απεικόνιση διάδοσης retweets	81
6.4 Απεικόνιση δημοφιλών λέξεων κλειδιών στον χάρτη	85
6.5 Απεικόνιση δημοφιλών hashtag στον χάρτη	87
ΚΕΦ.7: Συμπεράσματα	92
7.1 Συμπεράσματα έρευνας	92
7.2 Περιορισμοί έρευνας	94
7.3 Προτάσεις για μελλοντικές έρευνες	94
Βιβλιογραφία	96

Περίληψη

Το Twitter έχει εξελιχθεί σε κοινωνικό δίκτυο που χρησιμοποιείται για τη διάδοση πληροφοριών σχετικά με γεγονότα μικρότερης ή μεγαλύτερης κλίμακας. Επιπροσθέτως, από το 2009 τα μηνύματα που αναρτώνται στο Twitter δύναται να φέρουν χωροχρονική σήμανση, επιτρέποντας στους ερευνητές τη χωροχρονική απεικόνιση των σχετικών πληροφοριών.

Σκοπός της παρούσας εργασίας είναι μια περισσότερο αποτελεσματική και εύκολη στην κατανόηση αναπαράσταση και παρακολούθηση των πληροφοριών που φέρουν οι αναρτήσεις στο Twitter, συνδυάζοντας τις δυνατότητες που προσφέρουν οι προγραμματιστικές και απεικονιστικές διεπαφές του Twitter Streaming API και των Google Maps, αντίστοιχα.

Αντικείμενο της έρευνας αποτελεί η χωροχρονική απεικόνιση σε χάρτες Google Maps πληροφοριών που συνοδεύουν αναρτήσεις (tweets) στο Twitter με στόχο την αποτύπωση της δυναμικής των μηνυμάτων και του τρόπου με τον οποίο διαδίδονται γεωγραφικά μέσα από το ψηφιακό μέσο. Για τον σκοπό αυτό συλλέχθηκαν σε πραγματικό χρόνο tweets με χρήση της Python και του Twitter Streaming API, τα οποία στη συνέχεια επεξεργάστηκαν και ομαδοποιήθηκαν σε επίπεδο βάσης δεδομένων (MS SQL Server). Κατόπιν, αναπτύχθηκε ιστοσελίδα σε περιβάλλον .NET με χρήση της προγραμματιστικής διεπαφής (API) των Google Maps, με σκοπό την οπτική απεικόνιση (visualization) των αντίστοιχων tweets και πληροφοριών που εκείνα περιλαμβάνουν.

Η αντίστοιχη εφαρμογή χρησιμοποιήθηκε (i) για την οπτική απεικόνιση του συνόλου των tweets σε συστοιχίες σήμανσης (marker clusters) και οπτικούς χάρτες εστίασης (heatmap), (ii) για την οπτική απεικόνιση διάδοσης των δημοφιλέστερων hashtags σε χάρτη Google Map, (iii) για την οπτική απεικόνιση διάδοσης – ροής εξέλιξης των retweets διαφόρων tweets, (iv) για την οπτική απεικόνιση της ροής εξέλιξης των δημοφιλέστερων hashtags για διάφορες πόλεις των Ηνωμένων Πολιτειών και της Μεγάλης Βρετανίας και (v) για την οπτική απεικόνιση της ροής εξέλιξης των δημοφιλέστερων λέξεων χρήσης στο σύνολο των tweets για τις ίδιες πόλεις.

Με τη χρήση της ιστοσελίδας είναι δυνατή η άμεση παρακολούθηση της ροής διάδοσης ή εξέλιξης ή εναλλαγής των πληροφοριών που φέρουν τα διάφορα tweets, χωρίς την ανάγκη να ανατρέξει ο χρήστης σε κείμενα σχετικών αναλύσεων μηνυμάτων που αναρτώνται στο Twitter. Ενδεικτικά, παρακολουθώντας τη ροή εναλλαγής των δημοφιλών hashtags για διάφορες πόλεις των Ηνωμένων Πολιτειών, διαπιστώνεται πως σε συγκεκριμένη ημερομηνία, κυριαρχεί στις αναρτήσεις στο Twitter για το Σαν Φρανσίσκο το hashtag #traffic, μέρα κατά την οποία πραγματοποιήθηκε σημαντικό αυτοκινητιστικό ατύχημα στην πόλη του Σαν Φρανσίσκο.

Οπτική απεικόνιση διάδοσης και περιεχομένου αναρτήσεων στο Twitter

Λέξεις κλειδιά: Twitter, tweets, streaming data, online data analytics, hashtags, οπτική απεικόνιση, Google Maps

Abstract

Twitter has evolved into a social network that is used for spreading information about events of smaller or larger scale. In addition, since 2009, posting messages on Twitter may contain spatial signature, allowing researchers for the spatial representation of the relevant information.

The purpose of current research is to provide a more effective and easy to understand representation and tracking of tweets posted in Twitter by combining the capabilities offered by Twitter Streaming API Google Maps programming interfaces.

The subject of current survey is the spatial mapping of real-time information included in tweets published in Twitter with the use of Google Maps. For this purpose, tweets were collected in real-time by using Python and Twitter Streaming API. After that relevant data were edited and grouped with the use of MS SQL Server. Next, a web application was implemented using Google Maps API to visualize collected tweets and information contained in relevant tweets.

The corresponding application was used for: (i) the visualization of all tweets in marker clusters and heatmaps (ii) the visualization of the propagation of the most popular hashtags in Google Maps, (iii) the visualization of evolution of retweets in different cities or within a city, (iv) the visualization of the evolution of the most popular hashtags for different cities in United States and Great Britain, and (v) the visualization of the evolution of most popular used words in all tweets for the same cities.

With the use of the relevant website, it is possible to immediately monitor the propagation or evolution or the exchange of information contained in the relevant tweets without the need for the user to refer to the context of applied analyses on relevant tweets. For instance, following the flow - evolution of popular hashtags for various cities in the United States, it was noted that at a specific date, the hashtag #traffic had dominated tweets posted in San Francisco. The relevant day a major car accident took place in San Francisco.

Keywords: Twitter, tweets, streaming data, online data analytics, hashtags, visualization, Google Maps

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικόνα 1: Οπτικοποίηση δημοφιλέστερων hashtags στην Αττική στις 18/05/2017 23:15 από την εφαρμογή TrendsMaps	31
Εικόνα 2: Οπτικοποίηση δημοφιλέστερων λέξεων σε όλο τον κόσμο στις 18/05/2017 από την εφαρμογή TrendsMaps	32
Εικόνα 3: Η είδηση θανάτου του τραγουδιστή Chris Cornell στις 18 Μαΐου 2017 (BBC News, 2017).....	32
Εικόνα 4: Τελευταίες αναρτήσεις tweets για το hashtag #rpw2017	33
Εικόνα 5: Οπτική απεικόνιση των τελευταίων 1 εκατομμυρίων tweets που ελήφθησαν από το Twitter API στην ιστοσελίδα OneMillionTweetMap	33
Εικόνα 6: Οπτική απεικόνιση των tweets που περιλαμβάνουν το hashtag #job στην ιστοσελίδα OneMillionTweetMap (19/05/2017)	34
Εικόνα 7: Οπτική απεικόνιση των tweets που περιλαμβάνουν το hashtag #job στην Ευρώπη στην ιστοσελίδα OneMillionTweetMap (19/05/2017)	35
Εικόνα 8: Οπτική απεικόνιση με χάρτη εστίασης (heatmap) των τελευταίων 1 εκατομμυρίων tweets που ελήφθησαν από το Twitter API στην ιστοσελίδα OneMillionTweetMap.....	35
Εικόνα 9: Οπτικοποίηση των tweets που περιλαμβάνουν τη λέξη Greece για την περίοδο 19/02/2017 - 19/05/2017 στο σύστημα TweetMap.....	36
Εικόνα 10: Αριθμός δημοσιεύσεων tweets που περιλαμβάνουν τη λέξη Greece τη περίοδο 19/02/2017 - 19/05/2017 από το σύστημα TweetMap	36
Εικόνα 11: Οπτική απεικόνιση σε διεπαφή χάρτη των tweets που περιλαμβάνουν το hashtag #greece στις 20/05/2017 στην εφαρμογή KeyHole.....	37
Εικόνα 12: Επισήμανση τρεχουσών ειδήσεων στο χάρτη – TwitterStand	38
Εικόνα 13: Εμφάνιση πληροφοριών τρεχουσών ειδήσεων στο TwitterStand.....	38
Εικόνα 14: Παρουσίαση επιλεγμένης είδησης στο χρήστη στο σύστημα TwitterStand.....	39
Εικόνα 15: Επισήμανση τρεχουσών ειδήσεων στο χάρτη (Αττική, 19/05/2017) – NewsStand...39	
Εικόνα 16: Οπτικοποίηση περιοχής στο σύστημα Limosa (Vosecky et al., 2013)	40
Εικόνα 17: Οπτικοποίηση χρηστών στο σύστημα Limosa (Vosecky et al., 2013).....	40
Εικόνα 18: Οπτικοποίηση θέματος στο σύστημα Limosa (Vosecky et al., 2013)	41
Εικόνα 19: Οπτικοποίηση όρων αναζήτησης στο σύστημα Limosa (Vosecky et al., 2013).....	41
Εικόνα 20: Οπτικοποίηση εικόνων ανάρτησης σε tweets για επιλεγμένα θέματα ειδήσεων στο Τόκιο της Ιαπωνίας (Nakaji & Yanai, 2012)	42
Εικόνα 21: Οπτική απεικόνιση σημαντικών γεγονότων σε χάρτη Google Map (Kaneko & Yanai, 2016).....	43
Εικόνα 22: Εμφάνιση συνοδευτικών εικόνων για επιλεγμένο γεγονός (Kaneko & Yanai, 2016)	43
Εικόνα 23: Οπτική απεικόνιση σε χάρτη στατιστικών στοιχείων σχετικά με τη γλώσσα των αναρτημένων tweets στη Σαουδική Αραβία (Ghanem et al., 2014)	43
Εικόνα 24: Οπτική απεικόνιση σε χάρτη στατιστικών στοιχείων σχετικά με τη λειτουργικό σύστημα ανάρτησης tweets στη Σαουδική Αραβία (Ghanem et al., 2014)	43
Εικόνα 25: Οπτική απεικόνιση σε χάρτη στατιστικών στοιχείων σχετικά με τη γλώσσα των αναρτημένων tweets στη πόλη Jeddah της Σαουδικής Αραβίας με βάση τον διαφορετικό αριθμό tweets (Ghanem et al., 2014)	43
Εικόνα 26: Οπτική απεικόνιση σε χάρτη στατιστικών στοιχείων σχετικά με τη γλώσσα των αναρτημένων tweets στη πόλη Jeddah της Σαουδικής Αραβίας με βάση τον διαφορετικό αριθμό χρηστών (Ghanem et al., 2014)	43

Εικόνα 27: Οπτική απεικόνιση περιοχών ανάρτησης tweets – GeoScope (Budak et al., 2014) ..	44
Εικόνα 28: Οπτική απεικόνιση περιοχών αναφοράς tweets – GeoScope (Budak et al., 2014) ...	44
Εικόνα 29: Απεικόνιση γεωγραφική χάρτη με χρήση των Google Maps API	56
Εικόνα 30: Απεικόνιση γεωγραφικού σημείου	57
Εικόνα 31: Απεικόνιση χάρτη εστίασης	57
Εικόνα 32: Ομαδοποιημένη απεικόνιση γεωγραφικών σημείων (συστοιχίες σήμανσης)	57
Εικόνα 33: Διεπαφή ιστοσελίδας οπτικής απεικόνισης tweets	72
Εικόνα 34: Γεωγραφική απεικόνιση όλων των tweets σε χάρτη (χωρίς μεγέθυνση)	74
Εικόνα 35: Γεωγραφική απεικόνιση όλων των tweets σε χάρτη (με εφαρμογή μεγέθυνσης σε επίπεδο πολιτειών)	75
Εικόνα 36: Γεωγραφική απεικόνιση όλων των tweets σε χάρτη (με εφαρμογή μεγέθυνσης σε επίπεδο πολιτείας - Καλιφόρνια).....	76
Εικόνα 37: Γεωγραφική απεικόνιση όλων των tweets σε χάρτη (με εφαρμογή μεγέθυνσης σε επίπεδο δρόμου πόλεων)	76
Εικόνα 38: Οπτικός χάρτης εστίασης απεικόνισης του συνόλου των tweets που συλλέχτηκαν ..	77
Εικόνα 39: Οπτικός χάρτης εστίασης απεικόνισης του συνόλου των tweets που συλλέχτηκαν (ανατολικές ακτές ΗΠΑ)	77
Εικόνα 40: Οπτικός χάρτης εστίασης απεικόνισης του συνόλου των tweets που συλλέχτηκαν (περιοχή Μανχάταν – Νέα Υόρκη)	78
Εικόνα 41: Λίστα με τα δέκα (10) δημοφιλέστερα hashtags	78
Εικόνα 42: Επισήμανση στον χάρτη του πρώτου χρονολογικά tweet που περιλαμβάνει συγκεκριμένο hashtag.....	79
Εικόνα 43: Οπτική απεικόνιση tweets που περιλαμβάνουν το hashtag #job στις 20/12/2016 ..	79
Εικόνα 44: Οπτική απεικόνιση tweets που περιλαμβάνουν το hashtag #job στις 21/12/2016 ..	80
Εικόνα 45: Οπτική απεικόνιση tweets που περιλαμβάνουν το hashtag #job στις 09/01/2017 ..	80
Εικόνα 46: Οπτική απεικόνιση tweets που περιλαμβάνουν το hashtag #job στις 20/12/2016 στις ανατολικές ακτές των ΗΠΑ	81
Εικόνα 47: Οπτική απεικόνιση tweets που περιλαμβάνουν το hashtag #job στις 22/01/2017 στην Ουάσινγκτον	81
Εικόνα 48: Τα δέκα (10) tweets με τα περισσότερα retweets	82
Εικόνα 49: Επισήμανση στον χάρτη της πρώτης δημοσίευσης ενός tweet	82
Εικόνα 50: Οπτική απεικόνιση όλων των retweet στον χάρτη (σημείο προέλευσης, σημείο δημοσίευσης) μέχρι τις 24/12/2016 (Θέμα tweet: Incident on #ToLine)	83
Εικόνα 51: Οπτική απεικόνιση όλων των retweet στον χάρτη (σημείο προέλευσης, σημείο δημοσίευσης) μέχρι τις 14/01/2017 (Θέμα tweet: Incident on #ToLine)	83
Εικόνα 52: Οπτική απεικόνιση όλων των retweet στον χάρτη (σημείο προέλευσης, σημείο δημοσίευσης) μέχρι τις 21/01/2017 (Θέμα tweet: Incident on #ToLine)	84
Εικόνα 53: Είδηση προσπάθειας κλοπής σπαθιού σαμουράι εντός σταθμού μετρό της Νέας Υόρκης στις 22 Δεκέμβρη 2016 (CBS, 2016)	84
Εικόνα 54: Οπτική απεικόνιση δημοφιλέστερων λέξεων ανά πόλη για το σύνολο των tweets που συλλέχτηκαν	85
Εικόνα 55: Οπτική απεικόνιση δημοφιλέστερων λέξεων ανά πόλη για το σύνολο των tweets στις 25/12/2016	86
Εικόνα 56: Οπτική απεικόνιση δημοφιλέστερων λέξεων ανά πόλη για το σύνολο των tweets στις 29/12/2016	87
Εικόνα 57: Οπτική απεικόνιση δημοφιλέστερων λέξεων ανά πόλη για το σύνολο των tweets στις 02/01/2017	87

Εικόνα 58: Οπτική απεικόνιση δημοφιλέστερων hashtags ανά πόλη για το σύνολο των tweets που συλλέχτηκαν	88
Εικόνα 59: Οπτική απεικόνιση δημοφιλέστερων hashtags ανά πόλη στις 21/12/2016	89
Εικόνα 60: Οπτική απεικόνιση δημοφιλέστερων hashtags ανά πόλη στις 25/12/2016	89
Εικόνα 61: Οπτική απεικόνιση δημοφιλέστερων hashtags ανά πόλη στις 29/12/2016	90
Εικόνα 62: Οπτική απεικόνιση δημοφιλέστερων hashtags ανά πόλη στις 13/01/2017	90
Εικόνα 63: Είδηση τροχαίου στο Σαν Φρανσίσκο στις 13 Ιανουαρίου 2017 (News, 2017)	91

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 1: Δυνατότητες ιστοτόπων οπτικής απεικόνισης διάδοσης και περιεχομένου αναρτήσεων στο Twitter	45
Πίνακας 2: Πλεονεκτήματα και μειονεκτήματα εφαρμογών οπτικής απεικόνισης περιεχομένου αναρτήσεων στο Twitter που αναπτύχθηκαν στο πλαίσιο σχετικών ερευνών	46
Πίνακας 3: Πεδία δεδομένων tweets χρησιμοποιώντας το Twitter API (Twitter, 2012b)	50
Πίνακας 4: Σύγκριση επίδοσης MySQL, Postgre, SQL Server και Oracle	55
Πίνακας 5: Οι χρήστες με τις περισσότερες δημοσιεύσεις - tweets	69
Πίνακας 6: Τα δέκα (10) δημοφιλέστερα hashtags	70
Πίνακας 7: Οι δέκα (10) δημοφιλέστερες λέξεις	70
Πίνακας 8: Οι πόλεις με τον μεγαλύτερο αριθμό αναρτήσεων - tweets	71

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

Σχήμα 1: Αριθμός χρηστών ιστοτόπων κοινωνικής δικτύωσης 2010 – 2020 (Statista, 2016)	24
Σχήμα 2: Αριθμός ενεργών χρηστών για τους δημοφιλέστερους ιστότοπους κοινωνικής δικτύωσης (Statista, 2017)	25
Σχήμα 3: Σύνδεση στο Twitter API μέσω Streaming API (Twitter, 2012a).....	50
Σχήμα 4: Σύνδεση στο Twitter API μέσω REST API (Twitter, 2012a)	52
Σχήμα 5: Αρχιτεκτονική λειτουργίας της εφαρμογής οπτικής απεικόνισης tweets	60
Σχήμα 6: Προσδιορισμός βημάτων επεξεργασίας δεδομένων σε επίπεδο βάσης δεδομένων ..	63
Σχήμα 7: Το σχήμα - σχέσεις βάσης δεδομένων ιστότοπου.....	69

ΚΕΦ.1: Εισαγωγή

1.1 Αντικείμενο έρευνας

Οι πλατφόρμες κοινωνικών δικτύων προσφέρουν στους χρήστες τη δυνατότητα να επικοινωνούν, να μοιράζονται προσωπικά γεγονότα και προσωπικές απόψεις, ειδήσεις και συζητήσεις. Οι αντίστοιχες μορφές επικοινωνίας συνοδεύονται από ένα μεγάλο όγκο πληροφοριών, των οποίων η συνάθροιση αντιπροσωπεύει σύνολα δεδομένων τεράστιας κλίμακας. Εκτιμάται πως σε κάθε λεπτό αναρτώνται: (i) 350.000 tweets στο Twitter, (ii) τετρακόσιες (400) ώρες καινούργιων βίντεο στο YouTube, (iii) οι χρήστες στο Instagram δηλώνουν πως τους αρέσουν 2,5 εκατομμύρια αναρτήσεις άλλων χρηστών και (iv) οι χρήστες του Facebook δηλώνουν πως τους αρέσουν 4 εκατομμύρια αναρτήσεις άλλων χρηστών στην πλατφόρμα του Facebook. Αν ληφθεί υπόψη πως μόνο μια ημέρα διαθέτει 1440 λεπτά, ο όγκος των δεδομένων στα κοινωνικά δίκτυα είναι ασύλληπτα τεράστιος (Team Gwava, 2016). Συνεπώς, η συλλογή και ανάλυση δεδομένων από τα κοινωνικά δίκτυα μπορεί να πραγματοποιηθεί αποτελεσματικά μονάχα με μηχανικά και όσο το δυνατόν είναι εφικτό με αυτόματα μέσα, αλγορίθμους ή εργαλεία πληροφορικής.

Τα περισσότερα κοινωνικά δίκτυα προσφέρουν διασυνδέσεις προγραμματισμού (API – Application Programming Interface) μέσω των οποίων είναι δυνατή η συλλογή σχετικών δεδομένων. Υπό το πρίσμα της συνεχούς αυξανόμενης σπουδαιότητας και μόνιμης παρουσίας των ψηφιακών επικοινωνιών στην καθημερινή πραγματικότητα, οι υπολογιστικές μέθοδοι για την απόκτηση, επεξεργασία και οπτική απεικόνιση (visualization) αντίστοιχων δεδομένων αποτελούν δραστηριότητα πρωταρχικής σημασίας για τους ερευνητές (Sechelea et al., 2016).

Το Twitter αποτελεί μια ειδική πλατφόρμα κοινωνικών δικτύων (μικρο-blogging) με βασικό σκοπό τη διευκόλυνση της επικοινωνίας μεταξύ των χρηστών του. Στις μέρες μας το Twitter έχει εξελιχθεί σε ένα κοινωνικό φαινόμενο παγκόσμιας κλίμακας, που χρησιμοποιείται ως μέσο για τη διάδοση πληροφοριών σχετικά με σημαντικά αθλητικά γεγονότα, την οργάνωση εθνικών διαμαρτυριών και κάθε μορφής ανθρώπινης δραστηριότητας μεγαλύτερης ή μικρότερης κλίμακας (Marcus et al., 2011).

Τα δεδομένα που καταχωρούνται στο Twitter είναι ετερογενή μεταξύ τους και συνήθως ιδιαίτερα μεγάλα σε μέγεθος. Από το 2009 το Twitter επιτρέπει στα μηνύματα που αναρτώνται

στο κοινωνικό δίκτυο να φέρουν γεωγραφική επισήμανση της τοποθεσίας όπου πραγματοποιείται μια ανάρτηση περιεχομένου. Η σχετική γεωγραφική επισήμανση αντιστοιχεί σε ακριβή γεωγραφικά σημεία, όπου πραγματοποιήθηκε κάποια ανάρτηση περιεχομένου. Η σχετική δυνατότητα – πληροφορία που φέρουν τα μηνύματα που αναρτώνται στο Twitter διευκολύνει την οπτική αναπαράσταση (visualization) των αντίστοιχων μηνυμάτων (Graham et al., 2014).

Επιπροσθέτως, μια πληροφορία όταν δεν είναι συνδεδεμένη με συγκεκριμένη γεωγραφική περιοχή αρκετές φορές δεν έχει ανάλογη αξία συγκριτικά με την περίπτωση κατά την οποία είναι γνωστή η περιοχή που αντιπροσωπεύει η σχετική πληροφορία (Baik et al., 2016). Την αντίστοιχη απαίτηση για τις πληροφορίες που αναρτώνται στο Twitter καλύπτει η γνώση των γεωγραφικών συντεταγμένων της περιοχής που έγινε κάποια ανάρτηση σε αυτό. Η αντίστοιχη γνώση παρέχει μεγαλύτερη αξία στην ανάλυση των αντίστοιχων tweets και επιτρέπει την εξαγωγή χωροχρονικών προσδιορισμένων συμπερασμάτων.

Η χωροχρονική ανάλυση των tweets αποτελεί το αντικείμενο μελέτης της παρούσας έρευνας και ειδικότερα η οπτική χωροχρονική απεικόνιση σε χάρτη, πληροφοριών που περιλαμβάνονται σε αναρτήσεις - tweets στο Twitter.

1.2 Σκοπός έρευνας

Σκοπός της παρούσας έρευνας είναι να συνδυάσει τις πληροφορίες που περιλαμβάνουν τα tweets που αναρτώνται στην κοινωνική πλατφόρμα Twitter με τις δυνατότητες και χρησιμότητα των χαρτών Google Maps. Πιο συγκεκριμένα, σκοπός της έρευνας είναι να απεικονίζει σε χάρτες Google Maps πληροφορίες σχετικά με τα tweets που συλλέγονται στο πλαίσιο της παρούσας έρευνας. Πέραν της συγκεντρωτικής χωροχρονικής απεικόνισης των tweets και των αντίστοιχων πληροφοριών που μπορούν να αντληθούν από ένα σύνολο tweets (δημοφιλή θέματα, δημοφιλείς λέξεις), η έρευνα αποβλέπει στην απεικόνιση της ροής διάδοσης διαφόρων tweet (retweets) στη μονάδα του χρόνου, καθώς επίσης και απεικόνισης της ροής διατήρησης ή εναλλαγής δημοφιλών θεμάτων (hashtags) και χρήσης δημοφιλών λέξεων για διάφορες πόλεις στη μονάδα του χρόνου.

Συνοψίζοντας σκοπός της παρούσας μελέτης είναι (i) η οπτική χωροχρονική απεικόνιση πληροφοριών σχετικά με αναρτήσεις μηνυμάτων – tweets στην πλατφόρμα Twitter, καθώς

οπτική απεικόνιση οιαοοσης και περιεχομενου αναρτησεων στο i twitter

επίσης και (ii) η οπτική χωροχρονική απεικόνιση της ροής εξέλιξης, διάδοσης και εναλλαγής των σχετικών πληροφοριών.

1.3 Ερευνητικό κενό – Προσφορά έρευνας

Όπως αναφέρθηκε παραπάνω το Twitter έχει εξελιχθεί σε μεγάλη πηγή δεδομένων σχετικά με το τι συμβαίνει σε ολόκληρο τον κόσμο, είτε πρόκειται για σημαντικά θέματα όπως επίκαιρες ειδήσεις ή λιγότερο σημαντικά αντικείμενα συζήτησης όπως είναι περί της μουσικής, π.χ. τζαζ, κτλ. Οι υπάρχουσες μέθοδοι ανάλυσης των σχετικών δεδομένων επικεντρώνονται σε γραπτή παρουσίαση των αντίστοιχων πληροφοριών των αποτελεσμάτων εφαρμογής σχετικών αναλύσεων δεδομένων (Moeller & Sass, 2013). Η παρούσα έρευνα δημιουργεί μια περισσότερο ενδιαφέρουσα οπτική αναπαράσταση των σχετικών δεδομένων χρησιμοποιώντας τη διαθέσιμη διεπαφή προγραμματισμού του Twitter και του γεγονότος πως αρκετά από τα μηνύματα που αναρτώνται στο Twitter φέρουν πληροφορίες σχετικά με την γεωγραφική τοποθεσία του χρήστη την αντίστοιχη στιγμή, με σκοπό να απεικονίσει χωροχρονικά σε χάρτες τις αναρτήσεις στο Twitter και των σχετικών πληροφοριών που τα συνοδεύουν.

Η παρούσα μελέτη πέραν της στατιστικής απεικόνισης χωροχρονικών πληροφοριών που σχετίζονται με αναρτήσεις στο Twitter, αναπτύσσει εργαλεία για την απεικόνιση της ροής διάδοσης ή εξέλιξης των σχετικών πληροφοριών, επιτρέποντας μια αποτελεσματικότερη απεικόνιση των αντίστοιχων πληροφοριών αναφορικά με συγκεκριμένες πόλεις ή περιοχές του πλανήτη. Με τη χρήση της ιστοσελίδας που αναπτύχθηκε στο πλαίσιο της παρούσας έρευνας, ο χρήστης έχει τη δυνατότητα χωρίς αναζήτηση σε γραπτά κείμενα αποτελεσμάτων σχετικών αναλύσεων, με μια ματιά να παρακολουθεί την εξέλιξη διάδοσης διαφόρων πληροφοριών στο Twitter και άμεσα να διαπιστώνει επίκαιρα θέματα ή συζητήσεις ανά πόλη ή γεωγραφική περιοχή.

1.5 Δομή έρευνας

Στο παρόν κεφάλαιο παρουσιάζεται το αντικείμενο της τρέχουσας μελέτης, ο σκοπός της παρούσας έρευνας, το ερευνητικό κενό που εκείνη στοχεύει να καλύψει και η δομή της εργασίας που ακολουθήθηκε.

Κατόπιν, στο επόμενο (2^ο) κεφάλαιο παρουσιάζεται τι είναι κοινωνικά δίκτυα, οι κύριοι λόγοι που οδήγησαν στην ταχεία εξάπλωσή τους, τα πλεονεκτήματα και μειονεκτήματα που

Οπτική απεικόνιση διάδοσης και περιεχομένου αναρτήσεων στο Twitter

συνοδεύει η χρήση τους και γίνεται εισαγωγή στην πλατφόρμα Twitter και τις βασικές λειτουργίες της.

Στη συνέχεια στο τρίτο (3^ο) κεφάλαιο παρουσιάζονται έρευνες που έχουν πραγματοποιηθεί με αντικείμενο την πλατφόρμα Twitter και σχετίζονται με τη οπτική απεικόνιση αναρτήσεων στο Twitter και οπτικής απεικόνισης συσχετίσεων μεταξύ των διαφόρων tweets.

Έπειτα, στο τέταρτο (4^ο) κεφάλαιο περιγράφεται η μεθοδολογία της έρευνας και πιο συγκεκριμένα το δείγμα της έρευνας που χρησιμοποιήθηκε, τα εργαλεία που χρησιμοποιήθηκαν για τη συλλογή και επεξεργασία δεδομένων από την πλατφόρμα Twitter, καθώς επίσης και των εργαλείων που χρησιμοποιήθηκαν για την οπτική απεικόνιση των tweets.

Κατόπιν, στο πέμπτο (5^ο) κεφάλαιο παρουσιάζεται η δομή – μορφή των δεδομένων που συλλέχτηκαν, η επεξεργασία που εφαρμόστηκε στα δεδομένα σε επίπεδο βάσης δεδομένων, συγκεντρωτικά στατιστικά στοιχεία των tweets που συλλέχτηκαν, περιγραφή του ιστοτόπου που αναπτύχθηκε με σκοπό την οπτική παρουσίαση των tweets και το σχήμα της βάσης δεδομένων του ιστοτόπου.

Στη συνέχεια, στο έκτο (6^ο) κεφάλαιο ακολουθεί η οπτική παρουσίαση των tweets που συλλέχτηκαν. Πιο συγκεκριμένα παρουσιάζεται το σύνολο των tweets σε χάρτη με χρήση συστοιχιών σήμανσης και οπτικού χάρτη εστίασης. Στη συνέχεια παρουσιάζεται η εξέλιξη διάδοσης των hashtags με την πάροδο του χρόνου και απεικονίζεται σε χάρτη με τη μορφή βέλους η διάδοση των retweets. Τέλος, στο ίδιο κεφάλαιο παρουσιάζεται η απεικόνιση στο χάρτη των δημοφιλέστερων λέξεων και hashtags για πόλεις των Ηνωμένων Πολιτειών και του Λονδίνου. Η σχετική απεικόνιση προβάλλεται με τη μορφή ροής εναλλαγής των δημοφιλέστερων όρων για τις σχετικές πόλεις με την πάροδο του χρόνου.

Τέλος, στο έβδομο (7^ο) κεφάλαιο παρουσιάζονται τα συμπεράσματα της έρευνας, περιορισμοί της παρούσας έρευνας, καθώς επίσης και προτάσεις για σχετικές μελλοντικές επεκτάσεις.

ΚΕΦ.2: Κοινωνικά δίκτυα και Twitter

2.1 Web 2.0

Οι πρώτες σελίδες στο διαδίκτυο χαρακτηρίζονταν από μονόδρομη επικοινωνία, κατά την οποία ήταν διαθέσιμες στους χρήστες ιστοσελίδες με αποκλειστικό σκοπό την προσπέλαση του περιεχομένου στις σχετικές ιστοσελίδες. Κατόπιν, από τις αρχές του εικοστού πρώτου αιώνα έκαναν την εμφάνισή τους ιστότοποι που επέτρεπαν την αμφίδρομη επικοινωνία, την ανάρτηση περιεχομένου από τους ίδιους τους χρήστες και την άμεση διαθεσιμότητα του σχετικού περιεχομένου προς προσπέλαση από τους χρήστες ενός ιστοτόπου. Η καινούργια μορφή ιστοτόπων με βασικό χαρακτηριστικό την αμφίδρομη επικοινωνία αποκαλείται Web 2.0 και αποτελεί μια διαφορετική φιλοσοφία σχεδίασης και υλοποίησης ιστοτόπων (Goodchild, 2007).

Βασική διαφοροποίηση των ιστοτόπων Web 2.0 σε σχέση με τους πρώτους ιστότοπους που δημοσιεύτηκαν στο διαδίκτυο (αποκαλούμενοι συχνά και ως Web 1.0), είναι η παροχή περισσότερων ευκαιριών συνεργασίας μεταξύ των χρηστών και η ενίσχυση της κοινωνικής αλληλεπίδρασης μεταξύ τους. Επιπροσθέτως, στους ιστότοπους Web 2.0, οι χρήστες μπορούν συνήθως να προσαρμόζουν την εμφάνιση των ιστοτόπων και να παρεμβαίνουν στο περιεχόμενό τους (Rouse, 2013).

Χαρακτηριστικό παράδειγμα ιστοτόπων Web 2.0 αποτελούν τα κοινωνικά δίκτυα και ειδικότερα το κοινωνικό δίκτυο Twitter, που αποτελεί το αντικείμενο μελέτης της παρούσας έρευνας (Ajjan & Hartshorne, 2008).

2.2 Κοινωνικά δίκτυα

2.2.1 Τι είναι

Τα κοινωνικά δίκτυα ορίζονται ως *“μια ομάδα εφαρμογών μέσω διαδικτύου που βασίζονται στη φιλοσοφία ιστοσελίδων Web 2.0 και που επιτρέπουν τη δημιουργία και την ανταλλαγή περιεχομένου που παράγεται από τους χρήστες”* (Kaplan & Haenlein, 2010). Αντίστοιχη εργασία (Von Muhlen & Ohno-Machado, 2012) ορίζει τα κοινωνικά δίκτυα ως *“εφαρμογές μέσω διαδικτύου που επιτρέπουν σε άτομα και κοινότητες ατόμων να δημιουργούν ομάδες - δίκτυα, να επικοινωνούν, να μοιράζονται πληροφορίες, να συνεργάζονται ή και να παίζουν με τη χρήση προγραμμάτων περιήγησης στον ιστότοπο (φυλλομετρητές) ή κινητές συσκευές”*.

Βασικά χαρακτηριστικά των κοινωνικών δικτύων (Obar & Wildman, 2015):

- πρόκειται για διαδραστικές εφαρμογές διαδικτύου Web 2.0
- το περιεχόμενό τους δημιουργείται από τους χρήστες και αναφέρεται ενδεικτικά σε αναρτήσεις κειμένου ή σχόλια, ψηφιακές φωτογραφίες ή βίντεο και δεδομένα που προκύπτουν μέσω της διαδικτυακής κοινωνικής αλληλεπίδρασης
- οι χρήστες δημιουργούν ατομικό προφίλ χρήστη το οποίο διατηρείται από το μέσο κοινωνικής δικτύωσης
- μέσω της χρήσης τους διευκολύνεται η ανάπτυξη διαδικτυακών δικτύων ατόμων - ομάδες, συνδέοντας το προφίλ ενός χρήστη με το προφίλ άλλων ατόμων ή ομάδων ατόμων

Περίπου πριν από δυο δεκαετίες, όταν ιδρύθηκε η ηλεκτρονική εγκυκλοπαίδεια Wikipedia το 2001, ο αριθμός των ανθρώπων που πίστεψαν στην ιδέα ήταν πολύ μικρός και ελάχιστοι μπορούσαν να προβλέψουν την επιτυχία του σχετικού εγχειρήματος. Σήμερα η Wikipedia αριθμεί δεκάδες εκατομμύρια άρθρα και εκατοντάδες εκατομμύρια μοναδικούς επισκέπτες ανά μήνα, γεγονός που την καθιστά, εκ των πραγμάτων ως τον βασικότερο πάροχο πληροφοριών στο διαδίκτυο, οι οποίες ελέγχονται ως προς την αξιοπιστία τους απευθείας από τους ίδιους τους χρήστες (φαινόμενο πληθοπορισμού ή crowdsourcing). Ωστόσο, ο όρος κοινωνικά δίκτυα είναι περισσότερο ευρύς και δεν περιλαμβάνει μονάχα συμμετοχικές ηλεκτρονικές εγκυκλοπαίδειες, αλλά διάφορες κατηγορίες εργαλείων κοινωνικής δικτύωσης, οι δημοφιλέστερες εκ των οποίων είναι οι εξής (Kaplan & Haenlein, 2010):

- **Ιστότοποι κοινωνικής δικτύωσης (social networks):** επιτρέπουν στους χρήστες να δημιουργούν προσωπικά προφίλ - ιστοσελίδες με περιεχόμενο που ορίζει ο χρήστης. Αυτές οι σελίδες χρησιμοποιούνται για τη σύνδεση των χρηστών με τους φίλους τους, τους συναδέλφους τους και άλλους χρήστες του ιστοτόπου κοινωνικής δικτύωσης, προκειμένου να διαμοιραστούν περιεχόμενο μεταξύ τους και για την υποστήριξη της μεταξύ τους επικοινωνίας. Παραδείγματα σχετικών ιστοτόπων αποτελούν το Facebook, το Twitter, το LinkedIn και το Google+.
- **Μπλογκς (blogs):** αποτελούν τη παλαιότερη μορφή κοινωνικών δικτύων, η οποία εξακολουθεί να είναι δημοφιλής στις μέρες μας. Η δομή των blogs είναι ανάλογη της δομής ενός ηλεκτρονικού περιοδικού και η προβολή των περιεχομένων πραγματοποιείται ανά ημέρα, μήνα, έτος ή κατηγορία περιεχομένου. Οι αναρτήσεις περιεχομένου των χρηστών

μπορούν να περιλαμβάνουν κείμενο, ήχο, εικόνα ή και βίντεο. Παραδείγματα σχετικών κοινωνικών δικτύων αποτελούν οι ιστοσελίδες WordPress και Blogger (Paquet, 2003).

- **Κοινότητες περιεχομένου (content communities):** αποτελούν εργαλεία διαμοίρασης διαφορετικού τύπου περιεχομένων, όπως ενδεικτικά αρχείων βίντεο ή εικόνας. Παραδείγματα κοινοτήτων περιεχομένου αποτελούν το YouTube και το Scribd.
- **Wikis:** αποτελούν ιστοσελίδες στις οποίες μια κοινότητα ανθρώπων διαχειρίζεται το περιεχόμενό τους. Το περιεχόμενο των σχετικών ιστοσελίδων μπορεί να προστεθεί ή να τροποποιηθεί από οποιονδήποτε χρήστη διαθέτει σχετικά δικαιώματα. Παράδειγμα ιστοσελίδας wiki αποτελεί η ηλεκτρονική εγκυκλοπαίδεια Wikipedia που το περιεχόμενό της έχει αναρτηθεί από τους ίδιους τους χρήστες της ιστοσελίδας (Ebersbach et al., 2006).
- **Εικονικοί κόσμοι (virtual worlds):** αποτελούν τρισδιάστατα εικονικά περιβάλλοντα που επιτρέπουν στους χρήστες να αλληλεπιδρούν ανεξαρτήτως της τοποθεσίας που βρίσκονται. Σε έναν εικονικό κόσμο οι χρήστες προσομοιώνονται μέσω της μορφής μιας εικονικής προσωπικότητας – αντικείμενο που αποκαλείται avatar. Μέσω των avatar οι χρήστες μπορούν να συμμετέχουν σε δραστηριότητες, να κοινωνικοποιούνται και να δημιουργούν – “χτίζουν” το περιεχόμενο – περιβάλλον του εικονικού κόσμου. Η δημοφιλέστερη πλατφόρμα εικονικού κόσμου είναι η πλατφόρμα Second Life (Harris & Rea, 2009).
- **Περιοχές συζήτησης (forum):** αποτελούν ιστοσελίδες συζήτησης μεταξύ των χρηστών, στις οποίες οι χρήστες συμμετέχουν σε συνομιλίες με διάφορα περιεχόμενα με τη μορφή μηνυμάτων. Διαφέρουν από τα chat rooms στο ότι τα μηνύματα είναι τις περισσότερες φορές μεγαλύτερα από μία γραμμή κειμένου και αποθηκεύονται μόνιμα. Επίσης, ανάλογα με το επίπεδο πρόσβασης ενός χρήστη ή τη διαρρύθμιση ενός φόρουμ, πιθανόν να απαιτείται έγκριση από τον διαχειριστή του ιστοτόπου, πριν την εμφάνισή του σχετικού μηνύματος στην αντίστοιχη ιστοσελίδα – θέμα συζήτησης. Ενδεικτικό παράδειγμα περιοχών συζήτησης στο διαδίκτυο, αποτελεί το φόρουμ στο Yahoo.
- **Μίκρο-Μπλόγκς (microblogs):** αποτελεί υπηρεσία προσφερόμενη από αρκετούς ιστότοπους κοινωνικής δικτύωσης (π.χ. Twitter και το Facebook), η οποία αναφέρεται στη δυνατότητα του χρήστη να μεταδώσει σύντομες ενημερώσεις κειμένου σχετικά με καθημερινές του δραστηριότητες, όπως ποιο βιβλίο διαβάζει, τι σκέφτεται, αν παρακολουθεί κάποια ταινία σε κινηματογράφο ή κοινοποίηση επίσκεψης σε σημείο γενικότερου ενδιαφέροντος (Zhao & Rosson, 2009).

2.2.2 Κίνητρα ανάπτυξης

Προηγούμενες έρευνες έχουν χρησιμοποιήσει τη θεωρία των κινήτρων για να εξηγήσουν τη συμπεριφορά των ατόμων σχετικά με την αποδοχή και χρήση της τεχνολογίας των πληροφοριών. Ο Deci (1975) διέκρινε τις κινητήριες δυνάμεις που διέπουν τη συμπεριφορά ενός ατόμου σε εξωγενή κίνητρα και εσωτερικά κίνητρα. Τα εξωγενή κίνητρα αναφέρονται σε υλοποίηση μιας δράσης λόγω της αντιληπτής συνεισφοράς του στην επίτευξη ενός συγκεκριμένου σκοπού, ενώ τα εσωτερικά κίνητρα αναφέρονται σε υλοποίηση μιας δράσης λόγω του ενδιαφέροντος που παρουσιάζει η σχετική δράση για τα άτομα (Davis et al., 1992).

Σχετικά με την μεγάλη ανάπτυξη – αποδοχή που παρουσιάζουν τα κοινωνικά δίκτυα οι Kim et al. (2007) προτείνουν πως οφείλεται παράλληλα σε εξωτερικά κίνητρα των χρηστών (χρησιμότητα) και σε εσωτερικά κίνητρα (απόλαυση) από τη χρήση τους (Lin & Lu, 2011). Επίσης, η χρήση των κοινωνικών δικτύων είναι εύκολη για τους χρήστες και μέσω της χρήσης τους μπορούν εύκολα να διατηρήσουν επαφή με φίλους που βρίσκονται μακριά ή έχουν καιρό να συναντήσουν.

Ωστόσο, τα πιο σπουδαία κίνητρα που έχουν οδηγήσει στην μεγάλη δημοτικότητα χρήσης των κοινωνικών δικτύων είναι κατά κύριο λόγο ψυχολογικοί. Με τα κοινωνικά δίκτυα ο άνθρωπος μπορεί να κάνει πράγματα και να λαμβάνει εμπειρίες, που δεν θα μπορούσε διαφορετικά να πραγματοποιήσει. Ενώ κάποιος είναι μόνος του στο σπίτι, έχει τη δυνατότητα να μεταφέρει τις προσωπικές τους σκέψεις άμεσα και δημόσια σε όλους – κάτι που συνιστά μια ιστορικά πρωτοφανή ψυχολογική εμπειρία (Mc Mahon, 2015). Με τον τρόπο αυτό βιώνει ο άνθρωπος το φαινόμενο που καλείται “διαδικτυακή άρση αναστολών - online disinhibition” (Suler, 2004), κατά το οποίο οι ενέργειες ενός χρήστη στο διαδίκτυο πιθανότατα δεν θα υλοποιούνταν ποτέ σε επικοινωνία πρόσωπο-με-πρόσωπο με άλλα άτομα.

2.2.3 Πλεονεκτήματα χρήσης

Τα κοινωνικά δίκτυα αποτελούν περιβάλλοντα στον κυβερνοχώρο που επιτρέπουν στα άτομα να δημιουργούν το προσωπικό τους προφίλ, να μοιράζονται κείμενο, εικόνες και βίντεο, και να συνδέονται με άλλα μέλη του κοινωνικού δικτύου μέσω ομάδων χρηστών που έχουν δημιουργηθεί από άλλους χρήστες στο αντίστοιχο κοινωνικό δίκτυο (Tapscott, 2008). Ως εκ τούτου, ένα κοινωνικό δίκτυο επιτρέπει στους χρήστες να παρουσιάσουν τον εαυτό τους σε

άλλα άτομα με τον τρόπο που επιθυμούν, να αναπτύξουν και να διατηρήσουν διαδικτυακές σχέσεις με άλλους ανθρώπους (Kane et al., 2009).

Βασικοί λόγοι - πλεονεκτήματα υπέρ της χρήσης κοινωνικών δικτύων εκ μέρους των ατόμων είναι (Whiting & Williams, 2013) τα ακόλουθα:

- δυνατότητα άμεσης επικοινωνίας - διασύνδεσης με άλλους ανθρώπους
- δυνατότητα έκφρασης και απήχησης έκφρασης σε πολλά άτομα
- δυνατότητα αλληλεπίδρασης με άλλους ανθρώπους
- δυνατότητα εξεύρεσης χρήσιμων πληροφοριών για θέματα που απασχολούν έναν χρήστη
- δυνατότητα δημιουργικής και αλληλεπιδραστικής απασχόλησης του κενού χρόνου των χρηστών εντός της ημέρας
- ως μέσο διασκέδασης

Επίσης, σε σχετική έρευνα του GlobalWebIndex (2015), διαπιστώθηκε πως οι πέντε (5) κύριοι λόγοι υπέρ της χρήσης κοινωνικών δικτύων είναι οι εξής:

- διατήρηση επαφής με φίλους – γνωστούς (55%)
- ειδησεογραφική ενημέρωση (41%)
- απλά για την κάλυψη κενού χρόνου (41%)
- για διασκέδαση (39%)
- για έκφραση – διαμοίραση μιας άποψης (39%)

2.2.4 Κίνδυνοι από τη χρήση τους

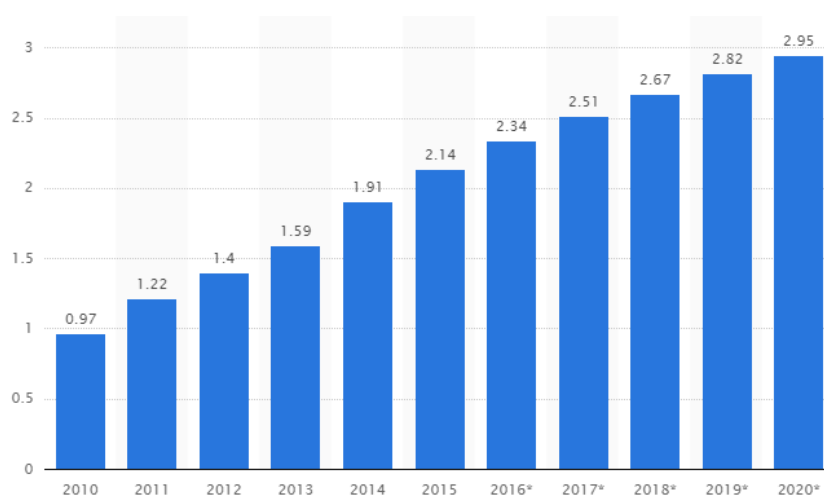
Βασικό κίνδυνο από τη χρήση κοινωνικών δικτύων αποτελεί η παραβίαση της ιδιωτικότητας των χρηστών και η πρόσβαση σε προσωπικές πληροφορίες των χρηστών χωρίς τη σύμφωνη γνώμη τους, που ενδεχομένως οι ίδιοι μπορεί να ανήρτησαν σε κάποιο κοινωνικό δίκτυο. Σε σχετική έρευνα των Krasnova et al. (2009) σε φοιτητές πανεπιστημίου του Βερολίνου, οι φοιτητές

απάντησαν πως θεωρούν ως σημαντικότερο κίνδυνο από τη χρήση κοινωνικών δικτύων την πρόσβαση σε προσωπικά δεδομένα των φοιτητών από τους καθηγητές του πανεπιστημίου, από μελλοντικούς εργοδότες, από άλλους φοιτητές, από την οικογένειά τους και γενικότερα από άτομα που συναναστρέφονται. Επιπροσθέτως, αρκετοί φοιτητές αναφέρθηκαν στον κίνδυνο παρενόχλησης ή εκφοβισμού μέσω της χρήσης κοινωνικών δικτύων.

Επίσης, σε ανάλογη έρευνα από τους Johnson et al. (2012) διαπιστώθηκε ο κίνδυνος χρήσης των κοινωνικών δικτύων για τον εντοπισμό και προσέγγιση ευαίσθητων ατόμων με σκοπό τη παραπλάνηση ή εκμετάλλευσή τους. Επιπροσθέτως, στη ίδια έρευνα διαφάνηκε ο κίνδυνος χρήσης πληροφοριών σε κοινωνικά δίκτυα σχετικά με τη ζωή των ατόμων με σκοπό τη διευκόλυνση εγκληματιών για την πραγματοποίηση παράνομων πράξεων. Τέλος, από την ίδια έρευνα διαφάνηκε ο κίνδυνος χρήσης των κοινωνικών δικτύων, ως μέσο αξιολόγησης υποψήφιων εργαζομένων από τους εργοδότες.

2.2.5 Στοιχεία χρήσης

Ο συνολικός εκτιμώμενος αριθμός χρηστών ιστοτόπων κοινωνικής δικτύωσης για το έτος 2017 εκτιμάται στο επίπεδο των 2,5 δισεκατομμυρίων χρηστών σε παγκόσμιο επίπεδο (Statista, 2016):

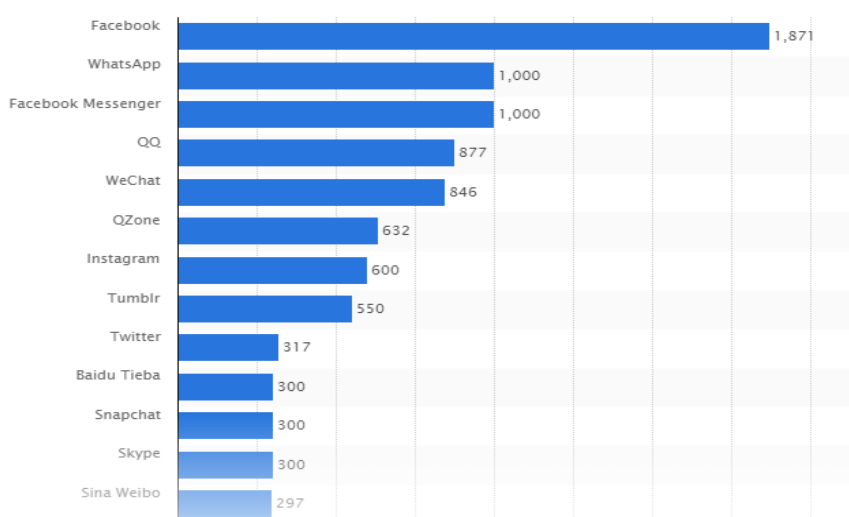


Σχήμα 1: Αριθμός χρηστών ιστοτόπων κοινωνικής δικτύωσης 2010 – 2020 (Statista, 2016)

Όπως παρατηρείται από το παραπάνω σχήμα, ο αριθμός χρηστών κοινωνικών δικτύων παρουσιάζει αύξουσα πορεία και μέσα σε πέντε χρόνια (από το 2010 μέχρι το 2015) ο αριθμός χρηστών έχει διπλασιαστεί (από 1 δισεκατομμύριο σε 2 δισεκατομμύρια χρήστες). Ο σχετικός

αριθμός χρηστών εκτιμάται πως θα φτάσει τα 3 δισεκατομμύρια χρήστες το 2020. Το γεγονός αυτό αντανakλά τη δυναμική των κοινωνικών δικτύων σε παγκόσμιο επίπεδο και το ερευνητικό ενδιαφέρον ανάλυσης των μηνυμάτων που αναρτώνται σε εκείνα.

Ο δημοφιλέστερος ιστότοπος κοινωνικής δικτύωσης είναι το Facebook, με τον αριθμό ενεργών χρηστών του να πλησιάζει τα 2 δισεκατομμύρια χρήστες, ενώ το Twitter που αποτελεί αντικείμενο της παρούσας έρευνας διαθέτει περί των 317 εκατομμύρια χρήστες σε παγκόσμιο επίπεδο (Statista, 2017):



Σχήμα 2: Αριθμός ενεργών χρηστών για τους δημοφιλέστερους ιστότοπους κοινωνικής δικτύωσης (Statista, 2017)

Ο μεγάλος αριθμός χρηστών του Twitter σε παγκόσμιο επίπεδο, αντανakλά τη δυναμική του κοινωνικού δικτύου Twitter και το ερευνητικό ενδιαφέρον ανάλυσης των μηνυμάτων που αναρτώνται στο αντίστοιχο κοινωνικό δίκτυο.

2.3 Το φαινόμενο Twitter

2.3.1 Τι είναι

Το Twitter αποτελεί ένα εργαλείο – ιστότοπο κοινωνικής δικτύωσης που έχει παρουσιάσει εντυπωσιακή ανάπτυξη από τη στιγμή που ιδρύθηκε το 2006 μέχρι και σήμερα. Πρακτικά, το Twitter αποτελεί μια κοινότητα ανταλλαγής περιεχομένου ή/και συναισθημάτων. Οι χρήστες συνήθως μοιράζονται χρήσιμες πληροφορίες όπως είναι τα προσωπικά τους ενδιαφέροντα ή τις ενασχολήσεις τους με διάφορα αντικείμενα (Wanichayarong et al., 2011). Ως αποστολή της πλατφόρμας Twitter, όπως αναφέρεται στην ιστοσελίδα του οργανισμού, είναι *“η παροχή της*

δυνατότητας σε όλους τους χρήστες άμεσης διαμοίρασης ιδεών και πληροφοριών, χωρίς κανένα εμπόδιο” (Twitter, 2013).

Αν και το Twitter είναι σχεδιασμένο με γνώμονα τη λειτουργία ενός κινητού τηλεφώνου, πολλοί από τους χρήστες του Twitter προσπελάζουν το περιεχόμενό του μέσω υπηρεσιών διαδικτύου ή άλλες εφαρμογές λογισμικού (third-party software applications) (Lotan et al., 2011).

Ενώ οι λόγοι που οι χρήστες επιλέγουν τη διασύνδεσή τους στο Twitter διαφέρουν μεταξύ τους, ορισμένες κοινές περιοχές ενδιαφέροντος για τους εγγεγραμμένους χρήστες στο Twitter είναι οι παρακάτω (Twitter, 2015):

- **Ειδήσεις και πολιτική:** άμεση ενημέρωση για τοπικές ειδήσεις και ειδήσεις σε παγκόσμιο επίπεδο. Δυνατότητα άμεσης παρακολούθησης εξέλιξης πολιτικών γεγονότων και δυνατότητα έκφρασης προσωπικής άποψης για γεγονότα που λαμβάνουν τόπο σε τοπικό ή παγκόσμιο επίπεδο.
- **Αθλητισμός:** ανάλογη ενημέρωση περί αθλητικών εκδηλώσεων μπορούν να λάβουν οι χρήστες με χρήση της εφαρμογής Twitter και διασύνδεσή τους με σχετικές ενημερωτικές ομάδες σχετικές με τον αθλητισμό.
- **Καλλιτεχνικά ενδιαφέροντα:** ενημέρωση σχετικά με καλλιτεχνικά γεγονότα, δυνατότητα διασύνδεσης με αγαπημένους καλλιτέχνες και δυνατότητα επικοινωνίας μαζί τους.
- **Διασύνδεση με άτομα επιρροής:** δυνατότητα διασύνδεσης με ηγέτες, φιλοσόφους, εμπειρογνώμονες του κλάδου που ενδιαφέρουν κάποιον χρήστη για τη λήψη χρήσιμων πληροφοριών και δυνατότητα επικοινωνίας μαζί τους.
- **Χρηστικές πληροφορίες:** λήψη πληροφοριών ή ενημερώσεων για θέματα ενδιαφέροντος ενός χρήστη, όπως είναι η επαφή με το τμήμα εξυπηρέτησης ενός οργανισμού, χωρίς να απαιτείται σχετική κλήση τηλεφώνου προς τον οργανισμό.

Η πρόσβαση των χρηστών στο Twitter μπορεί να πραγματοποιηθεί μέσω της ιστοσελίδας του Twitter, με μηνύματα στο κινητό τηλέφωνο (SMS) ή μέσω σχετικών εφαρμογών για κινητά τηλέφωνα. Εκτός των φυσικών ατόμων, λογαριασμό στο Twitter μπορούν να διαθέτουν ένας οργανισμός, μια επιχείρηση ή διάφορες ενώσεις ατόμων.

Σε αντίθεση με άλλα κοινωνικά δίκτυα που απαιτείται από τους χρήστες να επιβεβαιώσουν τη φιλία με άλλους χρήστες και την παραχώρηση δυνατότητας άμεσης αλληλεπίδρασής μαζί τους, το Twitter εφαρμόζει διαφορετικό μοντέλο κοινωνικής δικτύωσης που βασίζεται στην έννοια του “ακολουθεί (*follows*)”. Σύμφωνα με τη σχετική φιλοσοφία ο κάθε χρήστης έχει τη δυνατότητα να επιλέξει ποιους επιθυμεί να ακολουθήσει χωρίς να ζητήσει οποιαδήποτε άδεια και να παρακολουθεί άμεσα τις αναρτήσεις του στην πλατφόρμα του Twitter. Αντίστοιχα, ένας χρήστης μπορεί επίσης να ακολουθείται από άλλους χρήστες χωρίς την ανάγκη παραχώρησης σχετικής άδειας (Wend et al., 2010).

Το Twitter έχει αναδειχθεί ως ένα δημοφιλές κοινωνικό σύστημα όπου οι χρήστες μοιράζονται και συζητούν για τα πάντα, όπως μεταξύ άλλων, των ειδήσεων της επικαιρότητας, ανέκδοτα, την άποψη τους για κοινωνικά γεγονότα, καθώς επίσης και σχετικά με την ψυχολογική τους διάθεση. Με ένα απλό περιβάλλον, όπου μόνο μηνύματα με 140 χαρακτήρες μπορούν να αναρτηθούν, το Twitter γίνεται ολοένα και περισσότερο ένα σύστημα που εξυπηρετεί τη λήψη πληροφοριών άμεσα και σε πραγματικό χρόνο. Όταν ένας χρήστης δημοσιεύει ένα tweet, εκείνο αμέσως παρουσιάζεται σε όσους τον ακολουθούν, επιτρέποντας την άμεση διάδοση μιας ανάρτησης μέσω των «ακόλουθων»/«οπαδών» ενός χρήστη (Benevenuto et al., 2010).

2.3.2 Το tweet

Το tweet αποτελεί τη βάση λειτουργίας της πλατφόρμας Twitter, καθώς αποτελεί το μοναδικό τρόπο επικοινωνίας μεταξύ των χρηστών της πλατφόρμας, εάν εξαιρεθεί η δυνατότητα αποστολής προσωπικών μηνυμάτων μεταξύ των χρηστών της πλατφόρμας. Αρχικά τα tweets μπορούσαν να περιλαμβάνουν μονάχα κείμενο. Πλέον, τα tweets μπορούν να περιλαμβάνουν εικόνες, καθώς επίσης και μικρής διάρκειας αρχεία βίντεο.

Το tweet αποτελεί έναν αργκό όρο για την απάντηση στην ερώτηση “*Τι κάνεις αυτή τη στιγμή;*”. Ένα tweet (τιτίβισμα) είναι ουσιαστικά αυτό που πληκτρολογεί ο χρήστης στο σχετικό ερώτημα, χρησιμοποιώντας το πολύ 140 χαρακτήρες. Οι άνθρωποι αναρτούν με τη χρήση των tweet προσωπικά μηνύματα, τυχαίες σκέψεις, ή οτιδήποτε ταιριάζει στην προσωπικότητα ενός χρήστη και τη δράση του τη συγκεκριμένη στιγμή.

Το γεγονός πως τα περισσότερα tweets αποτελούνται από μια ή δυο προτάσεις, καθιστούν το Twitter ιδανικό εργαλείο για την εφαρμογή τεχνικών μηχανικής μάθησης (παρότι απαιτητικών

λόγω της αδόμητης φύσης του περιεχομένου που αναρτάται) για την ανάλυση του περιεχομένου των διαφόρων tweets.

2.3.3. To retweet

Προκειμένου ένας χρήστης να στείλει κάποιο tweet στην πλατφόρμα, δεν χρειάζεται να δημιουργήσει αναγκαστικά ένα καινούργιο - νέο περιεχόμενο - μήνυμα. Οι χρήστες μπορούν πολύ απλά να μοιραστούν τα tweets που άλλοι άνθρωποι έχουν δημοσιεύσει στην πλατφόρμα, προωθώντας τα σχετικά μηνύματα σε άλλους χρήστες. Η σχετική διαδικασία καλείται retweet.

Όταν κάποιος χρήστης επιλέγει να κάνει retweet σε μια ανάρτηση ενός άλλου χρήστη στο Twitter, αυτό συνήθως δηλώνει πως ο εν λόγω χρήστης θεωρεί το περιεχόμενο του tweet χρήσιμο και θεωρεί πως αν το κάνει retweet θα ωφελήσει άτομα που ακολουθούν τον ίδιο στην πλατφόρμα του Twitter. Η σχετική πράξη έχει ως αποτέλεσμα την αύξηση της εμβέλεια ενός tweet στην πλατφόρμα του Twitter και πιθανώς να οδηγήσει και στην αύξηση των χρηστών που ακολουθούν τον χρήστη που έκανε την αρχική δημοσίευση.

2.3.4 To follow

Για να στείλει ή να λάβει ένας χρήστης ένα tweet, θα πρέπει καταρχάς να έχει δημιουργήσει έναν δωρεάν λογαριασμό στο Twitter. Επίσης θα πρέπει να δηλώσει επαφές με άλλους λογαριασμούς χρηστών στο Twitter, διαφορετικά δεν υπάρχει κανένας παραλήπτης στα μηνύματα – tweet που αναρτά στην πλατφόρμα.

Ως ένα κοινωνικό δίκτυο, η αποτελεσματικότητα χρήσης του Twitter εξαρτάται από τον αριθμό οπαδών (followers) ενός χρήστη. Όταν ένα χρήστης επιλέγει να ακολουθήσει (follow) έναν άλλο χρήστη του Twitter, τα tweets του χρήστη εμφανίζονται με αντίστροφη χρονολογική σειρά στην κύρια σελίδα Twitter του χρήστη που επέλεξε να ακολουθήσει τον χρήστη που απέστειλε το σχετικό tweet. Αν κάποιος χρήστης ακολουθεί 50 άτομα, θα λαμβάνει ενημέρωση στη σελίδα του στο Twitter, σχετικά με τα tweets που ανάρτησαν οι σχετικοί χρήστες, όπως: ενημερώσεις δράσεις, προτεινόμενους συνδέσμους στο διαδίκτυο, προτάσεις για μουσική, προτάσεις για τον κινηματογράφο ή γενικότερες πολιτικές ή φιλοσοφικές απόψεις των χρηστών.

Συνοψίζοντας, ακολουθώντας άλλους χρήστες, ένα άτομο μπορεί να ενημερώνεται για τις αναρτήσεις tweets των σχετικών ατόμων ή ομάδων ατόμων. Επίσης, τα άτομα που ακολουθούν

έναν χρήστη λαμβάνουν ενημέρωση για τα tweets που δημοσιεύει ο σχετικός χρήστης στην πλατφόρμα του Twitter.

2.3.5 To hashtag

Ένας τρόπος για την κατηγοριοποίηση και την ομαδοποίηση μηνυμάτων, αποτελεί το σύμβολο δέσης που ακολουθείται από το θέμα ενός tweet. Για παράδειγμα, αν ένας χρήστης θέλει να αναρτήσει στην πλατφόρμα ένα tweet σχετικά με την αγορά ενός καινούργιου ρολογιού, μπορείτε να περιλάβει το κείμενο #newclock #longines ως μέρος του μηνύματος. Οι σχετικές αναρτήσεις καλούνται hashtags και έχουν ως αποτέλεσμα την αποτελεσματικότερη αναζήτηση μηνυμάτων από άλλους χρήστες. Πρακτικά για τη δημιουργία ενός hashtag αρκεί η προσθήκη του χαρακτήρα δέσης '#' μπροστά από μια λέξη στο μήνυμα – tweet ενός χρήστη.

Μια από τις σημαντικότερες δυνατότητες που προσφέρει το hashtag, είναι η δυνατότητα που παρέχει για την δημιουργία κοινοτήτων ανθρώπων με κοινά ενδιαφέροντα, επιτρέποντας την ευκολότερη και αποτελεσματικότερη διαμοίραση πληροφοριών που ενδιαφέρουν τα μέλη της κοινότητας.

ΚΕΦ.3: Πρότερες σχετικές έρευνες οπτικής απεικόνισης αναρτήσεων στο Twitter

Στη σύγχρονη εποχή των μεγάλων δεδομένων (big data) είναι όλο και πιο δύσκολο για έναν ερευνητή - αναλυτή να εξαγάγει ουσιαστική γνώση - πληροφορίες από έναν τεράστιο όγκο δεδομένων. Παράδειγμα μεγάλων δεδομένων αποτελούν αναμφισβήτητα τα δεδομένα των κοινωνικών δικτύων. Ενδεικτικά, μόνο τον Ιούνιο του 2012, μόνο το Twitter ανέφερε 340 εκατομμύρια tweets την ημέρα από 170 εκατομμύρια ενεργούς χρήστες. Στα αντίστοιχα tweets είναι δυνατόν να εντοπιστούν συζητήσεις σχετικά με γεγονότα που συγκέντρωσαν τη παγκόσμια προσοχή την εκάστοτε περίοδο. Τα δεδομένα των κοινωνικών δικτύων παρέχουν πλούσιες και εκτεταμένες πληροφορίες σχετικά με τον παλμό των γεγονότων και συζητήσεων παγκοσμίως ή ανά γεωγραφική περιοχή. Ωστόσο, προκειμένου να αποκτήσει κάποιος πρόσβαση σε σχετικές πληροφορίες, θα πρέπει να γνωρίζει ακριβώς τι ψάχνει (Morstatter et al, 2013).

Για τη διευκόλυνση παρουσίασης χαρακτηριστικών - τάσεων που περιλαμβάνουν τα tweets που δημοσιεύονται στο Twitter έχουν αναπτυχθεί ειδικοί ιστότοποι με σχετικό περιεχόμενο και έχουν υλοποιηθεί αντίστοιχες ειδικότερες έρευνες.

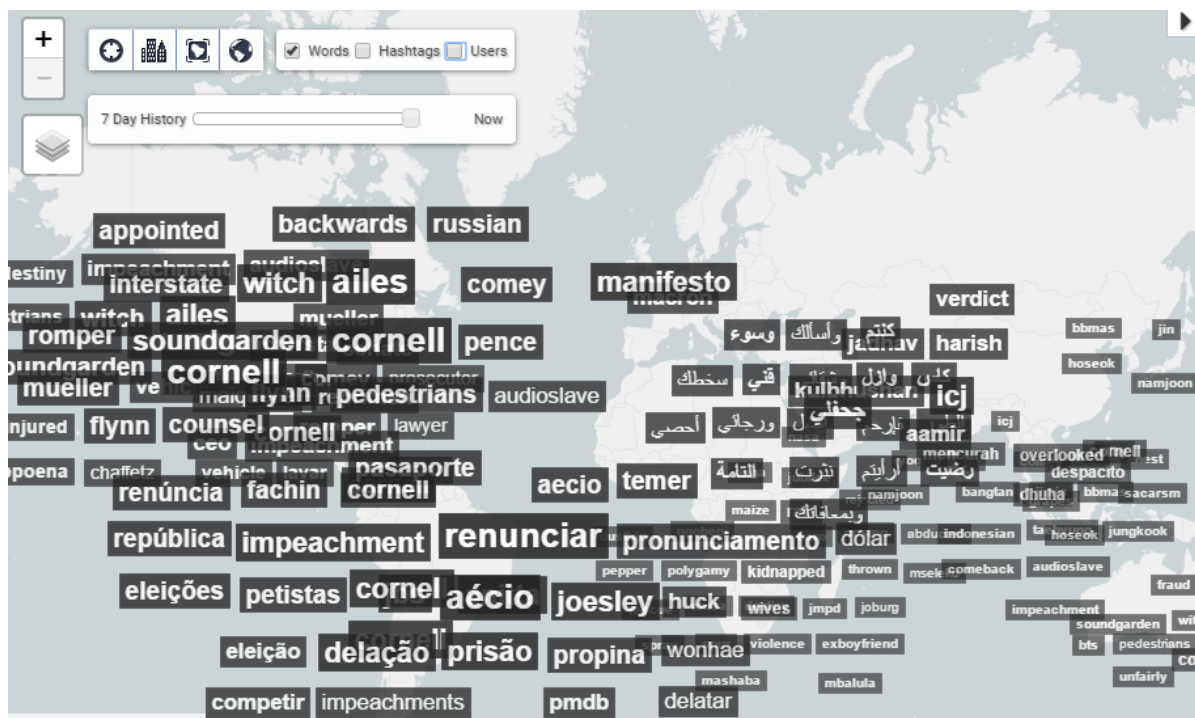
3.1 Διαθέσιμοι Ιστότοποι

Ιστοσελίδες όπως το TrendsMap, το OneMillionTweetMap και το TweetMap, με βάση διάφορες επιλογές του χρήστη παρέχουν ποικιλία οπτικής παρουσίασης και ανάλυσης των tweets που δημοσιεύονται στο Twitter. Στη συνέχεια ακολουθεί περιγραφή των δυνατοτήτων οπτικοποίησης των αντίστοιχων ιστοσελίδων.

3.1.1 TrendsMap

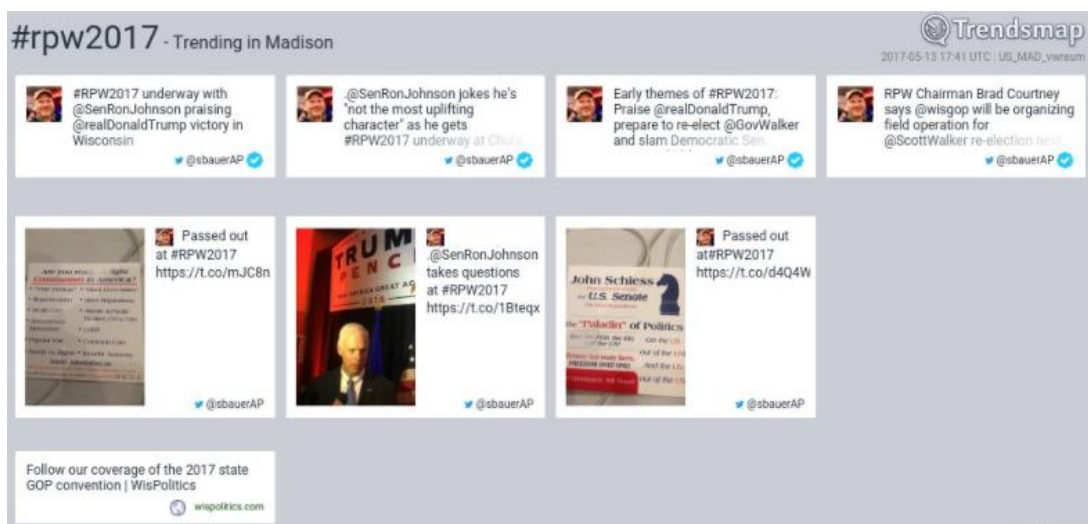
Το TrendsMap δημιουργήθηκε το 2009 και αποτελεί εφαρμογή επί πληρωμής, η οποία παρακολουθεί τα θέματα – tweets που αναρτώνται στο Twitter κατά γεωγραφική περιοχή, χρησιμοποιώντας δεδομένα που λαμβάνει από το Twitter API και στη συνέχεια τοποθετεί τα αντίστοιχα tweets σε χάρτη Google Map, όπου οι χρήστες μπορούν να φιλτράρουν τα tweets ανά πόλη ή περιοχή και να διαπιστώνουν τις δημοφιλέστερες λέξεις ή hashtags σε πραγματικό χρόνο ή για συγκεκριμένες χρονικές στιγμές στο παρελθόν. Σε καθημερινή βάση το TrendsMap αναλύει πάνω από πέντε εκατομμύρια tweets που αναρτώνται στο Twitter.

Η παρακάτω εικόνα απεικονίζει τα δημοφιλέστερα hashtags στην Αττική στις 18 Μαΐου 2017, ώρα 23:15:



Παρατηρείται πως στις 18 Μαΐου 2017 σε αρκετές περιοχές της Αμερικής, κυριαρχεί η χρήση της λέξη Cornell. Κατόπιν αναζήτησης στο διαδίκτυο διαπιστώθηκε πως στις 18 Μαΐου 2017 αυτοκτόνησε ο δημοφιλής τραγουδιστής στην Αμερική Chris Cornell, δικαιολογώντας την εμφάνιση του επιθέτου του καλλιτέχνη ως τη δημοφιλέστερη λέξη χρήσης στις 18 Μαΐου σε αρκετές περιοχές της Αμερικής (BBC News, 2017):

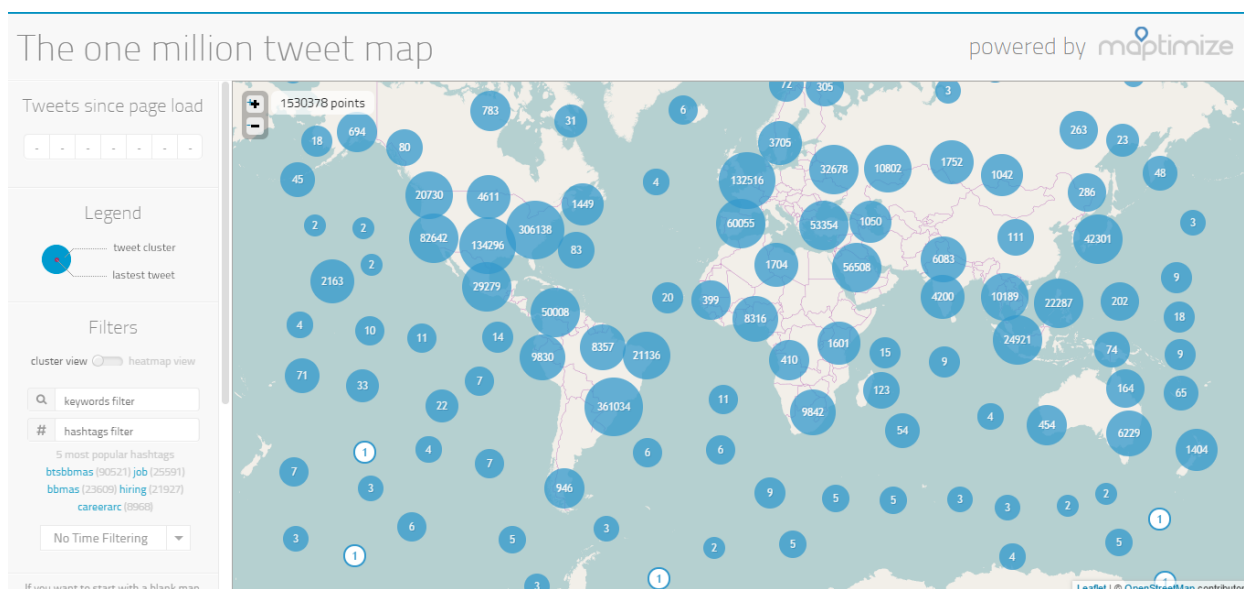
Τέλος, κάνοντας κλικ στα δημοφιλέστερα hashtags ή λέξεις, εμφανίζονται σχετικές πληροφορίες των αντίστοιχων tweets:



Εικόνα 4: Τελευταίες αναρτήσεις tweets για το hashtag #rpw2017

3.1.2 OneMillionTweetMap

Στο συγκεκριμένο ιστότοπο απεικονίζονται σε χάρτη Google Map τα τελευταία 1 εκατομμύρια tweets, όπως ελήφθησαν από το Twitter API:



Εικόνα 5: Οπτική απεικόνιση των τελευταίων 1 εκατομμυρίων tweets που ελήφθησαν από το Twitter API στην ιστοσελίδα OneMillionTweetMap

Επιπροσθέτως, ο χρήστης έχει τη δυνατότητα να αναζητήσει tweets που περιλαμβάνουν συγκεκριμένο hashtag ή συγκεκριμένες λέξεις κλειδιά. Επίσης, προβάλλονται στο αριστερό τμήμα της σελίδας τα πέντε δημοφιλέστερα hashtags για το σύνολο των τελευταίων 1 εκατομμυρίων tweets. Το πρωί στις 19 Μαΐου 2017 τα πέντε (5) δημοφιλέστερα hashtags σε παγκόσμιο επίπεδο ήταν τα εξής:

Οπτική απεικόνιση διάδοσης και περιεχομένου αναρτήσεων στο Twitter

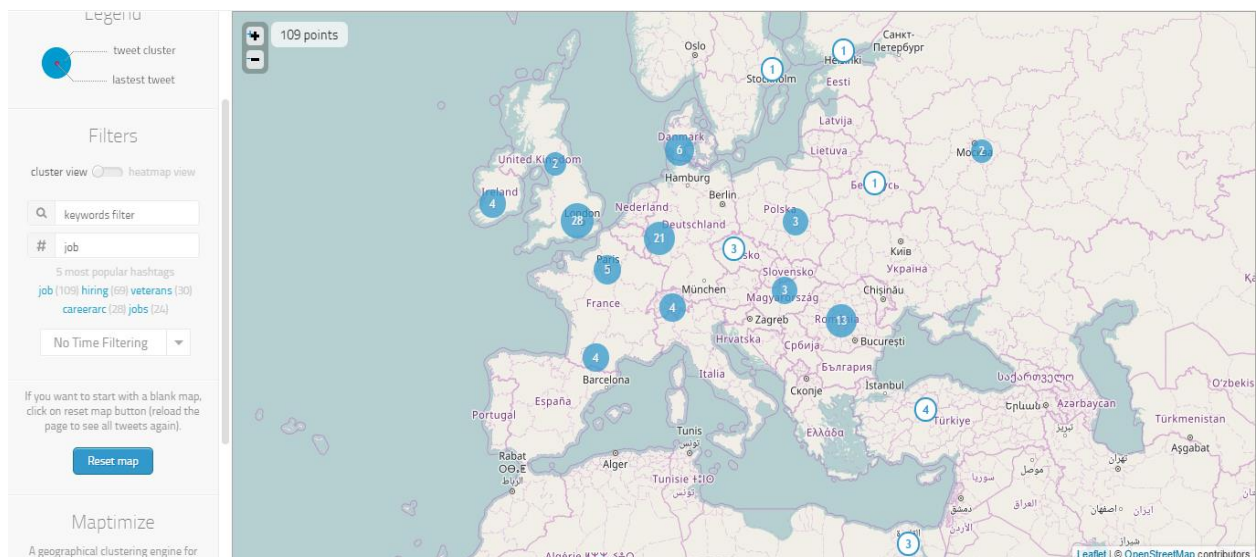
- btsbbmas (90.103 tweets)
- job (25.495 tweets)
- bbmas (23.526 tweets)
- hiring (21.841 tweets)
- careerarc (8.933 tweets)

Στη παρακάτω εικόνα απεικονίζονται σε παγκόσμιο επίπεδο τα tweets που περιλαμβάνουν το hashtag #job στις 19 Μαΐου 2017:



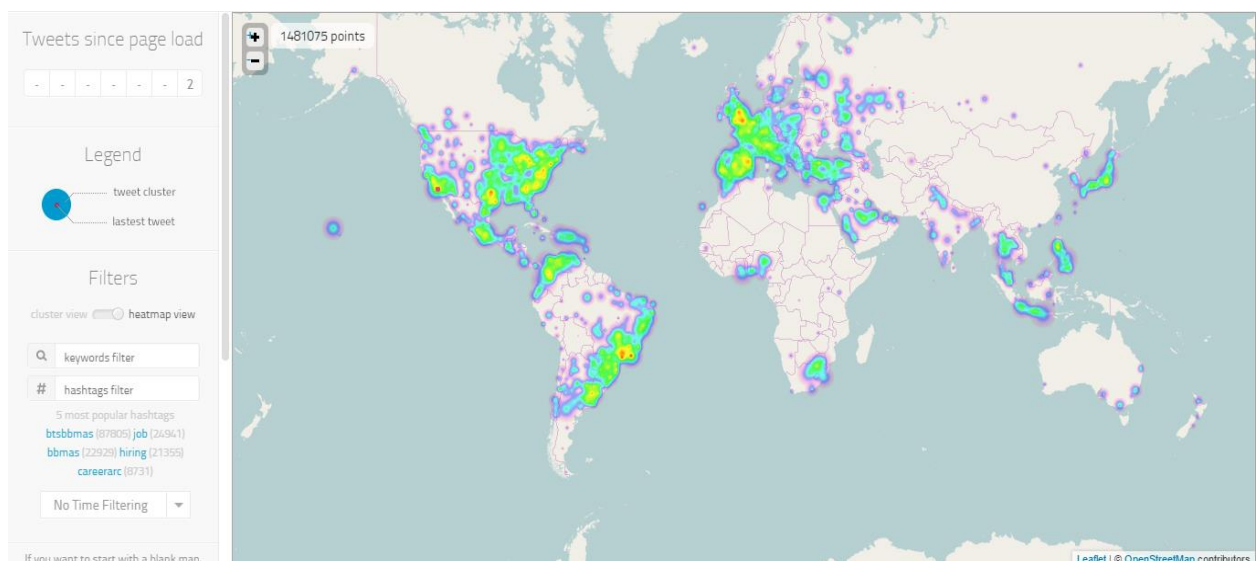
Εικόνα 6: Οπτική απεικόνιση των tweets που περιλαμβάνουν το hashtag #job στην ιστοσελίδα OneMillionTweetMap (19/05/2017)

Πατώντας ο χρήστης στις αντίστοιχες ετικέτες με τον αριθμό tweets που αντιστοιχεί στις διάφορες περιοχές, ο χάρτης εστιάζει αυτόματα στη συγκεκριμένη περιοχή και εμφανίζονται με μεγαλύτερη λεπτομέρεια οι περιοχές δημοσίευσης των αντίστοιχων tweets. Στην παρακάτω εικόνα απεικονίζονται οι αναρτήσεις tweets που περιλαμβάνουν το hashtag #job στην Ευρώπη στις 19 Μαΐου 2017:



Εικόνα 7: Οπτική απεικόνιση των tweets που περιλαμβάνουν το hashtag #job στην Ευρώπη στην ιστοσελίδα OneMillionTweetMap (19/05/2017)

Επίσης, η ίδια ιστοσελίδα απεικονίζει με τη μορφή χάρτη εστίασης (heatmap) τα 1 εκατομμύρια tweets που ελήφθησαν τελευταία από το Twitter API:



Εικόνα 8: Οπτική απεικόνιση με χάρτη εστίασης (heatmap) των τελευταίων 1 εκατομμυρίων tweets που ελήφθησαν από το Twitter API στην ιστοσελίδα OneMillionTweetMap

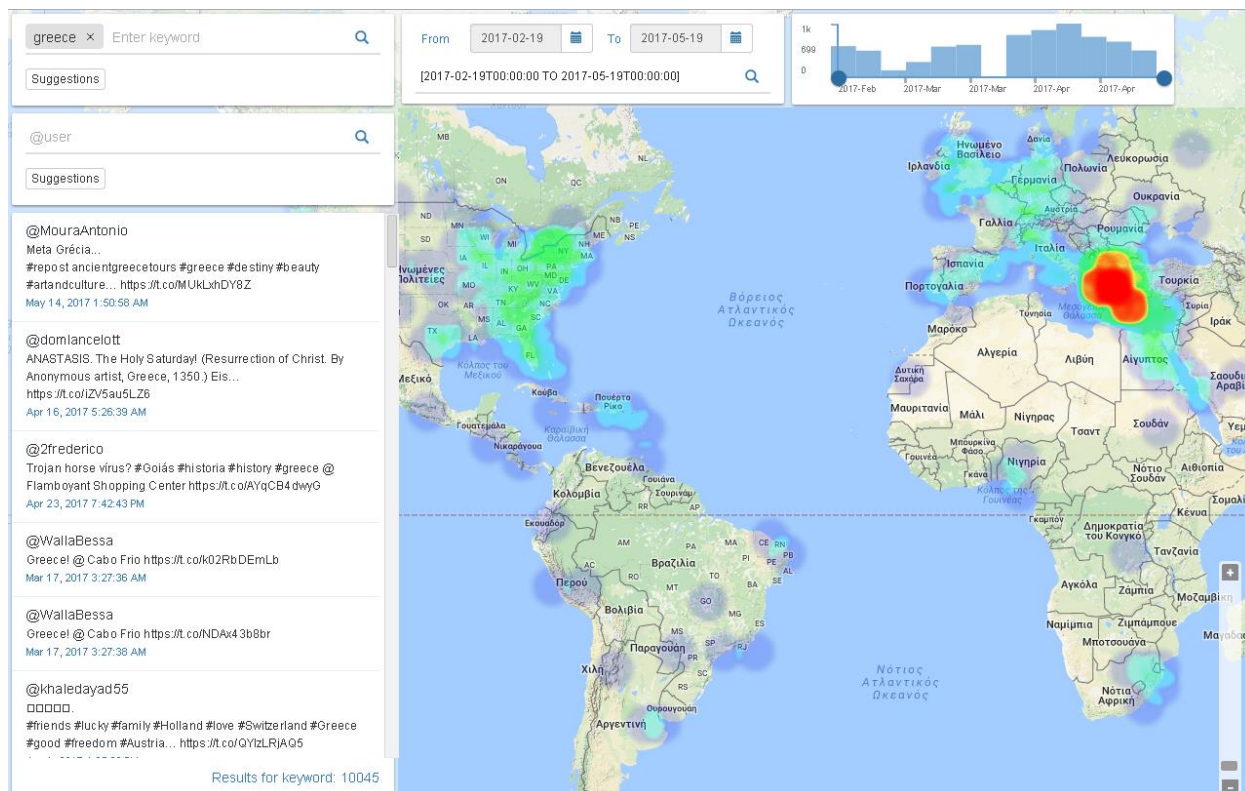
3.1.3 TweetMap

Αποτελεί εργαλείο – ιστότοπο που επιτρέπει τη δήλωση λέξεων κλειδιών ή χρηστών, καθώς επίσης και τη δήλωση χρονικής περιόδου και στη συνέχεια εμφανίζονται στο χρήστη με τη μορφή χάρτη εστίασης τα σχετικά tweets που δημοσιεύτηκαν την επιλεγμένη χρονική περίοδο. Επιπροσθέτως, σε διάγραμμα εμφανίζεται ο αριθμός των σχετικών tweets που δημοσιεύτηκαν κάθε μέρα για την επιλεγμένη χρονική περίοδο.

Το αντίστοιχο εργαλείο δημιουργήθηκε από φοιτητή του Πανεπιστημίου του MIT (Todd Mostak, 2012) στο πλαίσιο των μεταπτυχιακών του σπουδών και αποτέλεσε μετέπειτα τη βάση Οπτική απεικόνιση διάδοσης και περιεχομένου αναρτήσεων στο Twitter

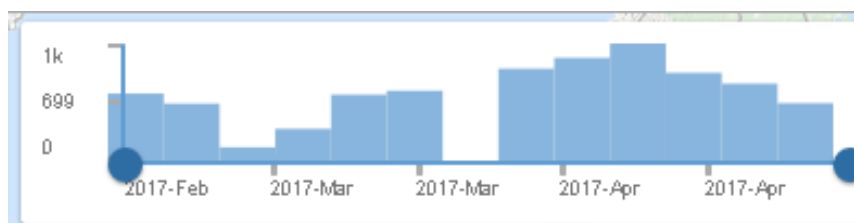
ανάπτυξης του συστήματος MapD (Massively Parallel Database), ενός συστήματος διαχείρισης βάσεων δεδομένων (Murphy, 2013).

Στην παρακάτω εικόνα απεικονίζονται σε χάρτη εστίασης τα tweets που δημοσιεύτηκαν τη περίοδο από 19/02/2017, έως 19/05/2017 και περιλαμβάνουν τη λέξη Greece:



Εικόνα 9: Οπτικοποίηση των tweets που περιλαμβάνουν τη λέξη Greece για την περίοδο 19/02/2017 - 19/05/2017 στο σύστημα TweetMap

Όπως διαπιστώνεται στην παραπάνω εικόνα, στο πάνω - δεξιά τμήμα της οθόνης εμφανίζεται το διάγραμμα του αριθμού δημοσιεύσεων σχετικών tweets κάθε μέρα για την επιλεγμένη χρονική περίοδο:



Εικόνα 10: Αριθμός δημοσιεύσεων tweets που περιλαμβάνουν τη λέξη Greece τη περίοδο 19/02/2017 - 19/05/2017 από το σύστημα TweetMap

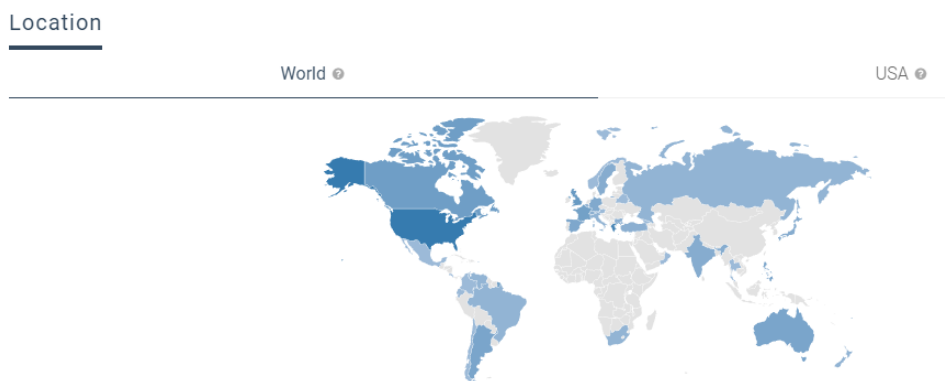
3.1.4 KeyHole

Η εφαρμογή KeyHole παρέχει στους χρήστες πληροφορίες σχετικά με hashtag, χρήστες, λέξεις κλειδιά και διευθύνσεις στο διαδίκτυο που δηλώνει ο ίδιος ο χρήστης. Πέραν των σχετικών πληροφοριών που προβάλλονται από την εφαρμογή KeyHole (αριθμό tweets, αριθμό χρηστών,

Οπτική απεικόνιση διάδοσης και περιεχομένου αναρτήσεων στο Twitter

αριθμό προβολών, σχετικά hashtags, σχετικές λέξεις, λειτουργικό σύστημα χρήστη, συναισθηματική ανάλυση tweets, δημογραφικά στοιχεία χρηστών), η εφαρμογή απεικονίζει σε διεπαφή χάρτη την τοποθεσία ανάρτησης των σχετικών tweets, σύμφωνα με τα κριτήρια αναζήτησης του χρήστη.

Στο παρακάτω διάγραμμα απεικονίζονται σε διεπαφή χάρτη οι αναρτήσεις tweets που περιλαμβάνουν το hashtag #greece σε πραγματικό χρόνο στις 20 Μαΐου 2017:



Εικόνα 11: Οπτική απεικόνιση σε διεπαφή χάρτη των tweets που περιλαμβάνουν το hashtag #greece στις 20/05/2017 στην εφαρμογή KeyHole

3.2 Σχετικές έρευνες

Μετά (i) τη διάθεση στο κοινό του Twitter API και της δυνατότητας συλλογής αναρτήσεων tweets σε πραγματικό χρόνο από τους χρήστες – ερευνητές, καθώς επίσης και (ii) της δυνατότητας συμπερίληψης γεωγραφικών πληροφοριών του σημείου ανάρτησης ενός tweet στο Twitter, αρκετές έρευνες υλοποιήθηκαν με αντικείμενο την οπτικοποίηση αναρτήσεων tweets και πληροφοριών που εκείνα περιλαμβάνουν. Ακολουθεί αναφορά σε ενδεικτικές σχετικές έρευνες που υλοποιήθηκαν μετά το 2008.

3.2.1 Οπτικοποίηση ειδήσεων σε χάρτη

Σε σχετική έρευνα των Sankaranarayanan et al. (2009) χρησιμοποιήθηκε το περιεχόμενο αναρτήσεων στο Twitter με σκοπό τη διαπίστωση των τελευταίων σημαντικών ειδήσεων ανά περιοχή και η απεικόνισή τους σε διεπαφή χάρτη για την εύκολη ανάγνωση των αντίστοιχων ειδήσεων. Το σύστημα που αναπτύχθηκε ονομάστηκε TwitterStand και λειτουργεί αποκλειστικά μόνο με τα tweets που δημοσιεύονται από τους χρήστες του Twitter για τον προσδιορισμό και απεικόνιση των τρεχουσών ειδήσεων σε χάρτη. Ως δείκτες προσδιορισμού των σημαντικότερων ειδήσεων χρησιμοποιήθηκε ο αριθμός των σχετικών tweets και η ταχύτητα εξάπλωσής τους. Όπως αναφέρουν οι ερευνητές βασικό πλεονέκτημα του εργαλείου που ανέπτυξαν αποτελεί η

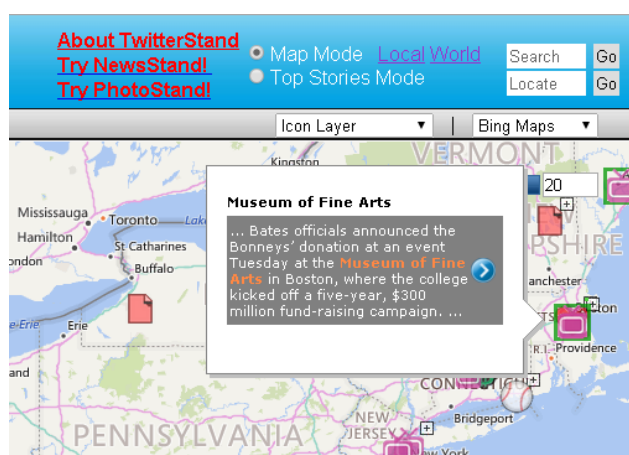
ταχύτητα παρακολούθησης των επίκαιρων ειδήσεων εκ μέρους των χρηστών και η έλλειψη αναμονής της επίσημης μετάδοσής τους από τα μέσα ενημέρωσης.

Οι ειδήσεις που αποτυπώνονται στο σύστημα TwitterStand ταξινομούνται σε πέντε βασικές κατηγορίες: (i) Γενικές ειδήσεις, (ii) Επιχειρηματικότητα, (iii) Επιστήμες, (iv) Διασκέδαση, (v) Υγεία και (vi) Αθλητικά. Στον παρακάτω χάρτη απεικονίζονται οι τρέχουσες επίκαιρες ειδήσεις (19 Μαΐου 2017) στις ανατολικές ακτές των Ηνωμένων Πολιτειών:



Εικόνα 12: Επισήμανση τρεχουσών ειδήσεων στο χάρτη – TwitterStand

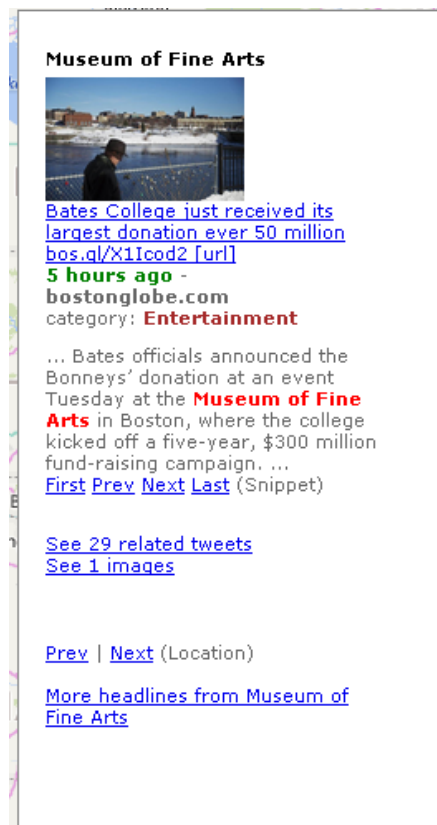
Για τη διευκόλυνση του χρήστη οι τρέχουσες ειδήσεις απεικονίζονται στο χάρτη με διαφορετικό εικονίδιο, σύμφωνα με τη κατηγορία που ανήκουν. Κάνοντας ο χρήστης κλικ στα αντίστοιχα εικονίδια, εμφανίζονται πληροφορίες για τις σχετικές ειδήσεις:



Εικόνα 13: Εμφάνιση πληροφοριών τρεχουσών ειδήσεων στο TwitterStand

Κάνοντας κλικ ο χρήστης στο σχετικό εικονίδιο (βελάκι) δίπλα στον τίτλο ειδήσεων, έχει ως αποτέλεσμα την προβολή των αντίστοιχων tweets:

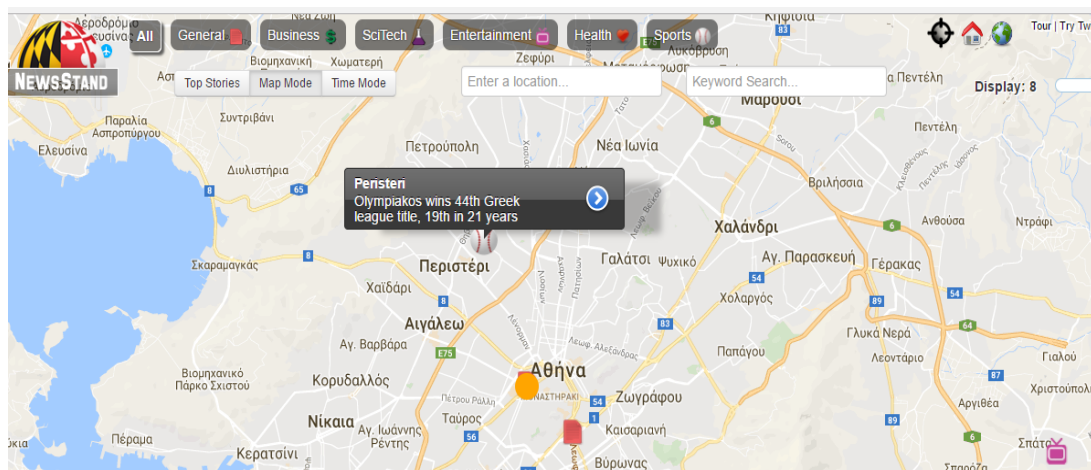
Οπτική απεικόνιση διάδοσης και περιεχομένου αναρτήσεων στο Twitter



Εικόνα 14: Παρουσίαση επιλεγμένης είδησης στο χρήστη στο σύστημα TwitterStand

Τέλος, το σύστημα TwitterStand παρέχει στο χρήστη εργαλεία για το φιλτράρισμα των κατηγοριών νέων ειδήσεων που προβάλλονται στη διεπαφή χάρτη.

Σε νεότερή τους έρευνα οι Samet et al. (2014) αναβάθμισαν το σύστημα TwitterStand, το οποίο πλέον καλείται NewsStand. Νέες λειτουργίες του συστήματος αποτελούν (i) η δυνατότητα καθορισμού του αριθμού νέων που προβάλλονται στο χάρτη Google Map, (ii) η επιλογή τοποθεσίας προβολής νέων και (iii) η αναζήτηση για νέα που περιλαμβάνουν συγκεκριμένους όρους. Στο παρακάτω σχήμα απεικονίζονται τα τρέχοντα νέα στην Αθήνα στις 19 Μαΐου 2017:



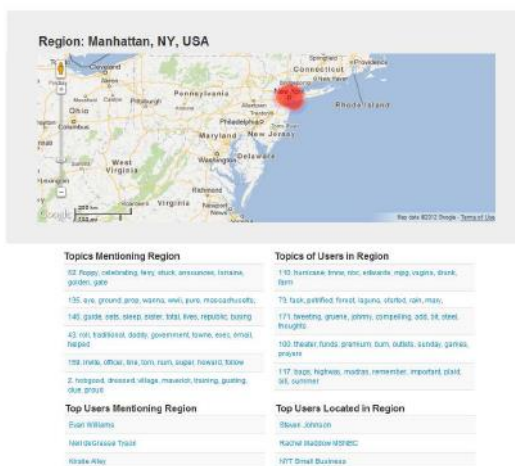
Εικόνα 15: Επισήμανση τρεχουσών ειδήσεων στο χάρτη (Αττική, 19/05/2017) – NewsStand

Οπτική απεικόνιση διάδοσης και περιεχομένου αναρτήσεων στο Twitter

3.2.2 Οπτικοποίηση γεωγραφικού ενδιαφέροντος χρηστών

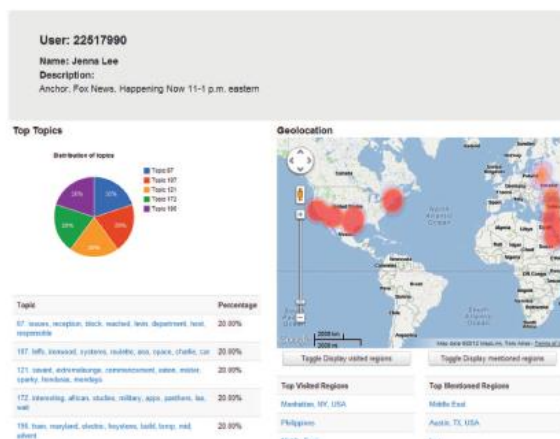
Οι Vosecky et al. (2013) ανέπτυξαν το σύστημα Limosa, ένα διαδραστικό σύστημα απεικόνισης του γεωγραφικού ενδιαφέροντος χρηστών στο Twitter. Ειδικότερα, το σύστημα Limosa περιλαμβάνει τέσσερις (4) διεπαφές:

- **Οπτικοποίηση περιοχής:** αρχικά ο χρήστης επιλέγει γεωγραφική περιοχή και στη συνέχεια αποτυπώνονται σε χάρτη τα μηνύματα που αναρτήθηκαν στη σχετική περιοχή με τη μορφή χάρτη εστίασης (heatmap) σύμφωνα με την πυκνότητα ανάρτησης μηνυμάτων Twitter για την επιλεγμένη περιοχή.



Εικόνα 16: Οπτικοποίηση περιοχής στο σύστημα Limosa (Vosecky et al., 2013)

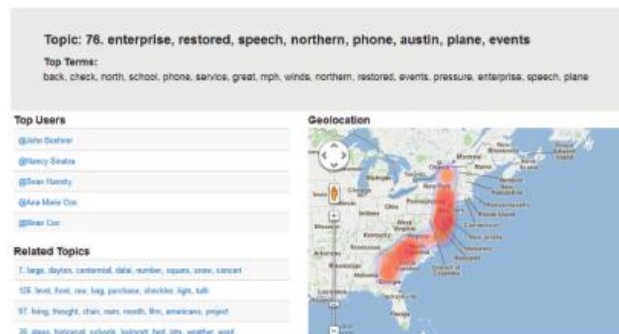
- **Οπτικοποίηση χρηστών:** αρχικά ο χρήστης επιλέγει το προφίλ ενός χρήστη στο Twitter και στη συνέχεια εμφανίζονται σε χάρτη: οι περιοχές που έχει επισκεφτεί ο χρήστης, καθώς επίσης και οι περιοχές ή θέματα που περιλαμβάνονται στις αναρτήσεις (tweets) του χρήστη.



Εικόνα 17: Οπτικοποίηση χρηστών στο σύστημα Limosa (Vosecky et al., 2013)

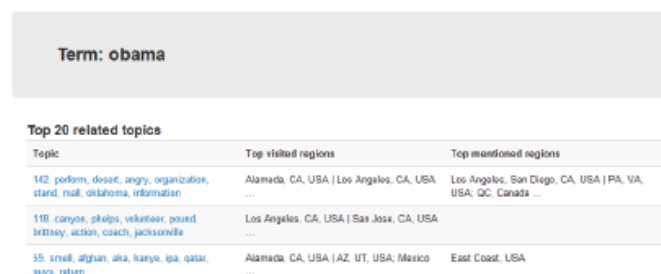
Οπτική απεικόνιση διάδοσης και περιεχομένου αναρτήσεων στο Twitter

- **Οπτικοποίηση θέματος:** αρχικά ο χρήστης επιλέγει κάποιο θέμα και στη συνέχεια εμφανίζονται σε χάρτη οι τοποθεσίες όπου το αντίστοιχο θέμα έχει αναφερθεί πιο συχνά, τους χρήστες που ανήρτησαν περιεχόμενο σχετικό με το επιλεγμένο θέμα και άλλες σχετικές πληροφορίες.



Εικόνα 18: Οπτικοποίηση θέματος στο σύστημα Limosa (Vosecky et al., 2013)

- **Οπτικοποίηση όρων αναζήτησης:** με βάση τους όρους αναζήτησης που δηλώνει ο χρήστης εμφανίζονται τα είκοσι (20) περισσότερο δημοφιλή θέματα σχετικά με τους όρους αναζήτησης που δήλωσε ο χρήστης.



Εικόνα 19: Οπτικοποίηση όρων αναζήτησης στο σύστημα Limosa (Vosecky et al., 2013)

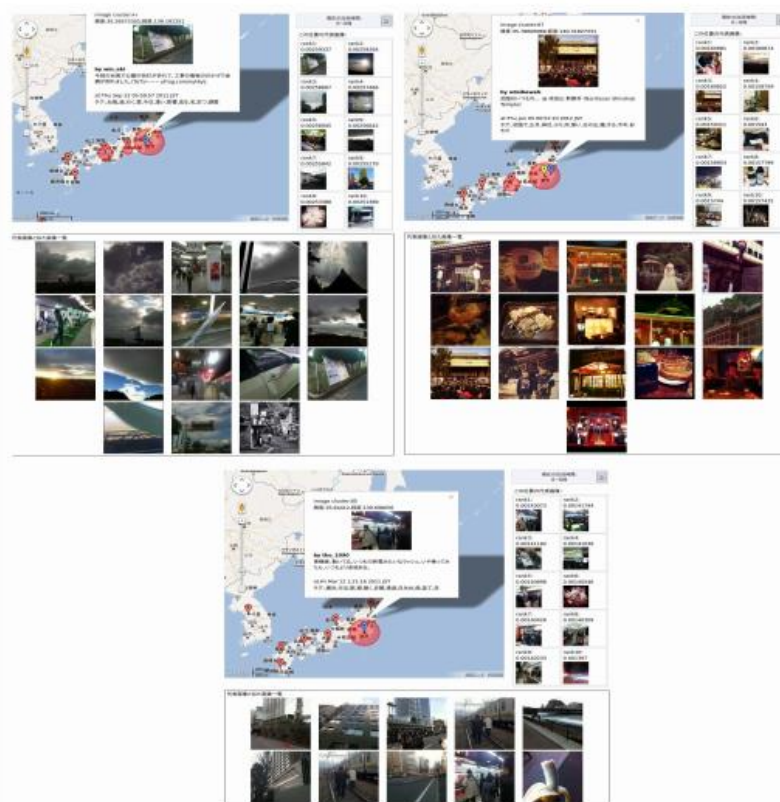
Το σύστημα Limosa επιτρέπει την οπτική απεικόνιση των ενδιαφερόντων των χρηστών και των δημοφιλών θεμάτων ανά γεωγραφική περιοχή, καθώς επίσης και της απήχησης επιλεγμένων θεμάτων σε διάφορες γεωγραφικές περιοχές.

3.2.3 Οπτικοποίηση συνοδευτικών εικόνων αναρτήσεων tweets

Σε σχετική έρευνα των Nakaji & Yanai (2012) συλλέχτηκαν tweets για περίπου 1.5 χρόνο και στη συνέχεια με βάση τα γεγονότα που κυριάρχησαν την αντίστοιχη περίοδο, αναζητήθηκαν tweets που αναφέρονταν σε σχετικά γεγονότα και περιλάμβαναν συνοδευτικά τουλάχιστον μια εικόνα. Κατόπιν, τα tweets που συνοδεύονταν από πληροφορίες σχετικά με το γεωγραφικό σημείο δημοσίευσής τους στο Twitter, απεικονίστηκαν σε χάρτη Google Map με τα σχετικά γεωγραφικά σημεία και τις συνοδευόμενες φωτογραφίες ανάρτησης. Όπως διαπιστώθηκε από την έρευνα,

υπήρχε διαφοροποίηση του θέματος και του ύφους των εικόνων που αναρτώνταν με βάση την περιοχή ανάρτησης και το χρόνο ανάρτησης στη διάρκεια της ημέρας.

Η παρακάτω εικόνα απεικονίζει την οπτικοποίηση tweets και των εικόνων που περιλαμβάνουν τα αντίστοιχα tweets για επιλεγμένα θέματα ειδήσεων (πρώτη εικόνα: Τυφώνας, δεύτερη εικόνα: Πρωτοχρονιά, Τρίτη εικόνα: Σεισμός) και αναρτήθηκαν στο Τόκιο της Ιαπωνίας:



Εικόνα 20: Οπτικοποίηση εικόνων ανάρτησης σε tweets για επιλεγμένα θέματα ειδήσεων στο Τόκιο της Ιαπωνίας (Nakaji & Yanai, 2012)

Σε ανάλογη έρευνα των Kaneko & Yanai (2016), κατόπιν της συλλογής περίπου 5.5 εκατομμυρίων tweets με χρήση τεχνικών εξόρυξης δεδομένων, προσδιορίστηκαν τα σημαντικότερα γεγονότα που περιλαμβάνονταν στα αντίστοιχα tweets, η γεωγραφική τοποθεσία των αντίστοιχων γεγονότων και στη συνέχεια επισημάνθηκαν οπτικά τα αντίστοιχα γεγονότα σε χάρτη Google Map. Στη συνέχεια επιλέγοντας ο χρήστης γεγονότα στο χάρτη Google Map, εμφανίζονται εικόνες που περιλαμβάνονταν στα αντίστοιχα γεγονότα – σχετικές αναρτήσεις tweets:



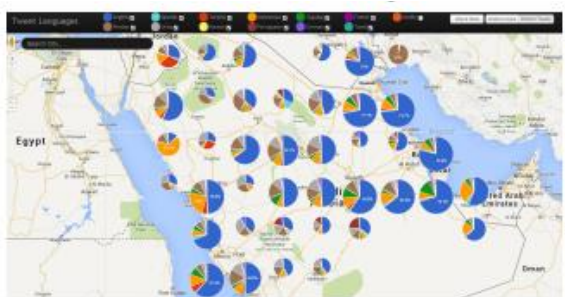
Εικόνα 21: Οπτική απεικόνιση σημαντικών γεγονότων σε χάρτη Google Map (Kaneko & Yanai, 2016)



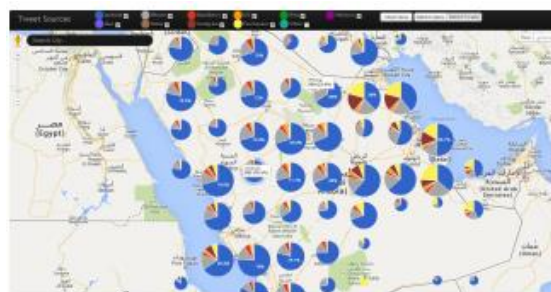
Εικόνα 22: Εμφάνιση συνοδευτικών εικόνων για επιλεγμένο γεγονός (Kaneko & Yanai, 2016)

3.2.4 Οπτικοποίηση χαρακτηριστικών αναρτήσεων tweets

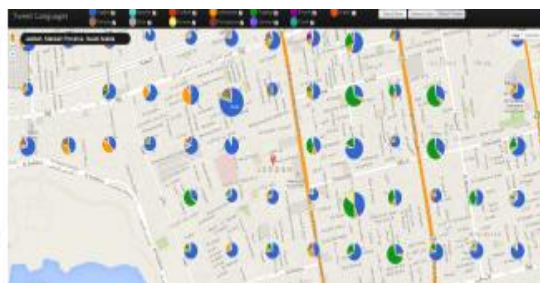
Σε σχετική έρευνα των Ghanem et al. (2014) παράλληλα με την οπτική παρουσίαση των tweets σε διεπαφή χάρτη, για κάθε περιοχή απεικονίζονται στατιστικά στοιχεία (με τη μορφή γραφήματος πίτας) της γλώσσας που είναι γραμμένα τα αναρτημένα tweets και του λειτουργικού συστήματος που χρησιμοποιήθηκε κατά την ανάρτησή τους στο Twitter:



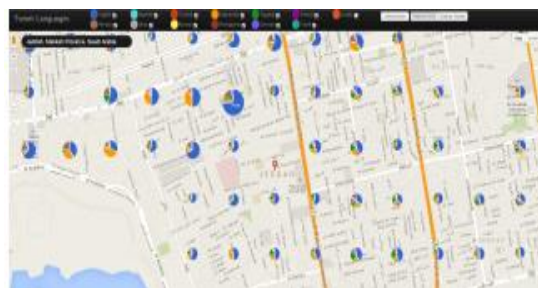
Εικόνα 23: Οπτική απεικόνιση σε χάρτη στατιστικών στοιχείων σχετικά με τη γλώσσα των αναρτημένων tweets στη Σαουδική Αραβία (Ghanem et al., 2014)



Εικόνα 24: Οπτική απεικόνιση σε χάρτη στατιστικών στοιχείων σχετικά με τη λειτουργικό σύστημα ανάρτησης tweets στη Σαουδική Αραβία (Ghanem et al., 2014)



Εικόνα 25: Οπτική απεικόνιση σε χάρτη στατιστικών στοιχείων σχετικά με τη γλώσσα των αναρτημένων tweets στην πόλη Jeddah της Σαουδικής Αραβίας με βάση τον διαφορετικό αριθμό tweets (Ghanem et al., 2014)



Εικόνα 26: Οπτική απεικόνιση σε χάρτη στατιστικών στοιχείων σχετικά με τη γλώσσα των αναρτημένων tweets στην πόλη Jeddah της Σαουδικής Αραβίας με βάση τον διαφορετικό αριθμό χρηστών (Ghanem et al., 2014)

Το σύστημα που αναπτύχθηκε καλείται VisCAT. Όπως αναφέρουν οι ερευνητές αντίστοιχες μορφές - πληροφορίες οπτικής παρουσίασης - ανάλυσης μπορεί να ενδιαφέρουν διάφορες

κατηγορίες χρηστών, όπως τις διοικητικές αρχές μιας χώρας για τον προσδιορισμό του αριθμού προσφύγων που βρίσκονται στη χώρα ή τους τουρίστες που επισκέπτονται μια χώρα. Επίσης, αντίστοιχες οπτικές παρουσιάσεις αναρτήσεων tweets μπορεί να ενδιαφέρουν οργανώσεις που παρακολουθούν τη χωρική κατανομή πλυθησμών εντός της επικράτειας ενός κράτους (Ghanem et al., 2014).

3.2.5 Οπτικοποίηση τοποθεσίας ανάρτησης και τοποθεσία αναφοράς tweets

Σε σχετική τους έρευνα οι Budak et al. (2014), ανέπτυξαν σύστημα (GeoScope) που απεικονίζει σε διεπαφή χάρτη εστίασης (heatmap) τις περιοχές όπου αναρτήθηκαν διάφορα tweets σε παγκόσμιο επίπεδο, καθώς επίσης σε ξεχωριστό χάρτη και τις περιοχές που αναφέρονται τα αντίστοιχα tweets:



Εικόνα 27: Οπτική απεικόνιση περιοχών ανάρτησης tweets – GeoScope (Budak et al., 2014)



Εικόνα 28: Οπτική απεικόνιση περιοχών αναφοράς tweets – GeoScope (Budak et al., 2014)

Όπως διαπιστώνεται από τους παραπάνω χάρτες, δεν υπάρχει πάντα αναλογία στην περιοχή ανάρτησης των διαφόρων tweets και στις περιοχές που αναφέρονται τα διάφορα tweets που αναρτώνται στο Twitter. Για παράδειγμα, στους παραπάνω χάρτες η Ιαπωνία αποτελεί αντικείμενο αναφοράς για μεγάλο αριθμό tweets σε παγκόσμιο επίπεδο και αυτό οφείλεται στους σεισμούς που πραγματοποιήθηκαν στην Ιαπωνία το διάστημα συλλογής των αντίστοιχων tweets. Επίσης, παρατηρείται στους παραπάνω χάρτες πως για χώρες όπως η Ινδονησία υπάρχει μεγάλος συνολικός αριθμός αναρτήσεων tweets, αλλά συγκριτικά μικρός αριθμός tweets που αναφέρονται στην Ινδονησία. Το γεγονός αυτό οφείλεται στο ότι η Ινδονησία είναι μια ιδιαίτερα δραστήρια χώρα στο Twitter, ενώ στη χώρα δε λαμβάνουν τόπο σημαντικά γεγονότα με παγκόσμιο ενδιαφέρον (Budak et al., 2014).

3.3 Σύνοψη - Σύγκριση

Παραπάνω αναφέρθηκαν ιστότοποι στο διαδίκτυο με αντικείμενο την οπτική απεικόνιση διάδοσης και περιεχομένου αναρτήσεων στο Twitter, καθώς επίσης και έρευνες με αντικείμενο την οπτική απεικόνιση, ειδικού ανά έρευνα, περιεχομένου αναρτήσεων στο Twitter. Οι ιστότοποι που παρουσιάστηκαν έχουν άμεση σχέση με τους σκοπούς της παρούσας μελέτης, ενώ οι έρευνες που παρουσιάστηκαν, μολονότι έχουν το ίδιο αντικείμενο έρευνας με τη παρούσα μελέτη, εστιάζουν σε συγκεκριμένους τομείς οπτικής απεικόνισης περιεχομένου.

Στον παρακάτω πίνακα συνοψίζονται οι δυνατότητες - λειτουργίες που διαθέτουν οι εφαρμογές – ιστότοποι που περιγράφηκαν παραπάνω σε σχέση με το αντικείμενο και τους στόχους της παρούσας μελέτης:

Πίνακας 1: Δυνατότητες ιστοτόπων οπτικής απεικόνισης διάδοσης και περιεχομένου αναρτήσεων στο Twitter

	TrendsMap	OneMillionTweetMap	TweetMap	KeyHole
Διαθεσιμότητα φίλτρων επιλογής των tweets προβολής		✓	✓	✓
Οπτική απεικόνιση συστοιχιών σήμανσης		✓		
Οπτική απεικόνιση χαρτών εστίασης		✓	✓	✓
Οπτική απεικόνιση δημοφιλών λέξεων	✓	✓		
Οπτική απεικόνιση δημοφιλών hashtags	✓	✓	✓	✓
Οπτική απεικόνιση διάδοσης hashtags	✓	✓	✓	
Οπτική απεικόνιση ροής εξέλιξης δημοφιλών λέξεων				
Οπτική απεικόνιση ροής εξέλιξης δημοφιλών hashtags				

Όπως διαπιστώνεται από τον παραπάνω πίνακα, οι ιστότοποι που παρουσιάστηκαν κατά κύριο λόγο απεικονίζουν οπτικά το περιεχόμενο αναρτήσεων στο Twitter με τη μορφή χάρτη εστίασης. Επίσης, κανένας από τους ιστότοπους δεν περιλαμβάνει ειδική λειτουργία προβολής των δημοφιλών λέξεων ή hashtags για διάφορες περιοχές με αυτοματοποιημένη μορφή προβολής των αντίστοιχων πληροφοριών ανά ημέρα προβολής.

Για τις έρευνες που αναφέρθηκαν παραπάνω επισημαίνονται στον παρακάτω πίνακα τα πλεονεκτήματα και μειονεκτήματα που παρουσιάζουν σχετικά με την οπτική απεικόνιση περιεχομένου αναρτήσεων στο Twitter:

Πίνακας 2: Πλεονεκτήματα και μειονεκτήματα εφαρμογών οπτικής απεικόνισης περιεχομένου αναρτήσεων στο Twitter που αναπτύχθηκαν στο πλαίσιο σχετικών ερευνών

	Πλεονεκτήματα	Μειονεκτήματα
Sankaranarayanan et al. (2009)	<ul style="list-style-type: none"> • Συσχέτιση αναρτήσεων στο Twitter με πραγματικά γεγονότα – ειδήσεις • Δυνατότητα εφαρμογής φίλτρων αναζήτησης • Εύκολη στη χρήση διεπαφή εφαρμογής • Εύκολη στην κατανόηση απεικόνιση περιεχομένου 	<ul style="list-style-type: none"> • Έλλειψη επισήμανσης νέων ειδήσεων • Έλλειψη απεικόνισης σημαντικότητας ή δημοφιλών ειδήσεων • Έλλειψη απεικόνισης ροής διάδοσης σχετικών ειδήσεων
Vosecky et al. (2013)	<ul style="list-style-type: none"> • Διασύνδεση απεικόνισης αναρτήσεων στο Twitter και χρηστών που πραγματοποιούν τις αντίστοιχες αναρτήσεις • Δυνατότητα εφαρμογής φίλτρων αναζήτησης 	<ul style="list-style-type: none"> • Δύσκολη στην κατανόηση απεικόνιση περιεχομένου • Δύσκολη στη χρήση διεπαφή εφαρμογής • Έλλειψη απεικόνισης ροής διάδοσης σχετικών πληροφοριών
Nakaji & Yanai (2012)	<ul style="list-style-type: none"> • Διασύνδεση αναρτήσεων στο Twitter με συνοδευτικές εικόνες • Εύκολη στη χρήση διεπαφή εφαρμογής • Εύκολη στην κατανόηση απεικόνιση περιεχομένου 	<ul style="list-style-type: none"> • Έλλειψη απεικόνισης ροής διάδοσης σχετικών πληροφοριών
Ghanem et al. (2014)	<ul style="list-style-type: none"> • Οπτική χωροταξική απεικόνιση στατιστικών στοιχείων αναρτήσεων στο Twitter 	<ul style="list-style-type: none"> • Δύσκολη στην κατανόηση απεικόνιση περιεχομένου

	<ul style="list-style-type: none"> • Δυνατότητα εφαρμογής φίλτρων αναζήτησης • Εύκολη στη χρήση διεπαφή εφαρμογής 	<ul style="list-style-type: none"> • Έλλειψη απεικόνισης ροής διάδοσης σχετικών πληροφοριών
Budak et al. (2014)	<ul style="list-style-type: none"> • Απεικόνιση σε χάρτη εστίασης της κατανομής αναρτήσεων στο Twitter, παράλληλα με απεικόνιση σε ξεχωριστό χάρτη της κατανομής των περιοχών που αναφέρονται τα αντίστοιχα tweets • Εύκολη στην κατανόηση απεικόνιση περιεχομένου 	<ul style="list-style-type: none"> • Έλλειψη απεικόνισης ροής διάδοσης σχετικών αναρτήσεων • Έλλειψη δυνατότητας εφαρμογής φίλτρων

ΚΕΦ.4: Μεθοδολογία έρευνας – Σχεδίαση έρευνας – Εργαλεία που χρησιμοποιήθηκαν

Το πρώτο βήμα για την οπτική απεικόνιση των μηνυμάτων που αναρτούν οι χρήστες στο κοινωνικό δίκτυο Twitter είναι η συλλογή αντίστοιχων δεδομένων - αναρτημένων μηνυμάτων στο Twitter. Για το σκοπό αυτό χρησιμοποιήθηκε η σχετική διεπαφή προγραμματισμού (API) του Twitter. Κατόπιν τα σχετικά δεδομένα αποθηκεύθηκαν σε βάση δεδομένων με σκοπό την εφαρμογή διαφόρων ομαδοποιήσεων – ταξινομήσεων των δεδομένων πριν την οπτική απεικόνισή τους. Τέλος, δημιουργήθηκε ιστότοπος για την υλοποίηση της οπτικής απεικόνισης των αρχικών tweets που συλλέχθηκαν και των αποτελεσμάτων της επεξεργασίας των δεδομένων στη βάση δεδομένων.

4.1 Δείγμα έρευνας

Ως δείγμα της έρευνας χρησιμοποιήθηκαν tweets που δημοσιεύτηκαν στο Twitter σε διάρκεια ενός μηνός και πιο συγκεκριμένα την περίοδο από 20 Δεκέμβρη 2016 μέχρι και 25 Ιανουαρίου 2017. Ειδικότερα, τα tweets που αναλύθηκαν δημοσιεύτηκαν από περιοχές στις πόλεις του Λονδίνου, της Νέας Υόρκης, της Ουάσινγκτον, του Σαν Φρανσίσκο και του Λος Άντζελες και αναφέρονται σε χρήστες που είχαν ενεργοποιημένη στη συσκευή τους, κατά το χρόνο χρήσης του Twitter, τη μετάδοση πληροφοριών σχετικά με τη τοποθεσία του χρήστη. Συνεπώς, η ανάλυση που εφαρμόστηκε, αφορά σε χρήστες που βρίσκονταν στις πόλεις που αναφέρθηκαν παραπάνω ή τα περίχωρά τους κατά την ανάρτηση των σχετικών tweets.

Ο αριθμός των μηνυμάτων tweets που αναλύθηκαν για την εξαγωγή των συμπερασμάτων της παρούσας μελέτης είναι ιδιαίτερα μεγάλος περίπου 1,2 εκατομμύρια tweets και αναφέρονται σε αναρτήσεις 186 χιλιάδων χρηστών του Twitter. Το δείγμα της έρευνας (σχετικές αναρτήσεις tweets) συλλέχθηκε σε πραγματικό χρόνο από τους χρήστες, καθώς η πρόσβαση σε tweets που δημοσιεύτηκαν σε παρελθόντα χρόνο θα απαιτούσε την καταβολή χρημάτων εκ μέρους του ερευνητή.

4.2 Εργαλεία συλλογής δεδομένων

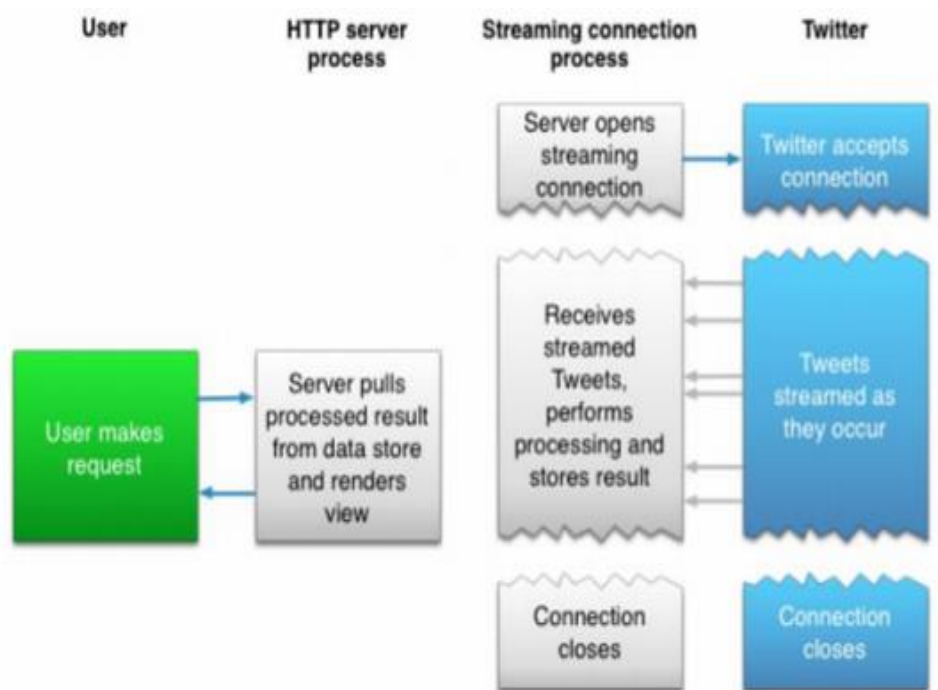
Για την συλλογή των δεδομένων χρησιμοποιήθηκαν: (i) το Twitter Streaming API, (ii) το REST API του Twitter, (iii) η γλώσσα Python και (iv) η βάση δεδομένων PostgreSQL. Στη συνέχεια ακολουθεί μια σύντομη περιγραφή - ανάλυση των δυνατοτήτων των αντίστοιχων εργαλείων.

4.2.1 Twitter Streaming API

Το Twitter Streaming API παρέχει στους προγραμματιστές πρόσβαση στη παγκόσμια ροή δεδομένων μέσω tweets στην πλατφόρμα Twitter. Επίσης, το Twitter Streaming API παρέχει πρόσβαση και σε άλλα συμβάντα που αφορούν σε αναρτήσεις χρηστών στο διαδίκτυο (tweets) σε πραγματικό χρόνο όπως διαγραφές, retweets και απαντήσεις. Επιπροσθέτως, παρέχονται φίλτρα που επιτρέπουν τη συλλογή δεδομένων που ικανοποιούν συγκεκριμένα κριτήρια, όπως tweets που αποστάλθηκαν από συγκεκριμένη γεωγραφική περιοχή ή tweets που περιέχουν κάποια συγκεκριμένη λέξη ή tweets που περιλαμβάνουν συγκεκριμένα hashtags (Riquelme & González-Cantergiani, 2016).

Με τη χρήση του Streaming API παρέχεται πρόσβαση στις αναρτήσεις των tweets σε πραγματικό χρόνο, χωρίς ωστόσο να παρέχεται πρόσβαση σε πληροφορίες σχετικά με tweets που έχουν δημοσιευτεί σε παρελθόντα χρόνο. Με τη χρήση του Streaming API εγκαθίσταται μια μόνιμη σύνδεση με το Twitter από την οποία περνούν – εμφανίζονται τα tweets. Ένας περιορισμός της χρήσης του Streaming API του Twitter αποτελεί το γεγονός πως ο αριθμός των tweets που εμφανίζονται μέσω της μόνιμης σύνδεσης δε ξεπερνούν το 1% του συνόλου δημοσιευμένων tweets. Ωστόσο, με τη χρήση φίλτρων – λέξεων κλειδιών ο χρήστης έχει τη δυνατότητα να καθορίσει συγκεκριμένα κριτήρια για τα tweets που θα περιλαμβάνει η μόνιμη σύνδεση με το Twitter. Τέλος, ένας χρήστης μπορεί να έχει ανοικτή μονάχα μια το πολύ σύνδεση με το Twitter με χρήση του Streaming API.

Η χρήση του Streaming API απαιτεί τη διατήρηση ενεργής σύνδεσης με τη διεπαφή διασύνδεσης. Η διατήρηση ενεργής σύνδεσης με το Twitter API πραγματοποιείται σε ξεχωριστό νήμα (thread), σε σχέση με την εφαρμογή που προσπελαύνει τα δεδομένα από το Twitter API. Η διαδικασία ροής (ενεργή σύνδεση) λαμβάνει τα tweets εισόδου και εκτελεί κάθε επιθυμητή ανάλυση, φιλτράρισμα ή οποιαδήποτε άλλη πράξη απαιτείται και στη συνέχεια αποθηκεύει τα τελικά δεδομένα σε μια αποθήκη δεδομένων. Κατόπιν, η διαδικασία χειρισμού http υποβάλλει ερωτήματα στη σχετική αποθήκη δεδομένων και επιστρέφει – εμφανίζει τα σχετικά δεδομένα στον χρήστη. Η σχετική διαδικασία παριστάνεται στο παρακάτω σχήμα (Twitter, 2012a):



Σχήμα 3: Σύνδεση στο Twitter API μέσω Streaming API (Twitter, 2012a)

Οι αποκρίσεις με χρήση του Twitter Streaming API σχετικά με τα διάφορα tweets περιλαμβάνουν μεταξύ άλλων τις εξής πληροφορίες:

Πίνακας 3: Πεδία δεδομένων tweets χρησιμοποιώντας το Twitter API (Twitter, 2012b)

Χαρακτηριστικό	Περιγραφή
contributors	Ο δημιουργούς ενός tweet
coordinates	Γεωγραφικές διαστάσεις της θέσης ανάρτησης του tweet
created_at	Το χρόνο UCT δημοσίευσης του tweet
entities	Πιθανά hashtags, url ή αναφορές χρηστών σε ένα tweet
favorite_count	Πόσες φορές έχει γίνει liked το σχετικό tweet
favorited	Αν έχει γίνει liked από τον δημιουργό του
id	Μοναδικό αναγνωριστικό κάθε tweet
id_str	Απόδοση με μορφή αλφαριθμητικού του μοναδικού αναγνωριστικού του tweet
lang	Η γλώσσα κειμένου που διαπιστώθηκε για το εξεταζόμενο tweet
place	Αν το tweet είναι συνδεδεμένο με μια τοποθεσία, οι σχετικές γεωγραφικές συντεταγμένες
retweet_count	Πόσες φορές έχει γίνει retweet κάποιο tweet
retweeted	Αν έχει γίνει ένα tweet retweeted από τον δημιουργό του

Οπτική απεικόνιση διάδοσης και περιεχομένου αναρτήσεων στο Twitter

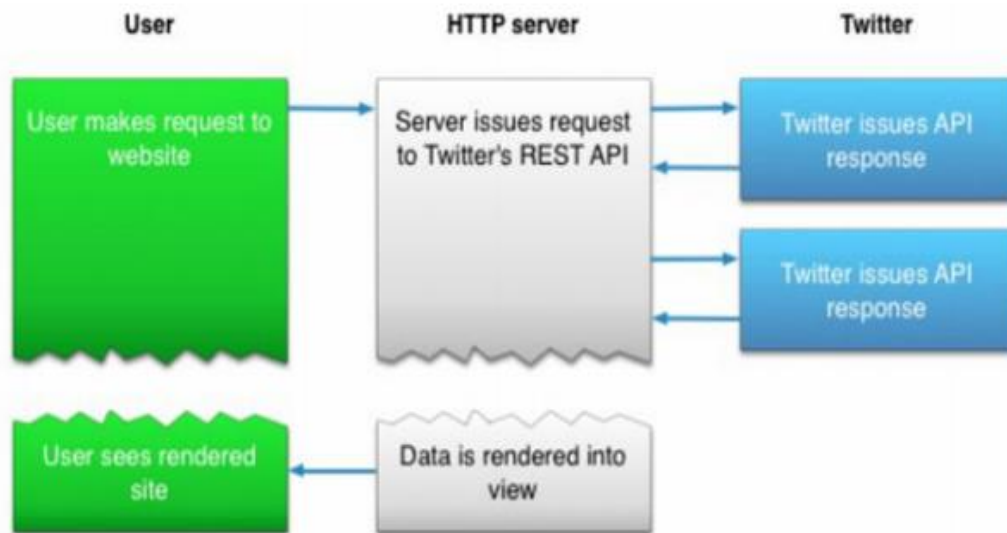
Χαρακτηριστικό	Περιγραφή
source	Η διεύθυνση στο διαδίκτυο (URL) για τη προσπέλαση του σχετικού tweet
user	Ο χρήστης που δημοσίευσε το tweet

4.2.2 REST API

Το REST (Representational State Transfer) API παρέχει πρόσβαση μέσω προγραμματισμού στην ανάγνωση και εγγραφή δεδομένων στο Twitter. Μεταξύ άλλων το REST API επιτρέπει τη δημιουργία ενός καινούργιου tweet, την προσπέλαση των στοιχείων του προφίλ ενός χρήστη, την προσπέλαση των στοιχείων των ατόμων που κάνουν follow κάποιον χρήστη και πολλά άλλα. Το REST API βασίζεται στην αρχιτεκτονική REST, η οποία είναι αρκετά δημοφιλής στο διαδίκτυο. Για τη συλλογή δεδομένων με το REST API είναι απαραίτητο να το ζητήσει ο χρήστης, δηλαδή εφαρμόζεται η μέθοδος PULL κατά τη συλλογή δεδομένων. Αντίθετα στο Streaming API εφαρμόζεται η μέθοδος PUSH, παρέχοντας μια συνεχή ροή δεδομένων από τη στιγμή της αποστολής και ανταπόκρισης στο αίτημα ενός χρήστη, χωρίς την ανάγκη τακτικών αιτήσεων λήψης δεδομένων.

Υπάρχει όριο των αποτελεσμάτων που επιστρέφονται με χρήση του REST API στα 100 tweets για κάθε ερώτημα χρήστη. Επίσης, για την αποφυγή κατάχρησης χρήσης του REST API, είναι επιτρεπτά το πολύ 180 ερωτήματα ανά 15 λεπτά. Ωστόσο, για περιπτώσεις ερωτημάτων που επιστρέφουν μεγάλο όγκο δεδομένων ή απαιτούν μεγάλη επεξεργασία εκ μέρους του Twitter, το όριο μειώνεται σε 15 ερωτήματα ανά 15 λεπτά.

Για τη χρήση του REST API δεν απαιτείται η εγκαθίδρυση μόνιμης σύνδεσης με το Twitter όσο διαρκεί η προσπέλαση δεδομένων. Μέσω του REST API υποβάλλονται αιτήματα στο API του Twitter, στη συνέχεια αποστέλλονται στο χρήστη οι πληροφορίες που ζητήθηκαν, ακολουθεί η μορφοποίηση των δεδομένων και τελικά προβάλλονται στο χρήστη οι μορφοποιημένες πληροφορίες ως απάντηση στην αρχική του αίτηση. Η σχετική διαδικασία παριστάνεται στο παρακάτω σχήμα (Twitter, 2012a):



Σχήμα 4: Σύνδεση στο Twitter API μέσω REST API (Twitter, 2012a)

4.2.3 Python

Η Python είναι μια αντικειμενοστραφής γλώσσα προγραμματισμού υψηλού επιπέδου, ιδιαίτερα ελκυστική για ταχεία ανάπτυξη εφαρμογών (Rapid Application Development) ή για τη διασύνδεση εφαρμογών. Η Python υποστηρίζει την ενσωμάτωση ενοτήτων (modules) και πακέτων (packages), γεγονός που επιτρέπει την εύκολη επαναχρησιμοποίηση προγραμματιστικού κώδικα. Ο διερμηνέας (interpreter) Python και οι βιβλιοθήκες της Python είναι διαθέσιμες σε δυαδική μορφή δωρεάν για όλες τις μεγάλες υπολογιστικές πλατφόρμες και μπορεί να διανεμηθούν ελεύθερα (Python, 2015).

Η Python επιλέχτηκε καθόσον είναι μια απλή στη σύνταξη γλώσσα προγραμματισμού και επιτρέπει τη συγγραφή τμήματος κώδικα σε πολύ λιγότερες γραμμές σε σχέση με άλλες γλώσσες προγραμματισμού (πχ Java).

4.2.4 PostgreSQL

Η PostgreSQL αποτελεί μια ανοικτού κώδικα σχεσιακή βάση δεδομένων. Η σχετική βάση δεδομένων μπορεί να εγκατασταθεί σε όλα τα βασικά λειτουργικά συστήματα υπολογιστών (Linux, UNIX, Windows) και υποστηρίζει τον ορισμό πρωτευόντων (primary keys) και δευτερευόντων κλειδιών (foreign keys), όψεις (views), συνενώσεις πινάκων (joins) και εναύσματα (triggers) (PostgreSQL, 2014).

Επίσης, η PostgreSQL επιτρέπει την εκτέλεση αποθηκευμένων διαδικασιών (stored procedures) με πολλές γλώσσες προγραμματισμού, μεταξύ των οποίων τη Java, Python, Perl, C/C++, καθώς

επίσης και τη δικιά της γλώσσα προγραμματισμού PL/pgSQL που λειτουργεί ανάλογα με την PL/SQ της Oracle (PostgreSQL, 2014).

Επιπροσθέτως, υποστηρίζει τη δυνατότητα αποθήκευσης όλων των βασικών τύπων δεδομένων σε μια βάση δεδομένων, όπως αριθμητικούς χαρακτήρες (numeric), πεδίο ψευδές/αληθές (Boolean), αλφαριθμητικά (varchar), ημερομηνίες (datetime), καθώς επίσης και δεδομένα σε δυαδική μορφή (binary data) (PostgreSQL, 2014).

Τέλος, η PostgreSQL υποστηρίζει δυνατότητες όπως αποκατάσταση βάσης δεδομένων σε συγκεκριμένη ημερομηνία, ασύγχρονη αντιγραφή δεδομένων, ένθετες συναλλαγές (savepoints), λήψη αντιγράφων μέσω διαδικτύου και περιβάλλον ανάλυσης – βελτιστοποίησης εκτέλεσης ερωτημάτων SQL (PostgreSQL, 2014).

4.3 Εργαλεία επεξεργασίας δεδομένων

Παρότι τα δεδομένα που συλλέχτηκαν αρχικά αποθηκεύτηκαν σε βάση δεδομένων PostgreSQL, στην πορεία διαπιστώθηκε πως είναι ευκολότερη η επεξεργασία των αντίστοιχων δεδομένων με χρήση του συστήματος MS SQL Server, οπότε αποφασίστηκε η μεταφορά των δεδομένων σε βάση δεδομένων SQL Server και η επεξεργασία τους στο αντίστοιχο σύστημα διαχείρισης βάσεων δεδομένων.

4.3.1 MS SQL Server

Ο κυριότερος λόγος που αποφασίστηκε η μεταφορά των δεδομένων από το σύστημα PostgreSQL στο σύστημα MS SQL Server είναι η υποστήριξη που παρέχει η έκδοση MS SQL Server 2016 για την επεξεργασία δεδομένων σε μορφή json. Τα δεδομένα με μορφή JSON αποτελούν σήμερα μια από τις περισσότερο συχνά χρησιμοποιούμενες μορφές αρχείων με σκοπό την ανταλλαγή δεδομένων. Οι περισσότερες σύγχρονες υπηρεσίες επιστρέφουν πληροφορίες μορφοποιημένες ως κείμενο JSON. Τα προγράμματα περιήγησης ιστού χρησιμοποιούν πληροφορίες μορφοποιημένες ως JSON. Το JSON αποτελεί μια μορφή δεδομένων που χρησιμοποιείται τακτικά όταν πρέπει να αποθηκευτούν ημιδομημένες πληροφορίες σε αρχεία ή βάσεις δεδομένων (Porovic, 2016).

Βασικό μειονέκτημα των αρχείων JSON είναι η δυσκολία εκτέλεσης ερωτημάτων σε σχετικά δεδομένα που βρίσκονται αποθηκευμένα σε μια βάση δεδομένων. Ωστόσο, η τελευταία έκδοση του MS SQL Server υποστηρίζει την αποτελεσματική διαχείριση αντίστοιχης μορφής δεδομένων (Porovic, 2016).

Πλέον, ο MS SQL Server υποστηρίζει την εύκολη εξαγωγή δεδομένων από ένα αρχείο JSON ή την επαλήθευση πως ένα αρχείο JSON είναι σωστά διαμορφωμένο χρησιμοποιώντας τις ενσωματωμένες εντολές JSON_VALUE, JSON_QUERY και ISJSON. Για πιο περίπλοκα ερωτήματα και αναλύσεις, η συνάρτηση OPENJSON μπορεί να μετατρέψει μια σειρά από αντικείμενα JSON σε ένα σύνολο γραμμών, που ταιριάζει καλύτερα στη φιλοσοφία αποθήκευσης δεδομένων σε μια σχεσιακή βάση δεδομένων. Οποιοδήποτε ερώτημα SQL μπορεί να εκτελεστεί σε δεδομένα της μορφής JSON με χρήση της εντολής OPENJSON. Τέλος, υπάρχει η εντολή FOR JSON που επιτρέπει τη διαμόρφωση των αποτελεσμάτων εκτέλεσης ενός ερωτήματος και παρουσίαση τους με την μορφή κειμένου JSON (Porovic, 2016).

Με σκοπό την αποτελεσματικότερη επεξεργασία των tweets που συλλέχθηκαν και την εξοικονόμηση χρόνου για την εκτέλεση ερωτημάτων SQL στα δεδομένα που συλλέχθηκαν και είναι της μορφής κειμένου JSON, αποφασίστηκε η μεταφορά των δεδομένων σε βάση δεδομένων MS SQL Server και κατόπιν η επεξεργασία τους στο ίδιο σχεσιακό σύστημα διαχείρισης βάσεων δεδομένων.

Ο SQL Server είναι ένα ολοκληρωμένο σχεσιακό σύστημα διαχείρισης βάσεων δεδομένων που έχει δημιουργηθεί από τη Microsoft. Το γεγονός αυτό υποδηλώνει πως ο SQL Server χρησιμοποιείται για την αποθήκευση των δεδομένων σε μια σχεσιακή βάση δεδομένων, αλλά και πως ο SQL Server αποτελεί ένα σύστημα διαχείρισης των δεδομένων και όχι απλά μια βάση δεδομένων. Ενδεικτικά παρέχει εργαλεία: (i) βελτιστοποίησης εκτέλεσης ερωτημάτων SQL (Query Optimization), (ii) διαχείρισης αντιγράφων ασφαλείας, (iii) εισαγωγής και εξαγωγής δεδομένων σε άλλα συστήματα διαχείρισης βάσεων δεδομένων ή άλλες μορφές αρχείων (xml, csv, xls, txt, json), (iv) υποστήριξης συναλλαγών (transactions), (v) δημιουργίας αναφορών με βάση τα δεδομένα (reporting services), (vi) δυνατότητα αναζήτησης πλήρους κειμένου (full text search) και (vii) εργαλεία εξόρυξης δεδομένων (business analytics).

Όπως αναφέρθηκε παραπάνω ο όγκος των δεδομένων που θα αποθηκευτούν στη βάση δεδομένων και θα εφαρμοστούν ερωτήματα SQL με σκοπό την οπτική απεικόνιση των αποτελεσμάτων είναι πολύ μεγάλος, πάνω από ένα εκατομμύριο εγγραφές. Συνεπώς, η ταχύτητα εκτέλεσης ερωτημάτων SQL στη βάση δεδομένων σχετίζεται άμεσα με την αποτελεσματικότητα της οπτικής απεικόνισης των αποτελεσμάτων στα ερωτήματα SQL που θα εκτελεστούν στο πλαίσιο της παρούσας έρευνας. Σε ανάλογη έρευνα που παρουσιάζεται στον

ιστότοπο Periscope Data¹, η απόδοση των ίδιων ερωτημάτων SQL σε βάση δεδομένων SQL Server υπερτερεί της αντίστοιχης απόδοσης – ταχύτητας για τα σχεσιακά συστήματα βάσεων δεδομένων MySQL, Postgresql και Oracle. Πιο συγκεκριμένα εκτελέστηκαν τρεις κατηγορίες ερωτημάτων σε βάση δεδομένων με εκατομμύρια εγγραφές (Periscope Data, 2013):

- **Ερώτημα 1:** συνένωση πινάκων, ομαδοποίηση και υπολογισμός αριθμού εγγραφών
- **Ερώτημα 2:** χρήση υπό ερωτήματος ομαδοποίησης και καταμέτρησης αριθμού εγγραφών και στη συνέχεια εφαρμογή συνένωσης πινάκων
- **Ερώτημα 3:** υπολογισμός αριθμού εγγραφών πριν την ομαδοποίηση εγγραφών και συνένωση πινάκων

Τα αποτελέσματα που προέκυψαν παρουσιάζονται στον παρακάτω πίνακα (Periscope Data, 2013):

Πίνακας 4: Σύγκριση επίδοσης MySQL, Postgre, SQL Server και Oracle

	MySQL	Postgres	SQL Server	Oracle
Ερώτημα 1	27.5s	348s	4.47s	6.09s
Ερώτημα 2	25.9s	10.6s	2.13s	3.70s
Ερώτημα 3	13.3s	7.13s	2.21s	3.71s

Από τα αντίστοιχα αποτελέσματα διαφαίνεται πως ορθώς επιλέχτηκε η χρήση της βάσης δεδομένων MS SQL Server, καθώς για την αποτελεσματική οπτική απεικόνιση των tweets και την εναλλαγή παρουσίασης των αποτελεσμάτων προβολής σε χρονολογικά σειρά απαιτείται η ταχύτερη δυνατή εκτέλεση ερωτημάτων SQL στα δεδομένα των tweets.

4.4 Εργαλεία υλοποίησης ιστότοπου

Για την υλοποίηση ενός σχετικού ιστότοπου θα χρησιμοποιηθούν τα ακόλουθα: (i) Google Maps API, (ii) ιστοσελίδες aspx (.NET), (iii) γλώσσα προγραμματισμού C# (server-side) και (iv) γλώσσα υλοποίησης - εκτέλεσης σεναρίων (scripting) jQuery στη πλευρά του πελάτη (client-side). Στη συνέχεια ακολουθεί μια σύντομη περιγραφή - ανάλυση των δυνατοτήτων των αντίστοιχων εργαλείων.

¹ <https://www.periscopedata.com/blog/count-distinct-in-mysql-postgres-sql-server-and-oracle.html>

Οπτική απεικόνιση διάδοσης και περιεχομένου αναρτήσεων στο Twitter

4.4.1 Google Maps API

Οι υπηρεσίες χαρτογράφησης που ξεκίνησε να παρέχει η Google από το 2005, έχουν αλλάξει σημαντικά την εμπειρία χαρτογράφησης σε ιστότοπους και εφαρμογές υπολογιστών. Όχι μόνο είναι δυνατόν να προβληθούν γεωγραφικοί χάρτες με γρήγορο και απλό τρόπο, αλλά παρέχεται η δυνατότητα ενσωμάτωσης αντικειμένων στους χάρτες με τη χρήση των Google Maps API (Schmidt & Weiser, 2012).

Ένα API (Application Programming Interface) είναι ένα σύνολο μεθόδων και εργαλείων που χρησιμοποιούνται για την υποστήριξη υλοποίησης εφαρμογών λογισμικού. Τα Google Maps API επιτρέπουν την εμφάνιση γεωγραφικών χαρτών σε έναν ιστότοπο (website) ή μια εφαρμογή (desktop application). Στην παρακάτω εικόνα απεικονίζεται ένας χάρτης, ο οποίος περιλαμβάνεται σε έναν ιστότοπο με χρήση των Google Maps API:



Εικόνα 29: Απεικόνιση γεωγραφική χάρτη με χρήση των Google Maps API

Βασικά πλεονεκτήματα της χρήσης των Google Maps API είναι (MobDevApp, 2015):

- **Μικρότερη κωδικοποίηση:** με τη χρήση των βιβλιοθηκών των Google Maps API, οι προγραμματιστές μπορούν με λίγες γραμμές κώδικα να εμφανίσουν χάρτες στις εφαρμογές ή ιστότοπους που αναπτύσσουν, χωρίς να απαιτούνται ιδιαίτερες δεξιότητες – ικανότητες για τη χρήση των βιβλιοθηκών που έχει αναπτύξει η Google.
- **Προσαρμοσμένος προγραμματισμός:** η ανάγκη συγγραφής μικρότερου σε όγκο προγραμματιστικού κώδικα, έχει ως αποτέλεσμα την απαλλαγή από επιπρόσθετες απαιτήσεις κωδικοποίησης - τεκμηρίωσης κώδικα. Αυτό διασφαλίζει μεταξύ άλλων τη βελτιωμένη διαχείριση ενός έργου πληροφορικής, γεγονός ιδιαίτερα σημαντικό όταν πρόκειται για έργο πληροφορικής μεγάλης κλίμακας στο οποίο πολλοί προγραμματιστές εργάζονται για την υλοποίησή του.

Οπτική απεικόνιση διάδοσης και περιεχομένου αναρτήσεων στο Twitter

- **Ταχύτερη ανάπτυξη:** λόγω της αυστηρής κωδικοποίησης των εντολών χρήσης των αντίστοιχων βιβλιοθηκών, ο συνολικός χρόνος ανάπτυξης εφαρμογών με χρήση των Google Maps API είναι περιορισμένος σε σύγκριση με άλλα εναλλακτικά αντίστοιχα εργαλεία - βιβλιοθήκες.
- **Χαμηλότερο κόστος:** οι μικρές απαιτήσεις σε προγραμματιστική προσπάθεια έχει ως αποτέλεσμα τον περιορισμό του κόστους ανάπτυξης εφαρμογών με χρήση των Google Maps API, συμβάλλοντας πρακτικά στον περιορισμό του χρόνου ανάπτυξης και στη βελτίωση των περιθωρίων κέρδους ενός οργανισμού.

Για τη χρήση των βιβλιοθηκών Google Maps API απαιτείται η χρήση ειδικού κωδικού, τον οποίο μπορεί να λάβει ένας χρήστης δωρεάν χρησιμοποιώντας ένα λογαριασμό χρήστη στο Google. Ωστόσο, αποτελεί μειονέκτημα η δυνατότητα περιορισμένων αιτημάτων προς τις βιβλιοθήκες του Google Maps API εκ μέρους των ιστοσελίδων ή εφαρμογών (1.000 αιτήματα/ημέρα), για την δωρεάν έκδοση σχετικών κωδικών χρήσης.

Με τη χρήση των Google Maps API είναι δυνατή (i) η επισήμανση γεωγραφικών σημείων στο χάρτη, (ii) η οπτική απεικόνιση χάρτη εστίασης (heatmap) και (iii) η ομαδοποιημένη οπτική απεικόνιση (συστοιχίες σήμανσης) σημείων στο χάρτη:



Εικόνα 30: Απεικόνιση γεωγραφικού σημείου



Εικόνα 31: Απεικόνιση χάρτη εστίασης



Εικόνα 32: Ομαδοποιημένη απεικόνιση γεωγραφικών σημείων (συστοιχίες σήμανσης)

Τρεις τύποι χωρικών δεδομένων μπορούν να ενσωματωθούν στους χάρτες Google: (i) σημεία, (ii) γραμμές και (iii) πολύγωνα, που ονομάζονται markers, polylines και polygons στην ορολογία του Google Map. Επίσης, είναι δυνατή η ενσωμάτωση απεικόνισης εικόνων σε χάρτες Google Map. Για τον προσδιορισμό των σημείων στο χάρτη ενσωμάτωσης σχετικών δεδομένων χρησιμοποιούνται οι γεωγραφικές συντεταγμένες (γεωγραφικό μήκος και γεωγραφικό πλάτος) των αντίστοιχων σημείων (Bildirici & Ulugtekin, 2010).

Η διεπαφή Google Maps API επιλέχτηκε ως το μέσο οπτικής απεικόνισης των πληροφοριών των tweets που έχουν συλλεχθεί και αναλύονται στο πλαίσιο της παρούσας μελέτης, εξαιτίας των

δυνατοτήτων που περιλαμβάνουν τα Google Maps API και των σχετικών γνώσεων που διαθέτει ο ερευνητής στη χρήση των αντίστοιχων διεπαφών προγραμματισμού (APIs).

4.4.2 Μοντέλο προγραμματισμού .NET

Το μοντέλο προγραμματισμού .NET Framework αποτελεί ένα ολοκληρωμένο μοντέλο προγραμματισμού για την υλοποίηση όλων των ειδών εφαρμογών σε λειτουργικά συστήματα Windows, καθώς επίσης και εφαρμογές για κινητά τηλέφωνα και ιστοσελίδες.

Βασικά πλεονεκτήματα της απόφασης χρήσης του μοντέλου προγραμματισμού .NET framework αποτελούν τα παρακάτω (ITeLearn, 2016):

- Το μέγεθος του κώδικα μιας εφαρμογής είναι περιορισμένο ιδιαίτερα σε εφαρμογές μεγάλης κλίμακας.
- Η ανάπτυξη εφαρμογών με το μοντέλο προγραμματισμού .NET θεωρείται ιδιαίτερα ασφαλής, καθώς υπάρχει αυτόματος διαχωρισμός της διεπαφής χρήσης και του κώδικα (επιχειρηματική λογική) μιας εφαρμογής ή ιστότοπου.
- Το .NET framework αποτελεί μια τεχνολογία διακομιστή, οπότε ο κώδικας .NET εκτελείται στον διακομιστή (για διαδικτυακές εφαρμογές) πριν κατευθυνθεί στον διακομιστή ιστού.
- Η πλατφόρμα .NET είναι συμβατή με πολλές γλώσσες προγραμματισμού (πχ C#, Java, C++, Visual Basic).
- Η πλατφόρμα .NET έχει σχεδιαστεί με γνώμονα την ευκολία χρήσης και επαναχρησιμοποίησης κώδικα και υποστηρίζει πολλές γλώσσες προγραμματισμού που μπορούν να αλληλεπιδρούν μεταξύ τους.
- Η πλατφόρμα .NET λειτουργεί σε αντικειμενοστραφές περιβάλλον προγραμματισμού, βοηθώντας στην εξάλειψη περιττών τμημάτων κώδικα και απαιτώντας μικρότερη σε έκταση κωδικοποίηση για τους προγραμματιστές.

Ως μειονέκτημα κατά την επιλογή χρήσης του μοντέλου προγραμματισμού .NET αποτελούν οι μεγαλύτερες υπολογιστικές απαιτήσεις κατά την εκτέλεση των εφαρμογών αναπτυγμένες με το μοντέλο .NET σε σχέση με εφαρμογές που έχουν αναπτυχθεί με άλλες γλώσσες προγραμματισμού και δύναται να χρησιμοποιούν περισσότερο αποτελεσματικά τους υπολογιστικούς πόρους ενός υπολογιστή (ITeLearn, 2016).

Ο λόγος που επιλέχτηκε η χρήση του μοντέλου προγραμματισμού .NET σχετίζεται με την χρήση της βάση δεδομένων MS SQL Server που επιτρέπει την ταχύτερη διασύνδεση με τη σχετική βάση δεδομένων, καθώς επίσης και η διάθεση σχετικών γνώσεων προγραμματισμού εκ μέρους του

ερευνητή, που διευκολύνει την ταχύτερη και αποτελεσματικότερη ανάπτυξη του ιστότοπου οπτικής απεικόνισης μηνυμάτων στο Twitter.

4.4.3 Η γλώσσα προγραμματισμού C#

Η γλώσσα προγραμματισμού C# είναι μια γλώσσα ειδικά σχεδιασμένη για την πλατφόρμα .NET, σε αντίθεση με τη Visual Basic, τη Java και C++ που έχουν προσαρμοστεί για να υποστηρίζουν την ανάπτυξη εφαρμογών με χρήση του μοντέλου προγραμματισμού .NET. Βασικά πλεονεκτήματα της C# (Niazi, 2014):

- Πρόκειται για μια αμιγώς αντικειμενοστραφή γλώσσα προγραμματισμού
- Διαθέτει ένα πολύ αποτελεσματικό σύστημα για τη διαγραφή και αφαίρεση όλων των κατάλοιπων χρήσης μνήμης του συστήματος ή δέσμευση υπολογιστικών πόρων εκ μέρους μιας εφαρμογής
- Αυτόματη διαχείριση μνήμης, χωρίς την απαίτηση σχετικής μέριμνας εκ μέρους των προγραμματιστών
- Ταχύτατη διασύνδεση με άλλες εφαρμογές που έχουν αναπτυχθεί με το μοντέλο προγραμματισμού .NET
- Ταχύτατη διασύνδεση με βάσεις δεδομένων και ειδικότερα με τη βάση δεδομένων MS SQL Server που θα χρησιμοποιηθεί για την προσπέλαση των σχετικών δεδομένων από τον ιστότοπο οπτικής απεικόνισης των tweets σε χάρτες

Όπως αναφέρθηκε παραπάνω η γλώσσα C# αποτελεί τη γλώσσα που ταιριάζει καλύτερα στις δυνατότητες που προσφέρει το μοντέλο προγραμματισμού .NET και το γεγονός αυτό αποτελεί τον λόγο επιλογής της αντίστοιχης γλώσσας προγραμματισμού. Με τη χρήση της C# θα υλοποιηθούν οι κλήσεις σε αποθηκευμένες διαδικασίες (stored procedures) στη βάση δεδομένων των tweets που συλλέχτηκαν στο πλαίσιο της παρούσας μελέτης.

4.4.4 Η γλώσσα σεναρίων jQuery

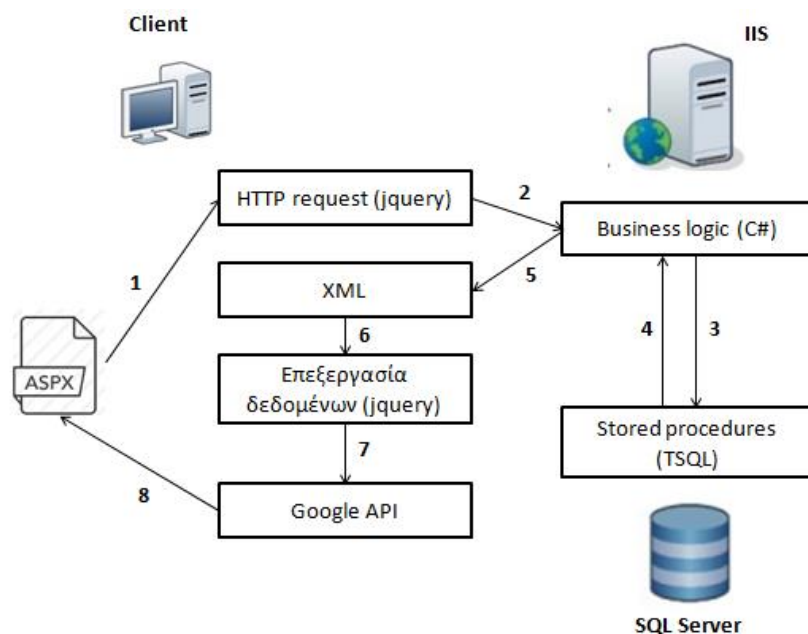
Η jQuery είναι μια γρήγορη στην εκτέλεση γλώσσα σεναρίων και συνοπτική βιβλιοθήκη JavaScript που απλοποιεί τη μετάδοση εγγράφων HTML, τον χειρισμό συμβάντων σε ιστοσελίδες και χρησιμοποιείται για την υποστήριξη ταχείας ανάπτυξης ιστοσελίδων (Rapid Application Development). Βασικά πλεονεκτήματα της γλώσσας σεναρίων jQuery είναι τα ακόλουθα (Asokan, 2013):

- Είναι δωρεάν διαθέσιμη

- Μικρή στο μέγεθος και συνεπώς ταχύτατη στην εκτέλεση χωρίς ιδιαίτερη επιβάρυνση του φυλλομετρητή
- Υποστηρίζεται από το σύνολο των διαθέσιμων φυλλομετρητών στην αγορά
- Ευκολία στη χρήση, καθώς πρόκειται πρακτικά για βιβλιοθήκη JavaScript
- Διαθεσιμότητα χιλιάδων δωρεάν σεναρίων jQuery για ενσωμάτωση σε ιστοσελίδες και προσθήκη σχετικών δυνατοτήτων σε εκείνες
- Απαιτήση για μικρότερες σε έκταση γραμμές κώδικα σε σχέση με άλλες γλώσσες σεναρίων όπως η JavaScript ή VBScript

Η γλώσσα σεναρίων jQuery επιλέχτηκε καθώς αποτελεί την προτεινόμενη γλώσσα σεναρίων κλήσεων προς μεθόδους των Google Maps API και η Google διαθέτει στον ιστότοπό της όλα τα σχετικά παραδείγματα διαχείρισης Google Maps στην αντίστοιχη γλώσσα σεναρίων. Η γλώσσα jQuery θα χρησιμοποιηθεί στην ιστοσελίδα οπτικής απεικόνισης των tweets για την ασύγχρονη λήψη σχετικών δεδομένων από τη βάση δεδομένων μέσω κλήσεων σε υπηρεσίες (web services) υλοποιημένες σε γλώσσα προγραμματισμού C# σε περιβάλλον .NET framework.

Στο παρακάτω διάγραμμα απεικονίζεται η αρχιτεκτονική λειτουργίας της εφαρμογής οπτικής απεικόνισης πληροφοριών που περιλαμβάνουν τα tweets που συλλέχτηκαν στο πλαίσιο της παρούσας έρευνας:



Σχήμα 5: Αρχιτεκτονική λειτουργίας της εφαρμογής οπτικής απεικόνισης tweets

ΚΕΦ.5: Υλοποίηση

5.1 Συλλογή δεδομένων

Για τη συλλογή δεδομένων αναπτύχθηκε εφαρμογή με χρήση της Python για την ανάκτηση αναρτήσεων tweets στο Twitter με χρήση του Twitter API. Τα σχετικά tweets συλλέχθηκαν σε πραγματικό χρόνο την περίοδο 20/12/2016 έως και 22/01/2017 και αποθηκεύτηκαν σε βάση δεδομένων με σκοπό τη μετέπειτα επεξεργασία τους. Σε δεύτερο χρόνο χρησιμοποιήθηκε το εργαλείο REST API, με σκοπό την συλλογή επιπροσθέτως πληροφοριών για τα tweets που συλλέχθηκαν με χρήση του Twitter API.

Η εφαρμογή δεδομένων λήψης των δεδομένων που αναπτύχθηκε περιλαμβάνει τα εξής αρχεία Python:

- **auth.py**: δήλωση διαπιστευτηρίων σύνδεσης χρήσης της εφαρμογής Twitter API.

```
consumer_key = "oM3MZdZBg7M6K00UWKyFFsqSx"
consumer_secret = "bnAC76XUIhPbMInmT9cAtieXmVzs4DIj2F0yvGpBenzaVLFaah"
access_token = "802493452218986496-5UBID2TqvSKnaw1FQP97Ua6ZzGQcGBx"
access_token_secret = "2NYW1WEo8Yh3ld51gQoUls1Mz4m4gILLsrPvqq1orVqIm"
```

- **database.py**: προσδιορισμός των παραμέτρων διασύνδεσης της εφαρμογής με τη βάση δεδομένων PostgreSQL. Ακολουθεί η δήλωση προσδιορισμού της βάσης δεδομένων που θα χρησιμοποιηθεί για την αποθήκευση των tweets:

```
create_engine('postgresql://tweetsql:tweetsql@127.0.0.1/tweetsql')
```

- **location_stream.py**: προσδιορισμός ορίων γεωγραφικών συντεταγμένων των tweets που θα αποθηκευτούν στη βάση δεδομένων. Η δήλωση των σχετικών ορίων γεωγραφικών συντεταγμένων (γεωγραφικό πλάτος, γεωγραφικό μήκος) παρουσιάζεται παρακάτω:

```
stream.filter(locations = [-74.25888888888889,40.4772222,-73.7,40.9175,-
0.3513888888888886,51.3847222,0.1480555555555555,51.6722222,-
77.119759,38.7916449,-76.909393,38.995548,-118.6681759,33.7036519,-
118.1552891,34.3373061])
```

Σε κάθε λήψη ενός καινούργιου tweet ελέγχεται αν περιλαμβάνει δήλωση συντεταγμένων, αν όχι διακόπτεται η επεξεργασία του αντίστοιχου tweet. Σε διαφορετική περίπτωση ελέγχονται αν οι συντεταγμένες του αντίστοιχου tweet βρίσκονται εντός των ορίων που έχουν δηλωθεί. Αν όχι διακόπτεται η επεξεργασία του αντίστοιχου tweet. Για όσα από τα tweets που λαμβάνει η εφαρμογή περιλαμβάνονται

πληροφορίες γεωγραφικών συντεταγμένων ανάρτησης και οι σχετικές γεωγραφικές συντεταγμένες είναι εντός των δηλωθέντων ορίων, η εφαρμογή προχωράει στην αποθήκευση των πληροφοριών που συνοδεύουν το αντίστοιχο tweet στη βάση δεδομένων του συστήματος.

- **model.py:** μοντελοποίηση - αντιστοίχιση πεδίων πίνακα βάσης δεδομένων και διαχείριση σχετικών εντολών ενημέρωσης της βάσης δεδομένων. Ακολουθεί η δήλωση των πεδίων του πίνακα που θα χρησιμοποιηθεί για την αποθήκευση των tweets:

```
__tablename__ = 'tweet'
id = Column(BigInteger, primary_key=True)
tweet_text = Column(String(300), nullable=False)
user_id = Column(String(50), nullable=False, index=True)
coordinates = Column(String(100), nullable=True)
created_at = Column(String(100), nullable=False)
place = Column(String(100), nullable=False)
retweet_count = Column(Integer, nullable=False, default=0, server_default='0')
favorite_count = Column(Integer, nullable=False, default=0, server_default='0')
reply_count = Column(Integer, nullable=False, default=0, server_default='0')
last_update = Column(DateTime, nullable=True, index=True)
json = Column(JSON, nullable=True)
```

- **refresh_tweet_data.py:** χρησιμοποιείται μετά τη συλλογή και αποθήκευση όλων των tweets στη βάση δεδομένων της εφαρμογής. Αρχικά προσπελούνται στη βάση δεδομένων το σύνολο των tweets και όσα διαπιστωθούν ότι αποτελούν απάντηση άλλων αποθηκευμένων tweets στη βάση tweets ενημερώνεται το πεδίο reply_count στη βάση δεδομένων του συστήματος:

```
for tweet_id in Tweet.all_with_json():
    tweet = Tweet.query.get(tweet_id)
    reply_to = tweet.json.get('in_reply_to_status_id')
    if reply_to is not None:
        Tweet.increment_replies(reply_to)
```

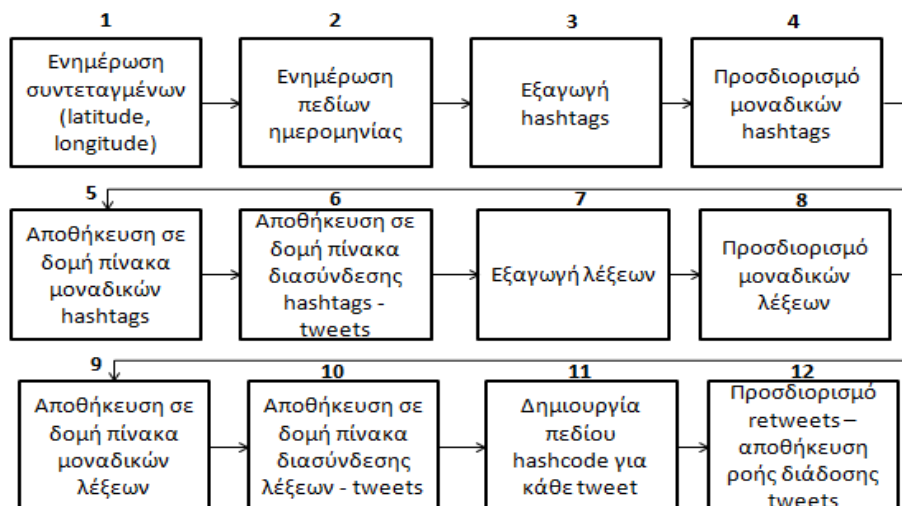
Κατόπιν, με χρήση του REST API προσπελούνται εκ νέου οι πληροφορίες όλων των αποθηκευμένων και ενημερώνονται οι πληροφορίες της τρέχουσας κατάστασης των αντίστοιχων tweets.

Μετά το πέρας λήψης των δεδομένων με χρήση της εφαρμογής που αναπτύχθηκε σε Python, τα δεδομένα των tweets αποθηκεύτηκαν σε έναν πίνακα σε βάση δεδομένων Postgresql με τα εξής πεδία τιμών:

- **id**: μοναδικός κωδικός για κάθε tweet
- **tweet_text**: το κείμενο ανάρτησης του αντίστοιχου tweet
- **user_id**: μοναδικός κωδικός του χρήστη που ανήρτησε το αντίστοιχο tweet
- **coordinates**: συντεταγμένες του γεωγραφικού σημείου ανάρτησης
- **created_at**: ημερομηνία δημοσίευσης του αντίστοιχου tweet
- **place**: τοποθεσία δημοσίευσης του σχετικού tweet
- **retweet_count**: πόσες φορές έχει γίνει retweet το αντίστοιχο tweet
- **favorite_count**: πόσες φορές το αντίστοιχο tweet έγινε liked από άλλους χρήστες στο Twitter
- **last_update**: ημερομηνία τελευταίας ενημέρωσης περιεχομένου του tweet
- **reply_count**: ο αριθμός απαντήσεων στο αντίστοιχο tweet
- **json**: το σύνολο των πληροφοριών που επέστρεψε το TWITTER API για το αντίστοιχο tweet σε μορφή αρχείου κειμένου τύπου json

5.2 Επεξεργασία δεδομένων

Τα δεδομένα με όλα τα tweets αποθηκεύτηκαν σε δομή πίνακα με το όνομα Twitter στη βάση δεδομένων. Όπως περιγράφονται παραπάνω τα δεδομένα που συλλέχτηκαν με χρήση του Twitter API δεν είναι πρακτικά έτοιμα για την οπτική απεικόνισή τους σε χάρτη Google Map. Στο παρακάτω σχήμα απεικονίζονται τα στάδια – βήματα επεξεργασίας των δεδομένων που συλλέχτηκαν πριν τη χρήση τους από την εφαρμογή που αναπτύχθηκε στο πλαίσιο της παρούσας έρευνας:



Σχήμα 6: Προσδιορισμός βημάτων επεξεργασίας δεδομένων σε επίπεδο βάσης δεδομένων

Οπτική απεικόνιση διάδοσης και περιεχομένου αναρτήσεων στο Twitter

Ως πρώτο βήμα για την επεξεργασία τους, μεταφέρθηκε το σύνολο των δεδομένων σε βάση δεδομένων MS SQL Server που υποστηρίζει τη διαχείριση κειμένου JSON. Μια πρώτη επεξεργασία που απαιτήθηκε ήταν ο προσδιορισμός ξεχωριστά του γεωγραφικού μήκους και πλάτους κάθε ανάρτησης tweet, καθώς το Twitter API είχε αποθηκεύσει τη σχετική πληροφορία στην βάση δεδομένων σε πεδίο με τύπο αλφαριθμητικού (varchar) και όχι πραγματικών αριθμών (real). Ειδικότερα οι συντεταγμένες ήταν αποθηκευμένες με κείμενο της μορφής: {-74.0007613,40.7207559}. Για τον προσδιορισμό των τιμών γεωγραφικού μήκους και πλάτους σε πραγματικές τιμές (real) χρησιμοποιήθηκε ερώτημα της μορφής (Βήμα 1):

```
SELECT cast(
substring(substring(coordinates,2,len(coordinates)-2),
0,CHARINDEX(',', substring(coordinates,2,len(coordinates)-2)))
as real) as latitude,
cast(
substring(substring(coordinates,2,len(coordinates)-2),
CHARINDEX(',', substring(coordinates,2,len(coordinates)-2))+1,100)
as real) as longitude
```

Επιπροσθέτως, η ημερομηνία ανάρτησης των σχετικών tweets ήταν επίσης αποθηκευμένη στη βάση δεδομένων με τη μορφή αλφαριθμητικού της εξής μορφής: *Mon Dec 26 05:01:43 +0000 2016*. Για την μετατροπή της ημερομηνίας σε πεδίο ημερομηνίας κατανοητής από τη βάση δεδομένων εκτελέστηκε το παρακάτω ερώτημα (Βήμα 2):

```
convert(datetime,
(substring (created_at,9,2)+ ' ' +substring (created_at,5,3)+ ' ' +
substring (created_at,27,4) + ' ' +substring (created_at,12,2) + ':' +
substring (created_at,15,2)+ ':' +substring (created_at,18,2) ) ) as imerominia
```

Επίσης, για κάθε tweet ήταν γνωστό μονάχα το κείμενο του tweet και όχι αναλυτικά ποια hashtags περιλαμβάνονται σε κάθε tweet ή ποιες συγκεκριμένες λέξεις περιλαμβάνονται σε κάποιο tweet, καθώς επίσης και ο αριθμός επανάληψής τους. Για το σκοπό αυτό δημιουργήθηκε συνάρτηση (function) στον SQL Server με σκοπό την εξαγωγή των hashtag που περιλαμβάνει κάθε κείμενο tweet (Βήμα 3):

```
CREATE FUNCTION [dbo].[SplitWords](@text varchar(8000))
RETURNS @words TABLE (
value varchar(8000)
)
AS
BEGIN
DECLARE
@wordi varchar(3000),
@pos smallint,
@i smallint,
@j smallint,
@s varchar(8000)
```



```

        SET @pos = 1
    WHILE @pos <= LEN(@text)
    BEGIN
        SET @i = CHARINDEX(' ', @text, @pos)
        SET @j = CHARINDEX(',', @text, @pos)
        IF @i > 0 OR @j > 0
        BEGIN
            IF @i = 0 OR (@j > 0 AND @j < @i)
                SET @i = @j

            IF @i > @pos
            BEGIN
                SET @s = SUBSTRING(@text, @pos, @i - @pos)
                if (left(@s,1)='#')
                begin
                    INSERT INTO @words
                    VALUES (@s)
                end
            END
            SET @pos = @i + 1

            WHILE @pos < LEN(@text)
                AND SUBSTRING(@text, @pos, 1) IN (' ', ',')
                SET @pos = @pos + 1
        END
    ELSE
    BEGIN
        set @wordi=SUBSTRING(@text, @pos, LEN(@text) - @pos + 1)
        if (left(@wordi,1)='#')
        begin
            INSERT INTO @words
            VALUES (@wordi)
        end
        SET @pos = LEN(@text) + 1
    END
END
RETURN
END

```

Στη συνέχεια με την εκτέλεση του παρακάτω ερωτήματος προσδιορίστηκαν τα σχετικά μοναδικά hashtag και αποθηκεύτηκαν σε ξεχωριστό πίνακα (gHashTags) στη βάση δεδομένων (Βήμα 4, 5):

```

insert into gHashTags(hashtag)
SELECT distinct Results.value
FROM Twitter AS Stuff
OUTER APPLY dbo.SplitWords(Stuff.tweet_text) AS Results
WHERE ISNULL(Results.value, '') <> ''

```

Κατόπιν για την αντιστοιχία ποια hashtags αντιστοιχούν σε ποιά tweets εκτελέστηκε το παρακάτω ερώτημα και τα αποτελέσματα αποθηκεύτηκαν σε ξεχωριστό ενδιάμεσο πίνακα (gLink) μεταξύ του πίνακα Twitter (όλα τα tweets) και του πίνακα gHashTags (όλα τα hashtags) (Βήμα 6):

```

insert into glink(hashtag,Twitter,imerominia,lat,lon)
SELECT distinct gHashTags.aa ,Stuff.aa, convert(datetime,
(substring (created_at,9,2)+ ' '+substring (created_at,5,3)+' '+
substring (created_at,27,4) + ' '+substring (created_at,12,2) + ':' +
substring (created_at,15,2)+ ':' +substring (created_at,18,2) ) ) ,
cast(
substring(substring(coordinates,2,len(coordinates)-2),

```

Οπτική απεικόνιση διάδοσης και περιεχομένου αναρτήσεων στο Twitter

```

0,CHARINDEX(',', substring(coordinates,2,len(coordinates)-2)))
as real),
cast(
substring(substring(coordinates,2,len(coordinates)-2),
CHARINDEX(',', substring(coordinates,2,len(coordinates)-2))+1,100)
as real)
FROM Twitter AS Stuff
OUTER APPLY dbo.SplitWords(Stuff.tweet_text) AS Results
inner join gHashTags on Results.value=gHashTags.hashtag
WHERE ISNULL(Results.value, '') <> ''

```

Με ανάλογο τρόπο δημιουργήθηκε πίνακας (gWord) με όλες τις λέξεις που περιλαμβάνονται στο σύνολο των tweets που συλλέχθηκαν και ενδιάμεσος πίνακας (gLinkWords) με πληροφορίες ποιες λέξεις περιλαμβάνονται σε κάθε tweet ξεχωριστά (Βήμα 7, 8, 9, 10).

Επίσης, καθώς με τη χρήση του Twitter API δεν κατέστη δυνατόν η άμεση λήψη της πληροφορίας ποιο τυχόν tweet αποτελεί retweet ενός άλλου, εναλλακτικά για κάθε tweet, για το κείμενο του αντίστοιχου tweet (tweet_text) υπολογίστηκε τιμή hashcode με χρήση της εντολής (Βήμα 11):

```

SELECT HashBytes('SHA1', tweet_text)
from Twitter

```

Κατόπιν συγκρίνοντας τις τιμές μεταξύ των hashcode, διαπιστώθηκαν ποια είναι όμοια tweets, δηλαδή αποτελούν retweets. Τα σχετικά tweets επισημάνθηκαν στον πίνακα tweets, θέτοντας την τιμή του πεδίου isTweet (τύπου bit) ίσο με αληθές (true).

Στη συνέχεια προσδιορίστηκε για κάθε ομάδα όμοιων tweet ποιο το πρώτο tweet που δημοσιεύτηκε και ακολούθησε η δημοσίευση (retweet) των λοιπών tweets (Βήμα 12a):

```

declare @hash as binary(50)
declare @mini as datetime

DECLARE vendor_cursor CURSOR FOR
select distinct hashcode from Twitter
where isTweet=1

OPEN vendor_cursor

FETCH NEXT FROM vendor_cursor INTO @hash

declare @a as integer
set @a=0

WHILE @@FETCH_STATUS = 0
BEGIN

select @mini=min(imerominia) from Twitter where hashcode=@hash

update Twitter set isFirst=1 where hashcode=@hash and imerominia=@mini
and aa in (select min(aa) from Twitter where hashcode=@hash and imerominia=@mini)

if (@a%100=0)
begin
print @a
end

set @a=@a+1

```

```

FETCH NEXT FROM vendor_cursor INTO @hash
end
CLOSE vendor_cursor;
DEALLOCATE vendor_cursor;

```

Τέλος, για κάθε ομάδα όμοιων tweet αποθηκεύτηκε σε πεδίο στον πίνακα Twitter αύξων αριθμός της σειράς δημοσίευσής τους, με βάση την ημερομηνία δημοσίευσης (Βήμα 12b):

```

declare @hash as binary(50)
declare @mini as datetime

declare @aa as bigint
declare @seira as bigint

DECLARE vendor_cursor CURSOR FOR
select distinct hashcode from Twitter
where isFirst=1

OPEN vendor_cursor

FETCH NEXT FROM vendor_cursor INTO @hash

declare @a as integer
set @a=0

WHILE @@FETCH_STATUS = 0
BEGIN

DECLARE cur CURSOR FOR
SELECT aa, ROW_NUMBER() OVER (ORDER BY imerominia asc) AS New_CODE_DEST
FROM Twitter where hashcode=@hash and isFirst=0

OPEN cur

FETCH NEXT FROM cur INTO @aa,@seira

WHILE @@FETCH_STATUS = 0
BEGIN

FETCH NEXT FROM cur INTO @aa,@seira
end

CLOSE cur;
DEALLOCATE cur;

UPDATE x
SET x.tweetOrder = x.New_CODE_DEST
FROM (
    SELECT tweetOrder, ROW_NUMBER() OVER (ORDER BY imerominia asc) AS New_CODE_DEST
    FROM Twitter where hashcode=@hash and isFirst=0
) x

if (@a%100=0)
begin
print @a
end

set @a=@a+1

```

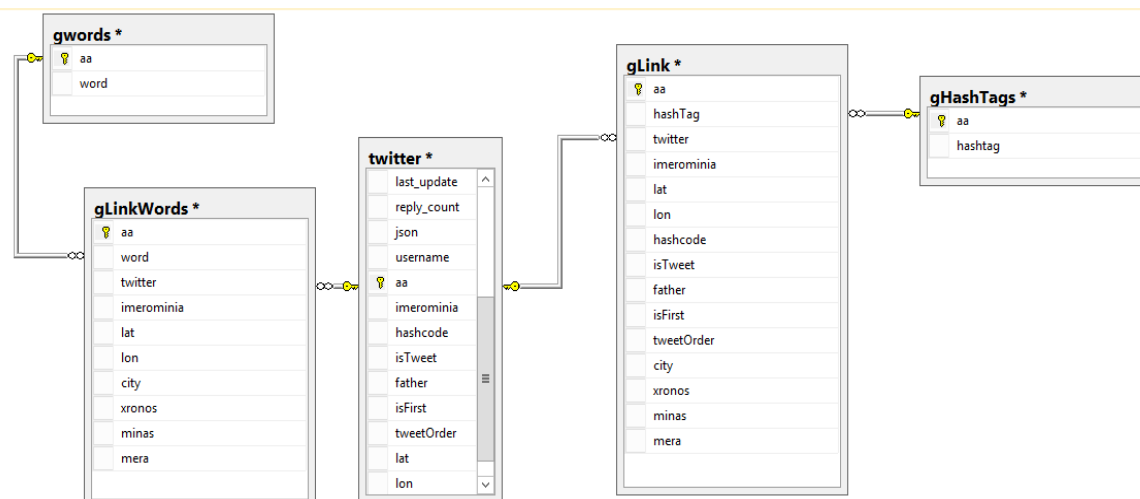
```
FETCH NEXT FROM vendor_cursor INTO @hash
end
CLOSE vendor_cursor;
DEALLOCATE vendor_cursor;
```

Επίσης, στη βάση δεδομένων δημιουργήθηκαν οι παρακάτω αποθηκευμένες διαδικασίες (stored procedures) για τη διευκόλυνση εκτέλεσης του ιστότοπου οπτικής απεικόνισης των tweets που περιγράφεται παρακάτω:

- **getAllByHashtag**: προσπέλαση των συντεταγμένων όλων των tweets που περιλαμβάνουν συγκεκριμένο hashtag και δημοσιεύτηκαν συγκεκριμένη ημερομηνία
- **getAllByHashCodeTweet**: συντεταγμένες όλων των retweets συγκεκριμένου tweet για συγκεκριμένη ημερομηνία
- **getAllTweets**: προσπέλαση των συντεταγμένων όλων των αποθηκευμένων tweets
- **getStartInfo**: αρχική και τελική ημερομηνία δημοσίευσης tweet που περιλαμβάνουν συγκεκριμένο hashtag, καθώς επίσης και το σύνολο των σχετικών tweets
- **getStartInfoReTweet**: αρχική και τελική ημερομηνία όλων των retweets συγκεκριμένου tweet, καθώς επίσης και το σύνολο των σχετικών retweets
- **getStartingList**: λίστα με τα δέκα (10) δημοφιλέστερα hashtag για το σύνολο των tweets που συλλέχτηκαν
- **getStartingListRetweet**: λίστα με τα δέκα (10) tweets με τον μεγαλύτερο αριθμό retweets
- **topHashAll**: προσπέλαση των δημοφιλέστερων hashtag ανά πόλη για συγκεκριμένη ημερομηνία ή συνολικά για όλα τα tweets
- **topWordsAll**: προσπέλαση των δημοφιλέστερων λέξεων ανά πόλη για συγκεκριμένη ημερομηνία ή συνολικά για όλα τα tweets

Επιπροσθέτως, για την αποτελεσματική προσπέλαση των δεδομένων και την εκτέλεση ερωτημάτων SQL, δημιουργήθηκαν ευρετήρια (indexes) σε όλα τα πεδία των πινάκων της βάσης που χρησιμοποιούνται στις αποθηκευμένες διαδικασίες που αναφέρθηκαν παραπάνω.

Στο παρακάτω διάγραμμα απεικονίζεται το σχήμα της βάσης δεδομένων που θα χρησιμοποιείται για την οπτική απεικόνιση των tweets που συλλέχτηκαν:



Σχήμα 7: Το σχήμα - σχέσεις βάσης δεδομένων ιστότοπου

5.3 Παρουσίαση δεδομένων

Συνολικά συλλέχθηκαν 1.215.023 tweets, με το πρώτο εξ αυτών να έχει δημοσιευτεί στην κοινωνική πλατφόρμα στις 2016-12-20 και ώρα 22:41:23 και το τελευταίο στις 2017-01-25 και ώρα 13:16:07. Τα αντίστοιχα tweets δημοσιεύτηκαν από 186.457 χρήστες με κατά μέσο όρο 6,51 tweets ανά χρήστη. Ο παρακάτω πίνακας παρουσιάζει τους δέκα (10) χρήστες με τον μεγαλύτερο αριθμό αναρτήσεων στο σύνολο των tweets που συλλέχθηκαν στο πλαίσιο της παρούσας μελέτης:

Πίνακας 5: Οι χρήστες με τις περισσότερες δημοσιεύσεις - tweets

Ψευδώνυμο χρήστη	Κωδικός χρήστη	Αριθμός tweets
Χρήστης 1	176046	23.670
Χρήστης 2	31068	15.384
Χρήστης 3	93238	14.091
Χρήστης 4	186402	13.366
Χρήστης 5	140390	9.474
Χρήστης 6	147024	6.807
Χρήστης 7	177685	6.579
Χρήστης 8	21201	6.246
Χρήστης 9	23802	5.399
Χρήστης 10	152903	4.817

Επίσης στο σύνολο των tweets που συλλέχθηκαν εμφανίζονται 255.395 μοναδικά hashtags, τα οποία συνολικά εμφανίζονται 1.995.437 φορές στο σύνολο των tweets που συλλέχθηκαν. Στον

παρακάτω πίνακα παρουσιάζονται τα δέκα (10) περισσότερο δημοφιλή hashtags με τον μεγαλύτερο αριθμό εμφανίσεων στα tweets που συλλέχτηκαν:

Πίνακας 6: Τα δέκα (10) δημοφιλέστερα hashtags

Hashtag	Αριθμός tweets
#job	138.788
#Hiring	119.031
#CareerArc	79.015
#Jobs	35.658
#NewYork	33.333
#hiring!	31.295
#LosAngeles	29.215
#Hospitality	24.948
#trndnl	22.510
#job:	16.303

Επιπροσθέτως, στο σύνολο των αναρτήσεων - tweets που συλλέχτηκαν περιλαμβάνονται συνολικά 13.384.363 λέξεις που περιλαμβάνουν 1.633.710 μοναδικές λέξεις. Στον παρακάτω πίνακα παρουσιάζονται οι δέκα (10) περισσότερο δημοφιλείς λέξεις (τουλάχιστον πέντε χαρακτήρων) με τον μεγαλύτερο αριθμό εμφανίσεων στα tweets που συλλέχτηκαν:

Πίνακας 7: Οι δέκα (10) δημοφιλέστερες λέξεις

Λέξη	Αριθμός tweets
latest	65.447
click	48.017
great	45.231
Incident	38.611
anyone	33.774
London	33.511
recommend	32.855
Angeles	27.376
photo	24.877
Christmas	22.301

Σχετικά με τις γεωγραφικές περιοχές ανάρτησης των tweets, οι δέκα (10) πόλεις – περιοχές με τον μεγαλύτερο αριθμό αναρτήσεων tweets είναι οι εξής:

Πίνακας 8: Οι πόλεις με τον μεγαλύτερο αριθμό αναρτήσεων - tweets

Πόλη	Αριθμός tweets
Manhattan, NY	236.603
Los Angeles, CA	173.842
London, England	80.050
Washington, DC	78.932
Brooklyn, NY	55.118
Queens, NY	38.106
Bronx, NY	16.169
Arlington, VA	13.113
South East, England	12.880
Long Beach, CA	11.705

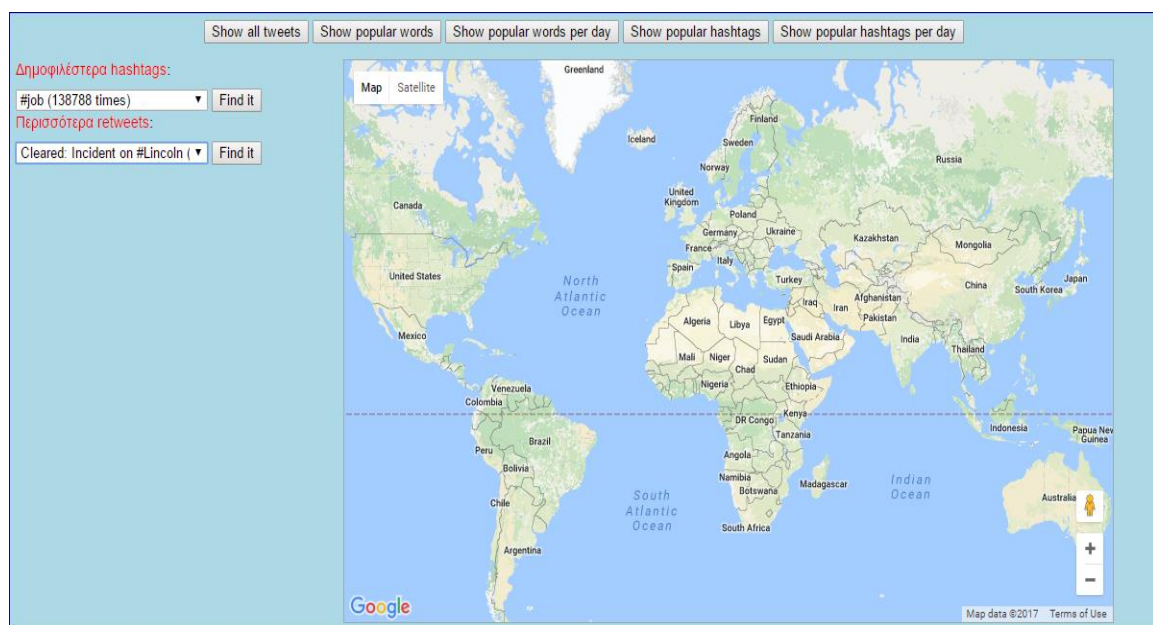
5.4 Υλοποίηση ιστοσελίδας απεικόνισης

Η ιστοσελίδα υλοποιήθηκε σε περιβάλλον Visual Studio 2015. Πέραν της ιστοσελίδας (**map.aspx**) οπτικής απεικόνισης των tweets, δημιουργήθηκαν οι παρακάτω σελίδες **aspx**, που λειτουργούν ως web services και καλούνται από την ιστοσελίδα (**map.aspx**) ασύγχρονα με χρήση της γλώσσας σεναρίων jQuery:

- **getAllByHashtag.aspx**: κλήση προς τη βάση δεδομένων περί των συντεταγμένων των tweets που περιλαμβάνουν συγκεκριμένο hashtag και δημοσιεύτηκαν σε συγκεκριμένη ημερομηνία
- **getAllByHashtagTweet.aspx**: κλήση προς τη βάση δεδομένων περί των συντεταγμένων των retweets για συγκεκριμένο tweet και δημοσιεύτηκαν σε συγκεκριμένη ημερομηνία
- **getAllTweets.aspx**: κλήση προς τη βάση δεδομένων περί των συντεταγμένων όλων των αποθηκευμένων tweets
- **getStartData.aspx**: κλήση προς τη βάση δεδομένων περί της αρχικής και τελικής ημερομηνίας δημοσίευσης tweets που περιλαμβάνουν συγκεκριμένο hashtag
- **getStartDataRetweet.aspx**: κλήση προς τη βάση δεδομένων περί της αρχικής και τελικής ημερομηνίας δημοσίευσης retweets για συγκεκριμένο tweet

- **topHashAll.aspx**: κλήση προς τη βάση δεδομένων περί δημοφιλέστερων hashtag ανά πόλη για συγκεκριμένη ημερομηνία ή συνολικά για όλα τα tweets
- **topWordsAll.aspx**: κλήση προς τη βάση δεδομένων περί δημοφιλέστερων λέξεων ανά πόλη για συγκεκριμένη ημερομηνία ή συνολικά για όλα τα tweets

Οι σχετικές σελίδες (web services), καλούνται μέσω της βιβλιοθήκης που υλοποιήθηκε (jQuery.js) που περιλαμβάνει ασύγχρονες κλήσεις προς τα στοιχεία που βρίσκονται αποθηκευμένα στη βάση δεδομένων και στη συνέχεια απεικονίζει οπτικά τα αντίστοιχα δεδομένα με χρήση των Google Maps API. Στην παρακάτω εικόνα απεικονίζεται η ιστοσελίδα που δημιουργήθηκε στο πλαίσιο της παρούσας έρευνας:



Εικόνα 33: Διεπαφή ιστοσελίδας οπτικής απεικόνισης tweets

Αριστερά στην οθόνη εμφανίζεται λίστα με τα δέκα (10) δημοφιλέστερα hashtags. Πατώντας ο χρήστης το κουμπί “Find it”, επισημαίνεται στον χάρτη το σημείο της πρώτης αντίστοιχης δημοσίευσης και πληροφορίες σχετικά με την πρώτη και τελευταία ανάρτηση αντίστοιχου tweet και ο συνολικός αριθμός των αντίστοιχων tweet. Κατόπιν για κάθε μέρα απεικονίζονται οπτικά τα σχετικά tweets στον χάρτη Google Map.

Ακριβώς από κάτω εμφανίζεται λίστα με τα δέκα (10) δημοφιλέστερα tweets, όσον αφορά στον αριθμό των retweets που ακολούθησαν. Πατώντας ο χρήστης το κουμπί “Find it”, επισημαίνεται στο χάρτη το σημείο της πρώτης αντίστοιχης δημοσίευσης και πληροφορίες σχετικά με την πρώτη και τελευταία ανάρτηση αντίστοιχου tweet και ο συνολικός αριθμός των αντίστοιχων retweets. Κατόπιν για κάθε μέρα απεικονίζονται οπτικά τα σχετικά tweets στον χάρτη Google Map, συνοδευόμενα από βέλη διάδοσης – πηγής προέλευσης του κάθε retweet.

Οπτική απεικόνιση διάδοσης και περιεχομένου αναρτήσεων στο Twitter

Τέλος, στο πάνω μέρος της οθόνης εμφανίζονται τα παρακάτω κουμπιά που επιτελούν τις εξής λειτουργίες:

- **Show all tweets:** απεικονίζει το σύνολο όλων των tweets στον χάρτη Google Map
- **Show popular words:** απεικονίζει στον χάρτη συνολικά την δημοφιλέστερη λέξη για πέντε πόλεις (Νέα Υόρκη, Λος Άντζελες, Ουάσιγκτον, Σαν Φρανσίσκο, Λονδίνο)
- **Show popular words per day:** απεικονίζει στον χάρτη την δημοφιλέστερη λέξη για πέντε πόλεις (Νέα Υόρκη, Λος Άντζελες, Ουάσιγκτον, Σαν Φρανσίσκο, Λονδίνο) για κάθε μια μέρα ξεχωριστά
- **Show popular hashtags:** απεικονίζει στον χάρτη συνολικά το δημοφιλέστερο hashtag για πέντε πόλεις (Νέα Υόρκη, Λος Άντζελες, Ουάσιγκτον, Σαν Φρανσίσκο, Λονδίνο)
- **Show popular hashtags per day:** απεικονίζει στον χάρτη το δημοφιλέστερο hashtag για πέντε πόλεις (Νέα Υόρκη, Λος Άντζελες, Ουάσιγκτον, Σαν Φρανσίσκο, Λονδίνο) για κάθε μια μέρα ξεχωριστά

ΚΕΦ.6: Οπτική απεικόνιση

6.1 Απεικόνιση όλων των tweets που συλλέχθηκαν

Στη συνέχεια απεικονίζονται σε χάρτη Google Map των σύνολο των tweets που συλλέχθηκαν με τις μεθόδους: (i) συστοιχίες σήμανσης και (ii) οπτικός χάρτης εστίασης.

6.1.1 Συστοιχίες σήμανσης (marker clusters)

Στην παρακάτω εικόνα απεικονίζονται σε χάρτη όλα τα tweets που συλλέχθηκαν στο πλαίσιο της παρούσας μελέτης:



Εικόνα 34: Γεωγραφική απεικόνιση όλων των tweets σε χάρτη (χωρίς μεγέθυνση)

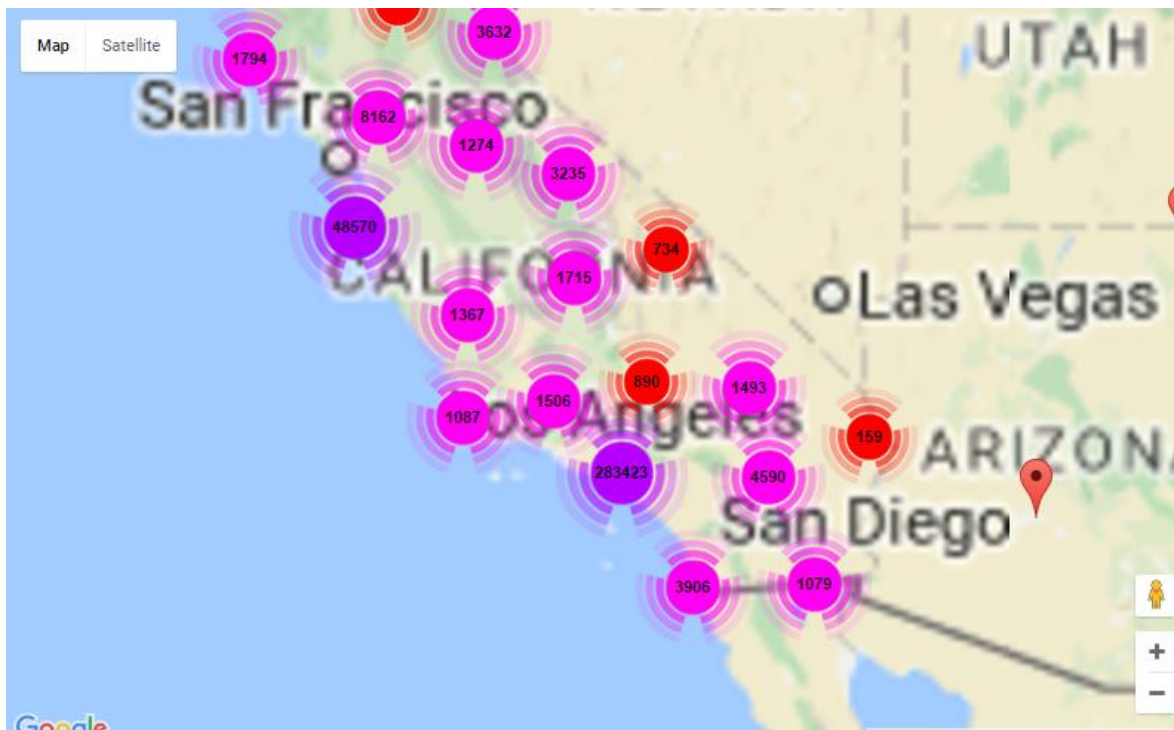
Στις βόρειες ανατολικές ακτές των ΗΠΑ (620.130) διαπιστώνονται τα περισσότερα tweets, ακολουθούμενα από τις νότιες δυτικές ακτές των ΗΠΑ (370.521) και την ευρύτερη περιοχή του Λονδίνου στο Ηνωμένο Βασίλειο (224.012). Όπως διαπιστώνεται από το παραπάνω διάγραμμα έχουν συλλεχθεί μικρά σε αριθμό tweets από περιοχές εκτός των ΗΠΑ και του Ηνωμένου Βασιλείου παρότι στον crawler συλλογής tweets είχε τεθεί σχετικός περιορισμός. Ωστόσο, ο αντίστοιχος αριθμός tweets είναι μικρότερος του 0,5% και τα σχετικά tweets θα αγνοηθούν κατά την οπτική απεικόνιση των tweets σε χάρτες Google Maps.

Επιλέγοντας, μεγέθυνση στον χάρτη του Google Map είναι δυνατή η προβολή με μεγαλύτερη ακρίβεια των πολιτειών ή πόλεων που αντιστοιχούν στα διάφορα tweets. Ενδεικτικά είναι δυνατός ο προσδιορισμός της κατανομής του αριθμού tweets μεταξύ των πόλεων Λος Άντζελες και Σαν Φρανσίσκο, οι οποίες χωρίς την μεγέθυνση χάρτη, αντιστοιχούσαν συνολικά σε αναρτήσεις – tweets στην ευρύτερη περιοχή της πολιτείας της Καλιφόρνια.



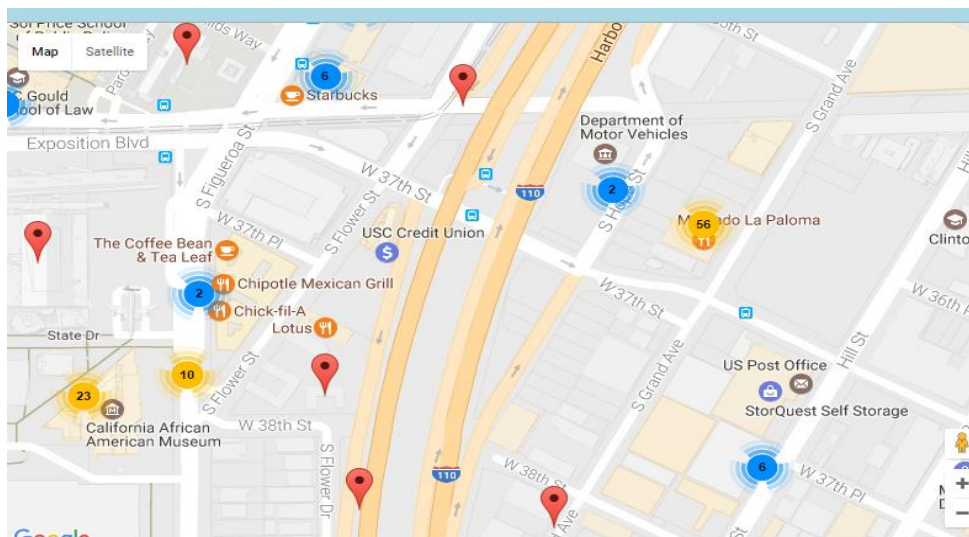
Εικόνα 35: Γεωγραφική απεικόνιση όλων των tweets σε χάρτη (με εφαρμογή μεγέθυνσης σε επίπεδο πολιτειών)

Εφαρμόζοντας μεγαλύτερη μεγέθυνση στον παραπάνω χάρτη είναι δυνατή η λεπτομερέστερη απεικόνιση της γεωγραφικής περιοχής που πραγματοποιήθηκε η ανάρτηση των σχετικών tweets όπως απεικονίζεται στην παρακάτω εικόνα σε επίπεδο πολιτείας (Καλιφόρνια):



Εικόνα 36: Γεωγραφική απεικόνιση όλων των tweets σε χάρτη (με εφαρμογή μεγέθυνσης σε επίπεδο πολιτείας - Καλιφόρνια)

Τέλος εφαρμόζοντας ακόμα μεγαλύτερη μεγέθυνση είναι δυνατή η απεικόνιση των tweets σε επίπεδο δρόμων πόλεων, όπως απεικονίζεται στην παρακάτω εικόνα:

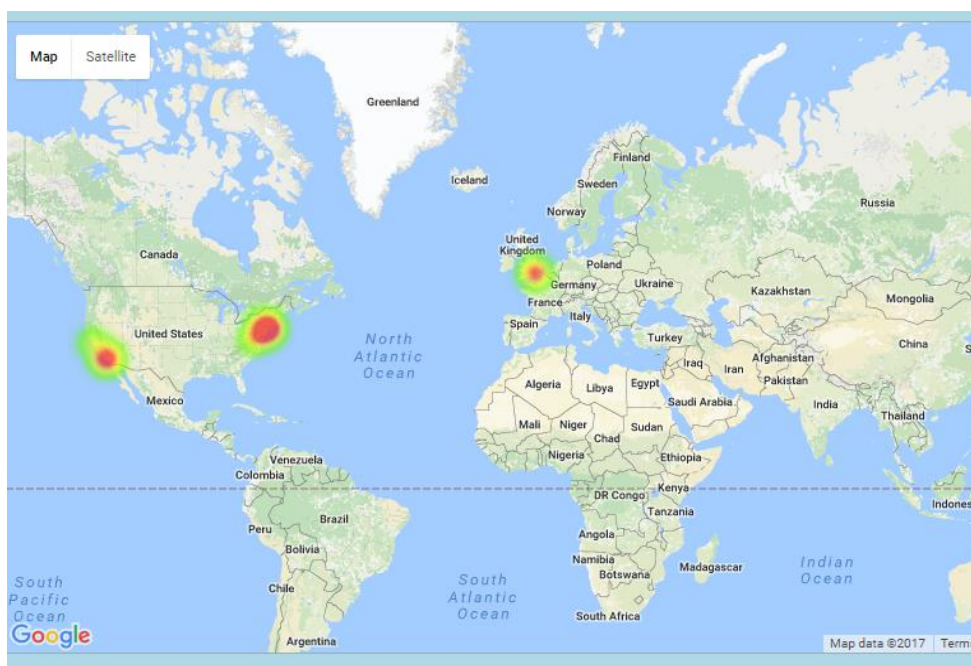


Εικόνα 37: Γεωγραφική απεικόνιση όλων των tweets σε χάρτη (με εφαρμογή μεγέθυνσης σε επίπεδο δρόμου πόλεων)

Από τον παραπάνω χάρτη παρατηρείται πως οι περισσότερες αναρτήσεις στο Twitter προέρχονται από δημόσιους χώρους αναψυχής, χώρους εστίασης και μουσεία. Πιθανόν, διότι οι χρήστες επιθυμούν να κοινοποιήσουν την παρουσία τους στους αντίστοιχους χώρους, καθώς επίσης και διότι είναι δυσκολότερη η ανάρτηση tweets εν κινήσει σε σχέση με χώρους αναψυχής, χώρους εστίασης και μουσεία.

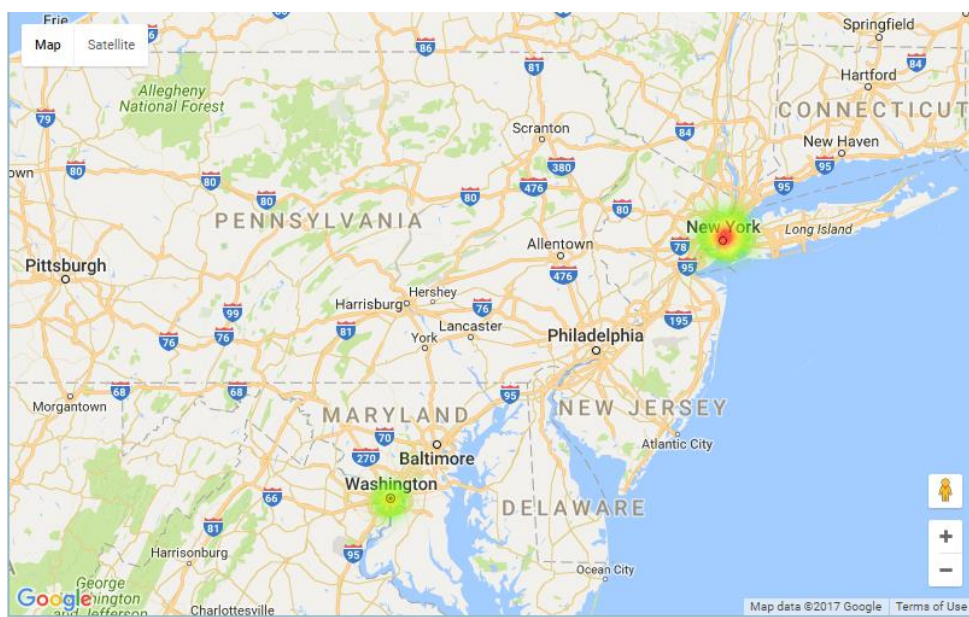
6.1.2 Οπτικός χάρτης εστίασης (heatmap)

Στη συνέχεια παρουσιάζονται όλα τα tweets που συλλέχθηκαν με η μορφή οπτικού χάρτη εστίασης:



Εικόνα 38: Οπτικός χάρτης εστίασης απεικόνισης του συνόλου των tweets που συλλέχθηκαν

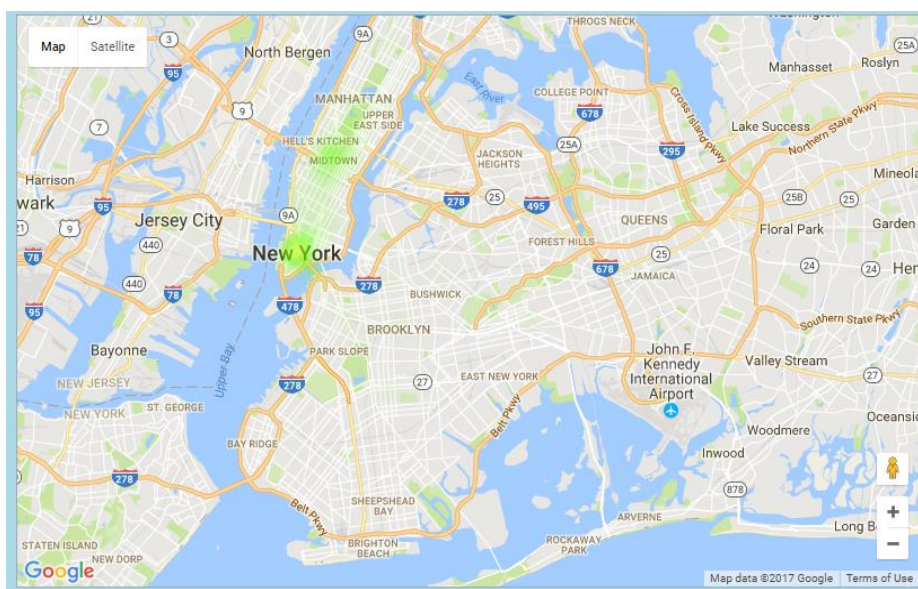
Στον παραπάνω χάρτη οι περιοχές με τον μεγαλύτερο αριθμό tweets απεικονίζονται με σκούρο κόκκινο χρώμα και καθώς ο σχετικός αριθμός γίνεται μικρότερος, το χρώμα των περιοχών σταδιακά γίνεται κίτρινο και πράσινο. Εστιάζοντας στον παραπάνω χάρτη στις ανατολικές ακτές των Ηνωμένων Πολιτειών, λαμβάνεται η παρακάτω απεικόνιση:



Εικόνα 39: Οπτικός χάρτης εστίασης απεικόνισης του συνόλου των tweets που συλλέχθηκαν (ανατολικές ακτές ΗΠΑ)

Οπτική απεικόνιση διάδοσης και περιεχομένου αναρτήσεων στο Twitter

Κατόπιν, εστιάζοντας στην περιοχή Μανχάταν της Νέας Υόρκης λαμβάνεται η εξής οπτική απεικόνιση εστίασης:



Εικόνα 40: Οπτικός χάρτης εστίασης απεικόνισης του συνόλου των tweets που συλλέχτηκαν (περιοχή Μανχάταν – Νέα Υόρκη)

6.2 Απεικόνιση διάδοσης hashtags

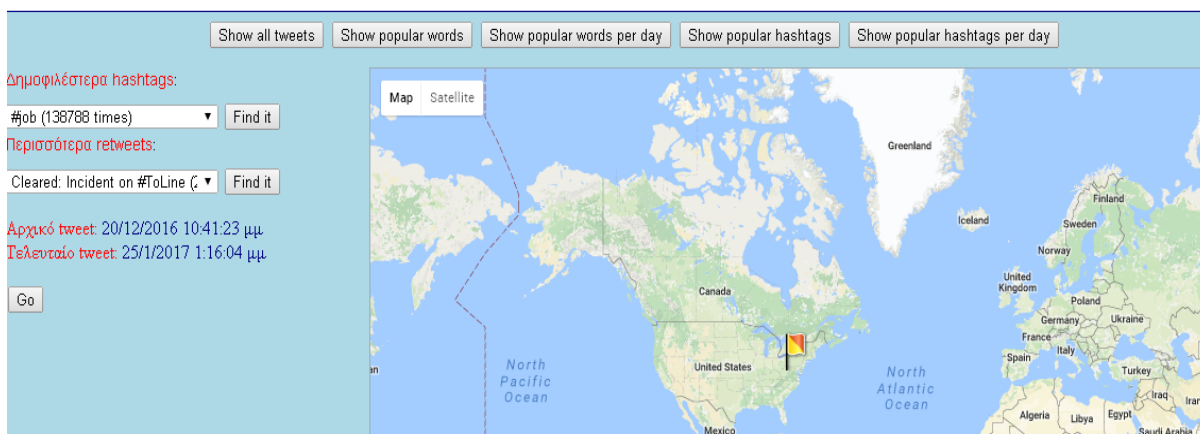
Στην ιστοσελίδα που υλοποιήθηκε στο πλαίσιο της παρούσας μελέτης στο πάνω αριστερό τμήμα περιλαμβάνεται λίστα με τα δέκα (10) hashtags με τον μεγαλύτερο αριθμό tweets που περιλαμβάνουν τα αντίστοιχα hashtags:



Εικόνα 41: Λίστα με τα δέκα (10) δημοφιλέστερα hashtags

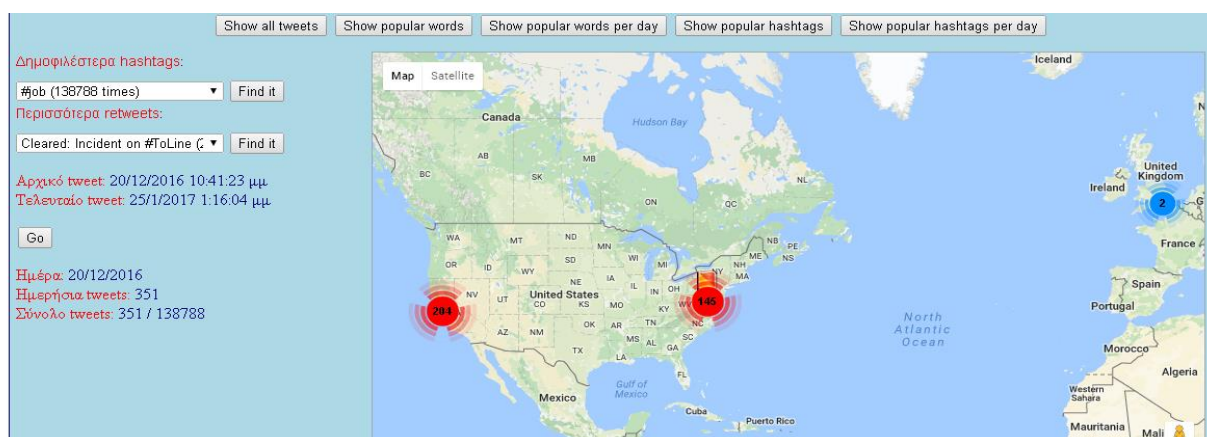
Επιλέγοντας κάποιο από τα δημοφιλέστερα hashtags για το σύνολο των tweets που συλλέχτηκαν και πατώντας το κουμπί “Find it”, εμφανίζονται πληροφορίες σχετικά με το χρόνο της πρώτης και τελευταίας ανάρτησης tweets που περιλαμβάνει το αντίστοιχο hashtag. Επίσης, υπάρχει επισήμανση στον χάρτη Google Map της τοποθεσίας της πρώτης αντίστοιχης δημοσίευσης:

Οπτική απεικόνιση διάδοσης και περιεχομένου αναρτήσεων στο Twitter



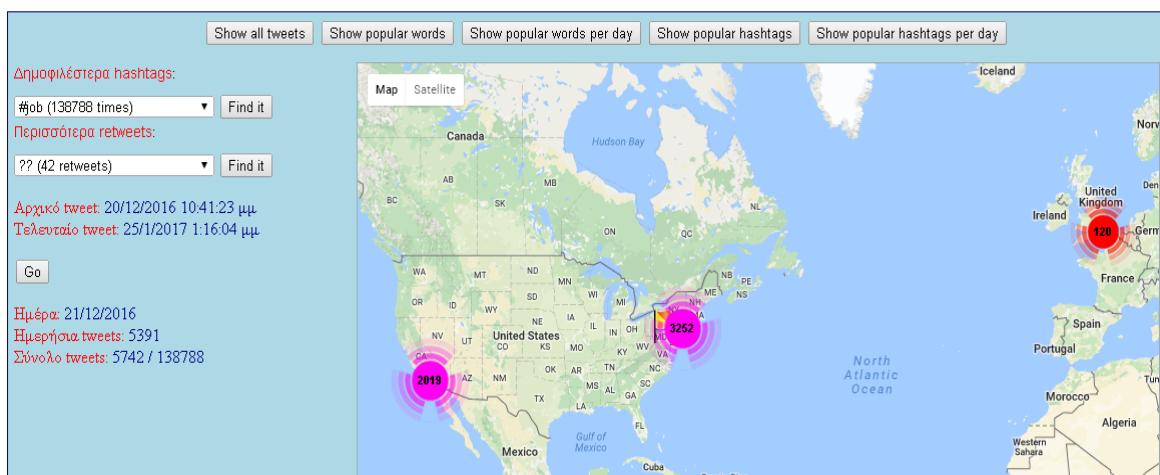
Εικόνα 42: Επισήμανση στον χάρτη του πρώτου χρονολογικά tweet που περιλαμβάνει συγκεκριμένο hashtag

Στη συνέχεια, αν ο χρήστης πατήσει το κουμπί “Go” , εμφανίζονται αναλυτικά στον χάρτη οι περιοχές δημοσίευσης των αντίστοιχων tweet για κάθε ημέρα ξεχωριστά. Επιπροσθέτως, σε κάθε ημέρα εμφανίζεται η πληροφορία του αριθμού των αντίστοιχων tweets, καθώς επίσης και του συνολικού αριθμού δημοσίευσης σχετικών tweet μέχρι την τρέχουσα κάθε φορά ημερομηνία. Η ιστοσελίδα ασύγχρονα φορτώνει τις αντίστοιχες πληροφορίες με χρήση της γλώσσας σεναρίων jQuery και τις απεικονίζει οπτικά στον χάρτη Google Map. Μετά το πέρας πέντε (5) δευτερολέπτων, επαναλαμβάνεται η ασύγχρονη κλήση μέσω jQuery για τη λήψη των αντίστοιχων πληροφοριών για την επόμενη χρονολογικά ημέρα:



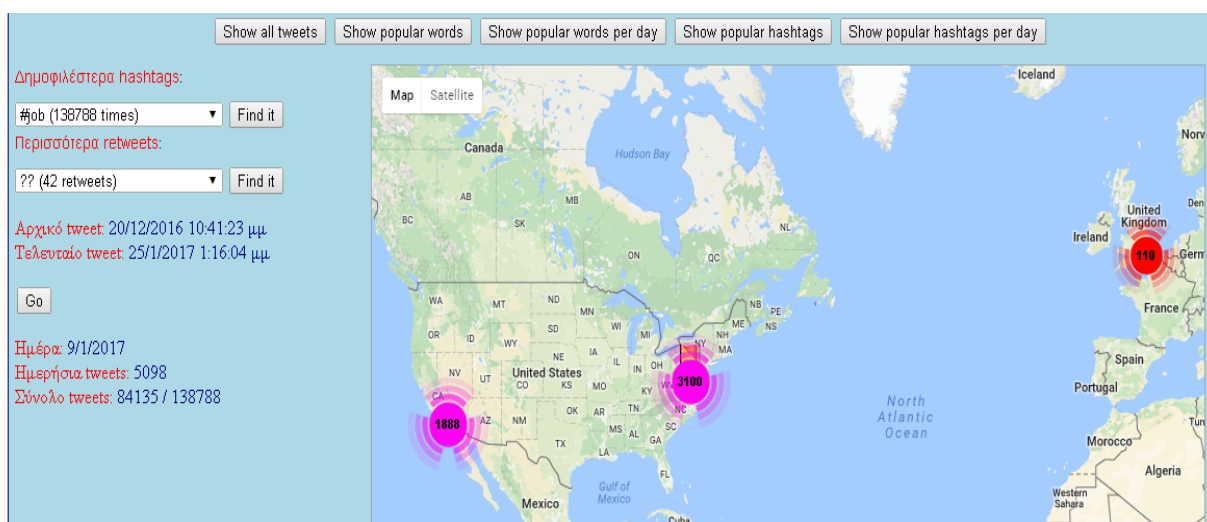
Εικόνα 43: Οπτική απεικόνιση tweets που περιλαμβάνουν το hashtag #job στις 20/12/2016

Κατά την εναλλαγή οπτικής προβολής των tweets για την επόμενη χρονολογικά ημέρα ενημερώνονται οι τιμές των αντίστοιχων ετικετών αριστερά περί της ημέρας οπτικής προβολής των αντίστοιχων tweets, το σύνολο των σχετικών ημερήσιων tweets και ο συνολικός αριθμός αντίστοιχων tweets που μέχρι την τρέχουσα ημερομηνία έχουν απεικονιστεί οπτικά στον χάρτη Google Map:



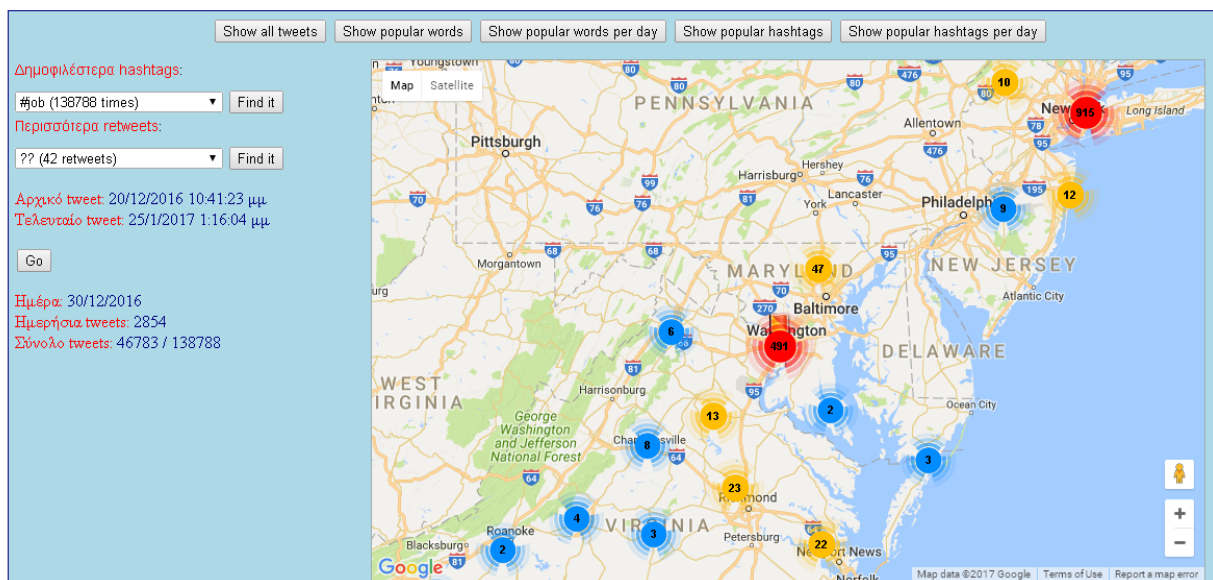
Εικόνα 44: Οπτική απεικόνιση tweets που περιλαμβάνουν το hashtag #job στις 21/12/2016

Από την παραπάνω εικόνα διαπιστώνεται πως την επόμενη ημέρα τα αντίστοιχα tweets υπέρ πολλαπλασιάστηκαν και για τις τρεις περιοχές απεικόνισης (δυτικές ακτές ΗΠΑ, ανατολικές ακτές ΗΠΑ και ευρύτερη περιοχή Λονδίνου). Ο αριθμός δημοσιευμάτων tweets διατηρούν ανάλογη συχνότητα εμφάνισης για τις περισσότερες ημέρες την περίοδο συλλογής των tweets:



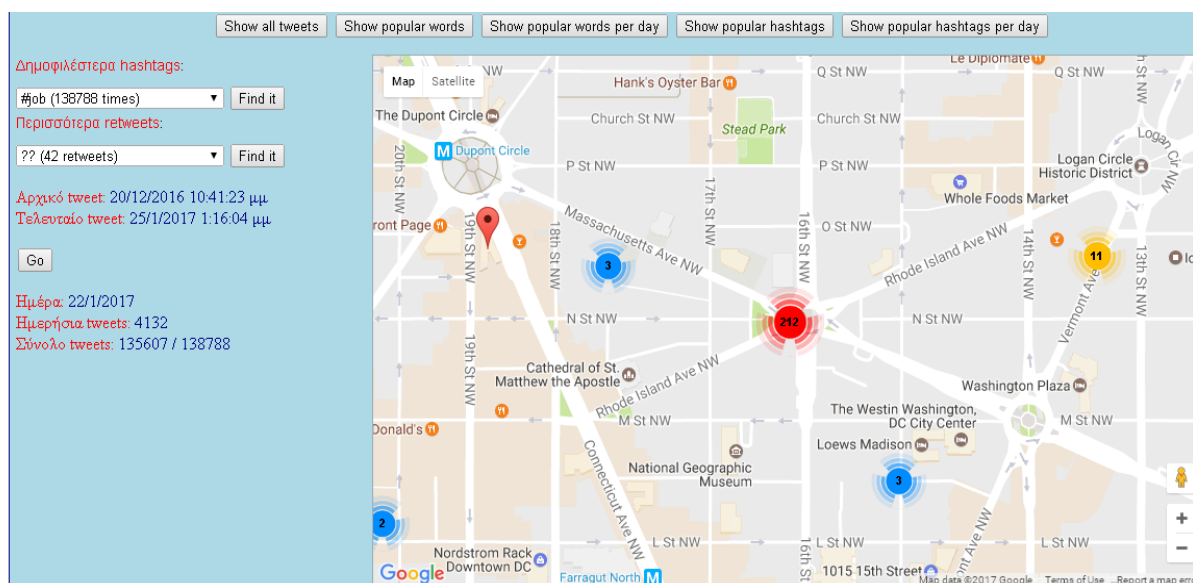
Εικόνα 45: Οπτική απεικόνιση tweets που περιλαμβάνουν το hashtag #job στις 09/01/2017

Κάθε στιγμή επιλέγοντας μεγέθυνση στον χάρτη Google Map είναι δυνατή η λεπτομερέστερη γεωγραφική απεικόνιση των αντίστοιχων tweets. Στην παρακάτω εικόνα απεικονίζονται τα tweets σε μεγέθυνση αποκλειστικά για τις ανατολικές ακτές των ΗΠΑ:



Εικόνα 46: Οπτική απεικόνιση tweets που περιλαμβάνουν το hashtag #job στις 20/12/2016 στις ανατολικές ακτές των ΗΠΑ

Η μεγέθυνση του χάρτη Google Map και η απεικόνιση των σχετικών tweets μπορεί να φτάσει μέχρι και επίπεδο δρόμου με πολύ μεγάλη κλίμακα απεικόνισης:



Εικόνα 47: Οπτική απεικόνιση tweets που περιλαμβάνουν το hashtag #job στις 22/01/2017 στην Ουάσινγκτον

Ο μεγάλος αριθμός αναρτήσεων στο Twitter που περιλαμβάνουν το hashtag #job στη περιοχή απεικόνισης στον παραπάνω χάρτη, επηρεάζεται από το John Hopkins Business School που βρίσκεται στην ευρύτερη περιοχή απεικόνισης του παραπάνω χάρτη, καθώς επίσης και των πρεσβειών της Αυστραλίας και της Ιαπωνίας σε κοντινή απόσταση.

6.3 Απεικόνιση διάδοση retweets

Όπως αναφέρθηκε παραπάνω δεν κατέστη δυνατόν η λήψη της πληροφορίας ποια τυχόν tweets αποτελούν retweet άλλων tweet και συγκεκριμένα ποιων tweets. Ως λύση στο συγκεκριμένο

Οπτική απεικόνιση διάδοσης και περιεχομένου αναρτήσεων στο Twitter

Θέμα ελέγχτηκε το κείμενο των αναρτήσεων στο Twitter και όσα εξ αυτών περιλαμβάνουν το ίδιο ακριβώς κείμενο επισημάνθηκαν στη βάση δεδομένων ως αναρτήσεις retweet. Στο σημείο αυτό αξίζει να αναφερθεί πως τα περισσότερα από τα tweets που επισημάνθηκαν ως retweets αποτελούσαν retweet από τον ίδιο χρήστη που είχε κάνει την αρχική ανάρτηση και για τον λόγο αυτόν αγνοήθηκαν κατά την οπτική απεικόνιση των retweets.

Στην ιστοσελίδα που δημιουργήθηκε, κάτω από τη λίστα με τα δημοφιλέστερα hashtags εμφανίζεται λίστα με τα tweets με τον μεγαλύτερο αριθμό retweets από άλλους χρήστες:



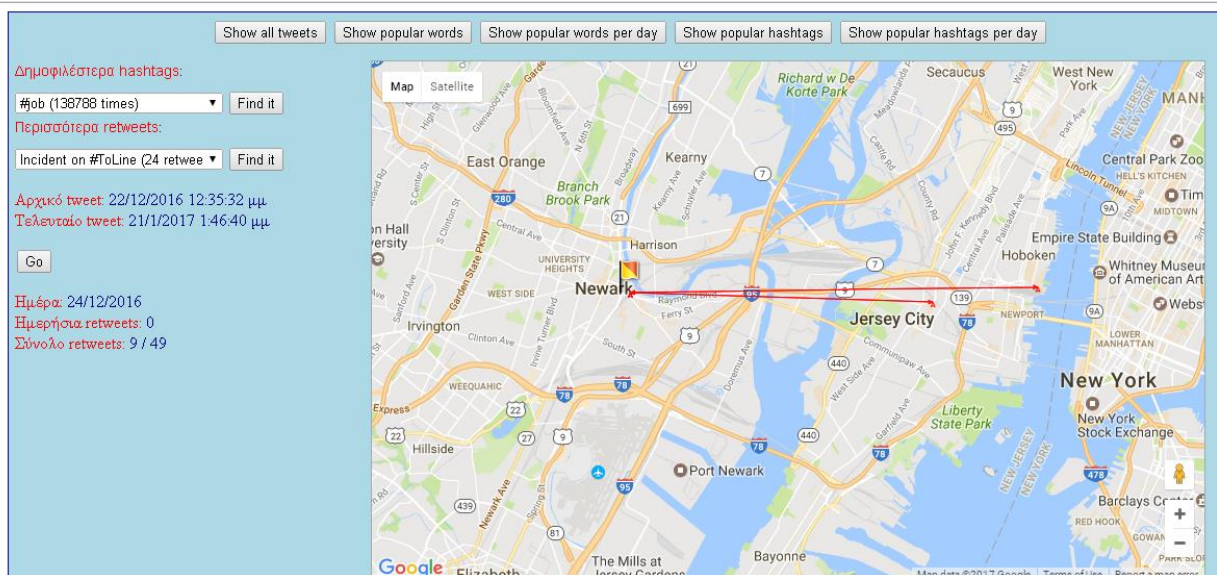
Εικόνα 48: Τα δέκα (10) tweets με τα περισσότερα retweets

Κατόπιν επιλέγοντας κάποιο από τα σχετικά tweets και πατώντας το κουμπί "Find it", επισημαίνεται στον χάρτη το γεωγραφικό σημείο που δημοσιεύτηκε για πρώτη φορά το αντίστοιχο tweet, και εμφανίζεται ο χρόνος της πρώτης και τελευταίας δημοσίευσης του αντίστοιχου tweet:



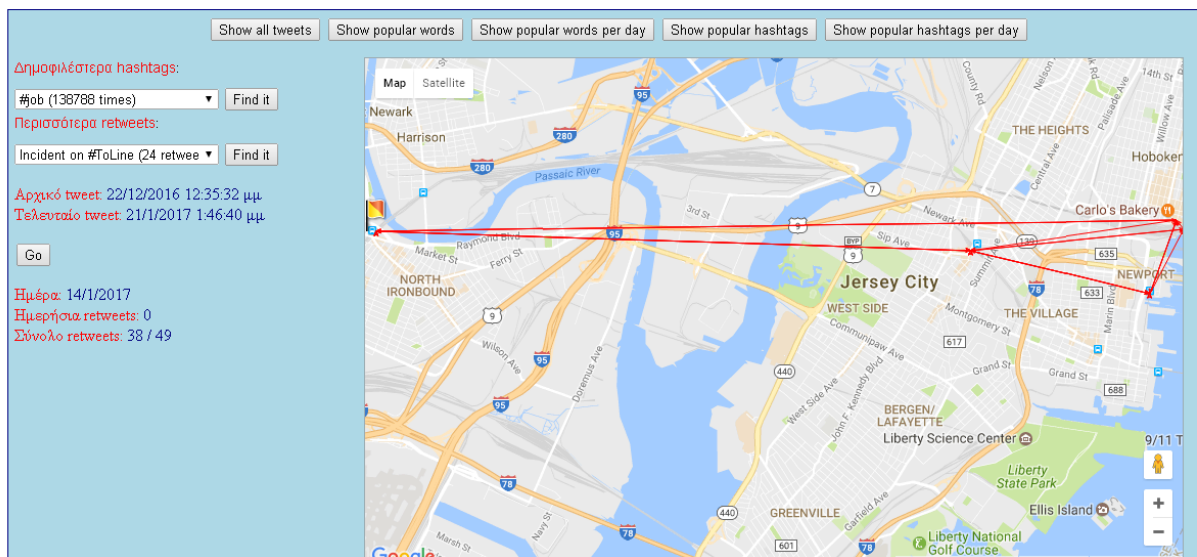
Εικόνα 49: Επισημάνση στον χάρτη της πρώτης δημοσίευσης ενός tweet

Κατόπιν, πατώντας το κουμπί "Go", επισημαίνονται με βέλη στον χάρτη η διασπορά των retweets με αύξουσα χρονολογική σειρά. Κάθε retweet συνδέεται με βέλος έχοντας ως αρχή του βέλους το γεωγραφικό σημείο του tweet από το οποίο έγινε το retweet και ως τέλος το γεωγραφικό σημείο που πραγματοποιήθηκε το retweet:



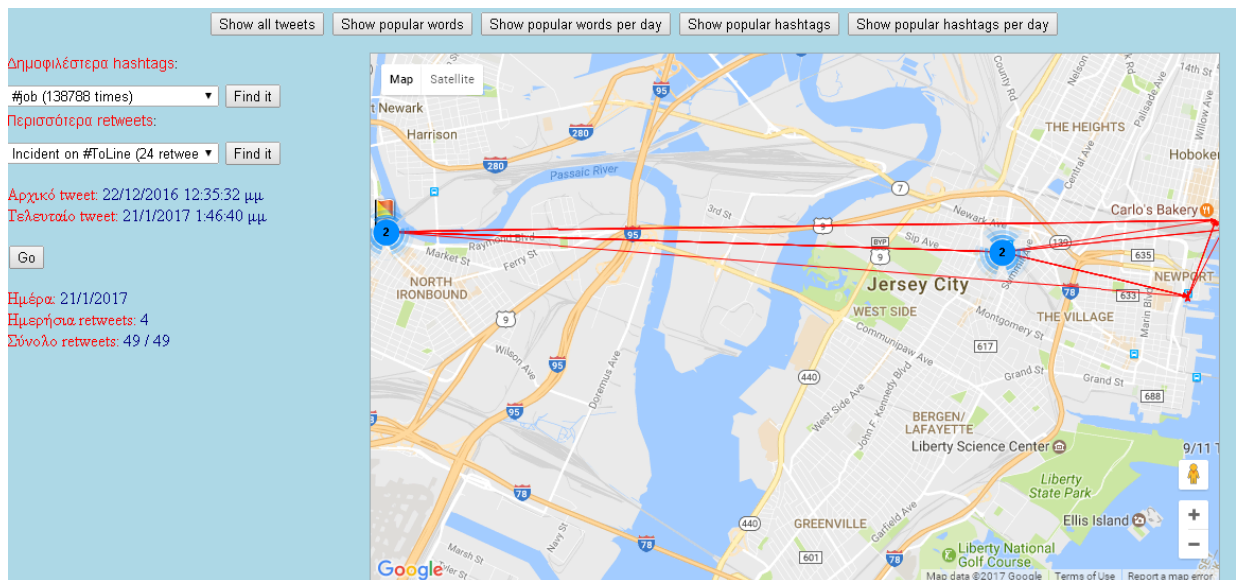
Εικόνα 50: Οπτική απεικόνιση όλων των retweet στον χάρτη (σημείο προέλευσης, σημείο δημοσίευσης) μέχρι τις 24/12/2016 (Θέμα tweet: Incident on #ToLine)

Από την παραπάνω εικόνα διαπιστώνεται η αρχική διάδοση ενός tweet στα όρια της Νέας Υόρκης. Ακολουθούμενα από επιπλέον retweet με την πάροδο των ημερών:



Εικόνα 51: Οπτική απεικόνιση όλων των retweet στον χάρτη (σημείο προέλευσης, σημείο δημοσίευσης) μέχρι τις 14/01/2017 (Θέμα tweet: Incident on #ToLine)

Η παρακάτω εικόνα απεικονίζει τον τελικό γράφο διάδοσης του tweet με κείμενο Incident on #ToLine για το σύνολο των tweets που συλλέχτηκαν στο πλαίσιο της παρούσας έρευνας:



Εικόνα 52: Οπτική απεικόνιση όλων των retweet στον χάρτη (σημείο προέλευσης, σημείο δημοσίευσης) μέχρι τις 21/01/2017 (Θέμα tweet: Incident on #ToLine)

Οι συγκεκριμένες αναρτήσεις στο Twitter συνδέονται με περιστατικό που έλαβε τόπο στο σταθμό του μετρό Newark Penn Station της Νέας Υόρκης στις 22 Δεκεμβρίου 2016, κατά το οποίο κάποιος προσπάθησε να κλέψει ένα σπαθί σαμουράι από επιβάτη του μετρό εντός του σταθμού Newark Penn Station, γεγονός που είχε ως αποτέλεσμα την καταδίωξη και σύλληψη του δράστη, αλλά και του κατόχου του σπαθιού εντός του αντίστοιχου σταθμού μετρό. Η πρώτη ανάρτηση του tweet που στη συνέχεια έγινε retweet από άλλους χρήστες, πραγματοποιήθηκε εντός του σταθμού μετρό Newark Penn Station:



Εικόνα 53: Είδηση προσπάθειας κλοπής σπαθιού σαμουράι εντός σταθμού μετρό της Νέας Υόρκης στις 22 Δεκέμβρη 2016 (CBS, 2016)

Με ανάλογο τρόπο είναι δυνατή η οπτική απεικόνιση της σχετικής πληροφορίας για κάθε tweet που ακολουθείτε από retweets για τα tweets που συλλέχτηκαν στο πλαίσιο της παρούσας μελέτης.

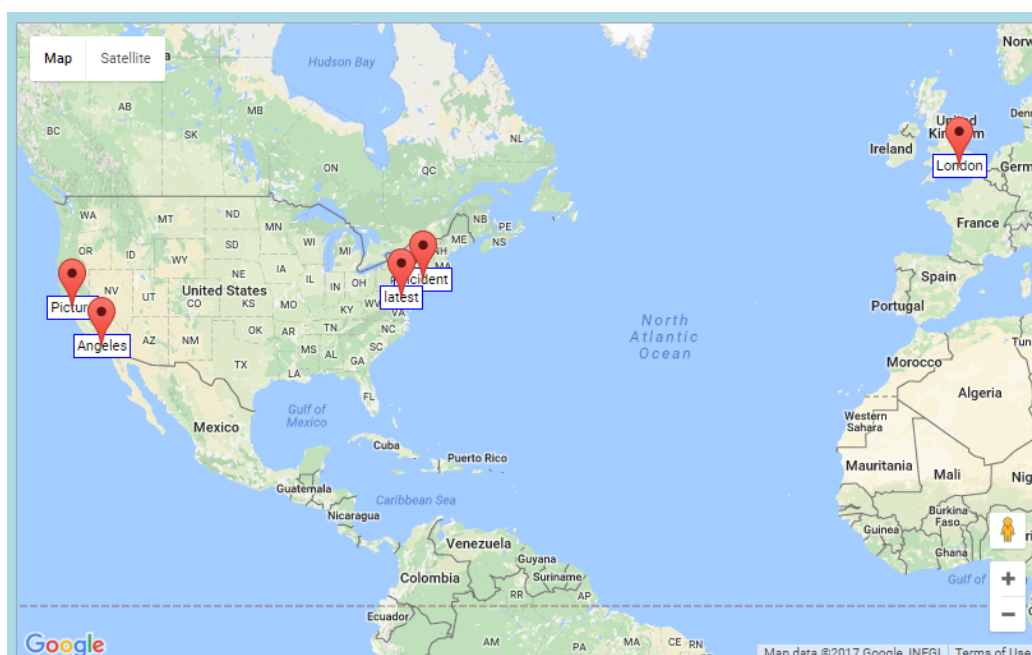
Οπτική απεικόνιση διάδοσης και περιεχομένου αναρτήσεων στο Twitter

6.4 Απεικόνιση δημοφιλών λέξεων κλειδιών στον χάρτη

Τα tweets που αποτέλεσαν τη βάση της παρούσας ανάλυσης προέρχονται από τις παρακάτω πέντε (5) πόλεις:

1. Νέα Υόρκη
2. Λος Άντζελες
3. Σαν Φρανσίσκο
4. Ουάσινγκτον
5. Λονδίνο

Στην παρακάτω εικόνα απεικονίζεται σε χάρτη η δημοφιλέστερη λέξη για κάθε μια από τις παραπάνω πόλεις για το σύνολο των tweets που συλλέχτηκαν:



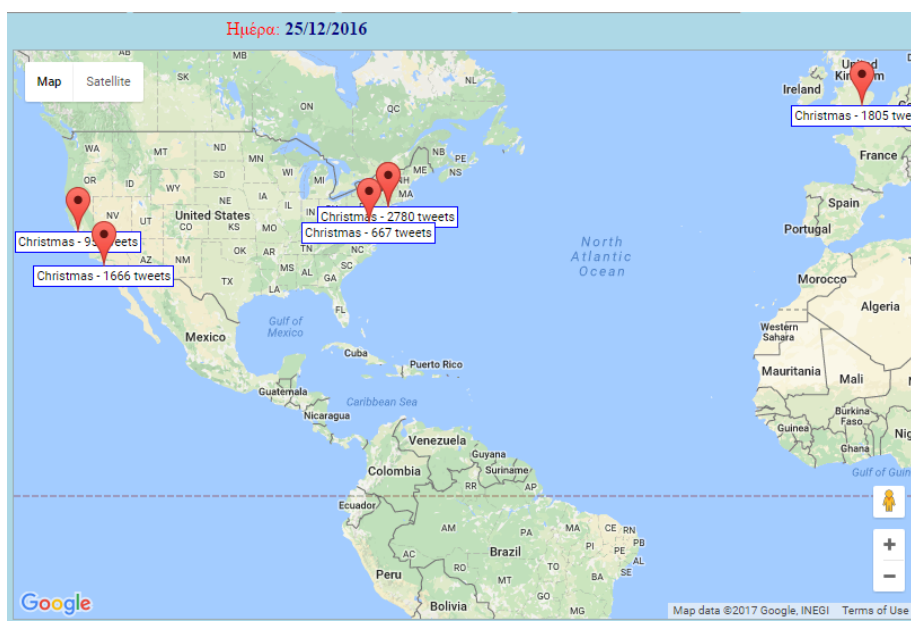
Εικόνα 54: Οπτική απεικόνιση δημοφιλέστερων λέξεων ανά πόλη για το σύνολο των tweets που συλλέχτηκαν

Όπως διαπιστώνεται από την παραπάνω εικόνα για την πόλη της Νέας Υόρκη η λέξη Incident εμφανίζεται σε μεγαλύτερο αριθμό tweets, για την Ουάσινγκτον η λέξη latest, για το Λος Άντζελες η λέξη Angeles, για το Σαν Φρανσίσκο η λέξη Picture και για το Λονδίνο η λέξη London. Παρατηρώντας τις δημοφιλέστερες λέξεις που εμφανίζονται στο σύνολο των tweets που συλλέχτηκαν στο πλαίσιο της παρούσας μελέτης, διαπιστώνεται πως εκείνες περιλαμβάνουν είτε το “όνομα της πόλης” που πραγματοποιήθηκε η ανάρτηση, είτε τη λέξη “συμβάν”, είτε τη λέξη “πρόσφατη – τελευταία”, είτε τη λέξη “εικόνα”. Οι σχετικές λέξεις υποδηλώνουν πως οι χρήστες χρησιμοποιούν το Twitter για την ανάρτηση πρόσφατων συμβάντων – γεγονότων τη στιγμή που εκείνα λαμβάνουν τόπο και αφορούν σε γεγονότα που λαμβάνουν τόπο στην

περιοχή που βρίσκονται κάθε στιγμή. Επίσης, σε μεγάλο ποσοστό οι σχετικές αναρτήσεις περιλαμβάνουν εικόνες των αντίστοιχων συμβάντων.

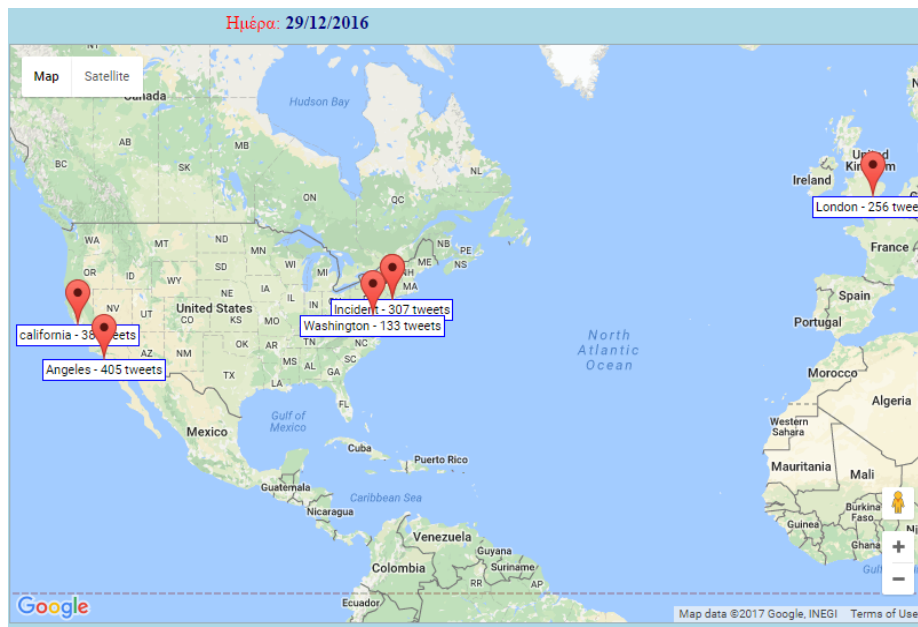
Κατόπιν αναπροσαρμόζεται ο κώδικας της εφαρμογής (ιστότοπος) που υλοποιήθηκε με σκοπό την οπτική απεικόνιση των tweets, έτσι ώστε να προβάλλονται αυτόματα για κάθε ημέρα ξεχωριστά οι δημοφιλέστερες λέξεις, ξεκινώντας από την ημέρα που δημοσιεύτηκε το πρώτο tweet, μέχρι την ημέρα που δημοσιεύτηκε το τελευταίο tweet που συλλέχτηκε με τη χρήση του Twitter API.

Για τις περισσότερες ημέρες, η δημοφιλέστερη λέξη για όλες τις πόλεις είναι η ίδια με τη δημοφιλέστερη λέξη για το σύνολο όλων των ημερών. Ωστόσο, σε διάφορες ημέρες, όπως τα Χριστούγεννα σε όλες τις πόλεις δημοφιλέστερη λέξη είναι η Christmas:



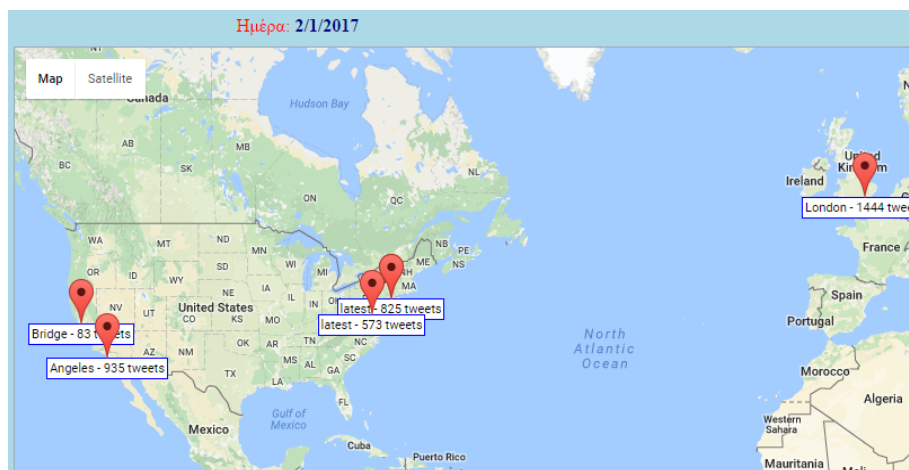
Εικόνα 55: Οπτική απεικόνιση δημοφιλέστερων λέξεων ανά πόλη για το σύνολο των tweets στις 25/12/2016

Ομοίως στις 29/12/2016 δημοφιλέστερη λέξη για το Σαν Φρανσίσκο είναι η California και για την Ουάσινγκτον η λέξη Washington:



Εικόνα 56: Οπτική απεικόνιση δημοφιλέστερων λέξεων ανά πόλη για το σύνολο των tweets στις 29/12/2016

Αντίστοιχα, για τις 02/02/2017 η δημοφιλέστερη λέξη στο Σαν Φρανσίσκο είναι Bridge, ενώ για τις υπόλοιπες πόλεις διατηρείται η συνολικά δημοφιλέστερη λέξη:



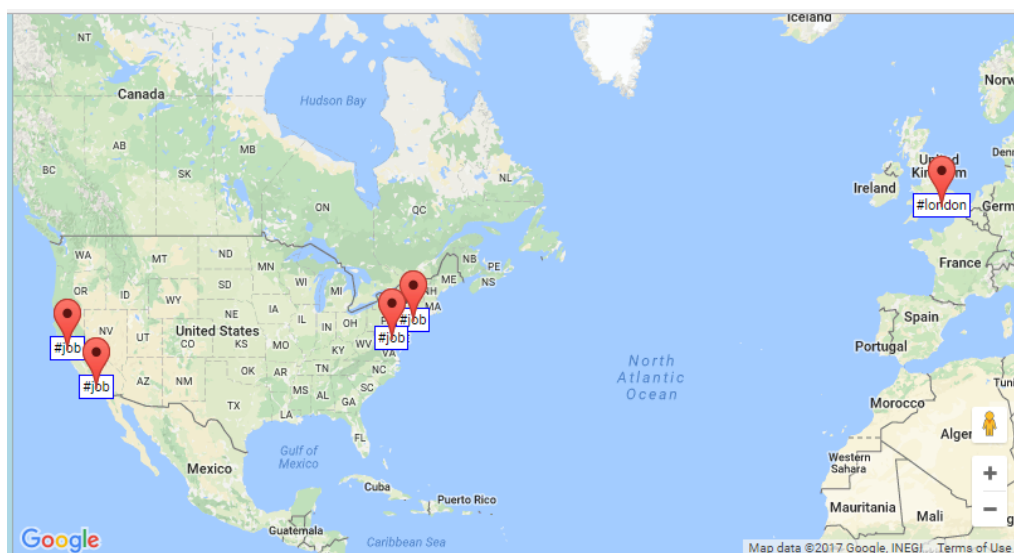
Εικόνα 57: Οπτική απεικόνιση δημοφιλέστερων λέξεων ανά πόλη για το σύνολο των tweets στις 02/01/2017

6.5 Απεικόνιση δημοφιλών hashtag στον χάρτη

Όπως αναφέρθηκε παραπάνω τα tweets κατά κύριο λόγο προέρχονται από τις εξής πέντε (5) πόλεις:

1. Νέα Υόρκη
2. Λος Άντζελες
3. Σαν Φρανσίσκο
4. Ουάσινγκτον
5. Λονδίνο

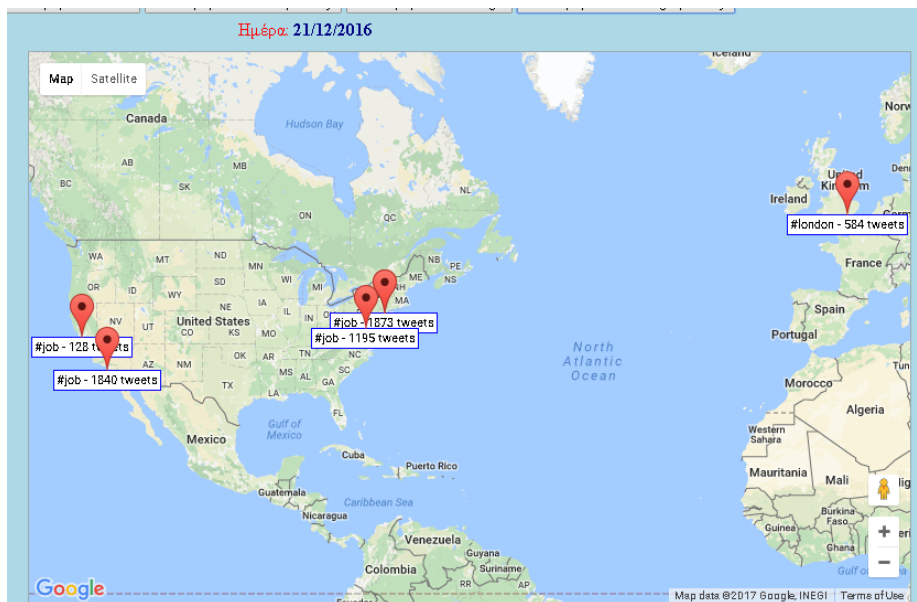
Στην παρακάτω εικόνα απεικονίζεται σε χάρτη το δημοφιλέστερο hashtag για κάθε μια από τις παραπάνω πόλεις για το σύνολο των tweets που συλλέχτηκαν:



Εικόνα 58: Οπτική απεικόνιση δημοφιλέστερων hashtags ανά πόλη για το σύνολο των tweets που συλλέχτηκαν

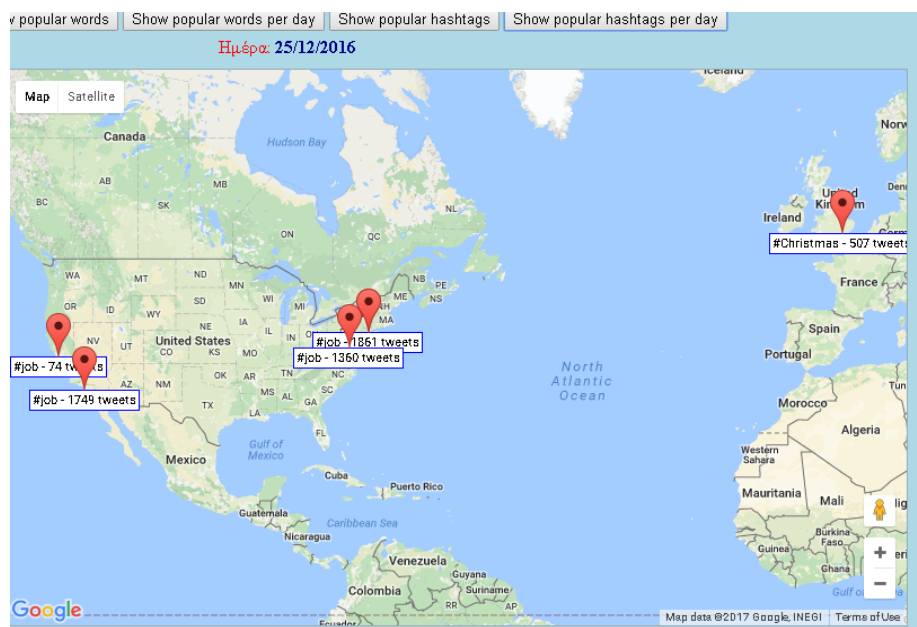
Από τη παραπάνω εικόνα διαπιστώνεται πως σε όλες τις πόλεις με εξαίρεση το Λονδίνο, το hashtag #job εμφανίζεται σε μεγαλύτερο αριθμό στα tweets που συλλέχτηκαν για τις αντίστοιχες γεωγραφικές περιοχές. Για την περιοχή του Λονδίνου το δημοφιλέστερο hashtag είναι το #london.

Κατόπιν αναπροσαρμόζεται ο κώδικας της εφαρμογής (ιστότοπος) που υλοποιήθηκε με σκοπό την οπτική απεικόνιση των tweets, έτσι ώστε να προβάλλει αυτόματα για κάθε ημέρα ξεχωριστά τα δημοφιλέστερα hashtags για κάθε μέρα, ξεκινώντας από την ημέρα που δημοσιεύτηκε το πρώτο tweet, μέχρι την ημέρα που δημοσιεύτηκε το τελευταίο tweets που συλλέχτηκε με τη χρήση του Twitter API. Για τις περισσότερες ημέρες τα δημοφιλέστερα hashtag είναι τα ίδια για όλες τις πόλεις. Ακολουθεί παρουσίαση των δημοφιλέστερων hashtag για διάφορες ημερομηνίες:



Εικόνα 59: Οπτική απεικόνιση δημοφιλέστερων hashtags ανά πόλη στις 21/12/2016

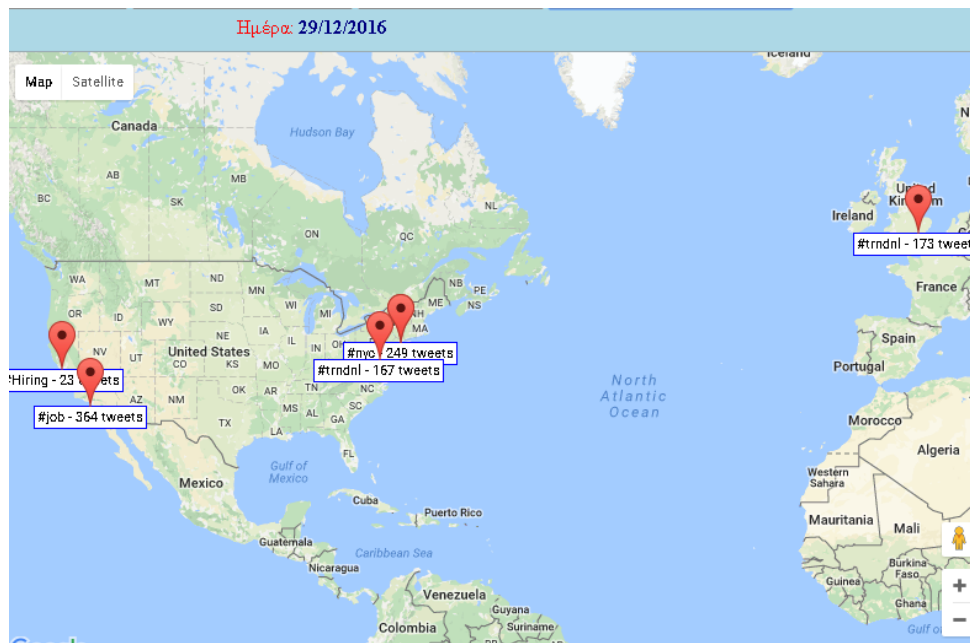
Ωστόσο, όπως διαπιστώνεται στην παρακάτω εικόνα την ημέρα των Χριστουγέννων, δημοφιλέστερο hashtag για το Λονδίνο είναι το #christmas, σε συμφωνία με την εορταστική περίοδο των Χριστουγέννων:



Εικόνα 60: Οπτική απεικόνιση δημοφιλέστερων hashtags ανά πόλη στις 25/12/2016

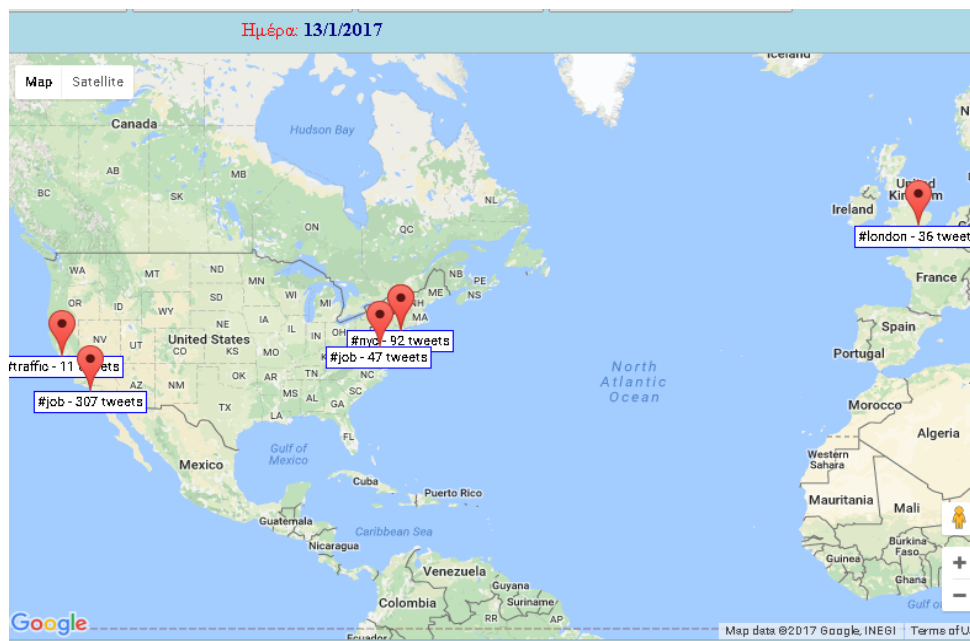
Επίσης, για τις 29/12/2017 διαπιστώνεται πως στο Λονδίνο και την Ουάσινγκτον δημοφιλέστερο hashtag είναι το #trndnl. Το σχετικό hashtag αντιστοιχεί στον όρο Trendinalia και η αντίστοιχη σελίδα περιλαμβάνει πληροφορίες – αναρτήσεις σχετικά με δημοφιλή θέματα – hashtags σε διάφορα μέρη του πλανήτη:

Οπτική απεικόνιση διάδοσης και περιεχομένου αναρτήσεων στο Twitter



Εικόνα 61: Οπτική απεικόνιση δημοφιλέστερων hashtags ανά πόλη στις 29/12/2016

Επίσης, για τις 13/01/2017 διαπιστώνεται ως δημοφιλέστερο hashtag στο Σαν Φρανσίσκο το hashtag #traffic:



Εικόνα 62: Οπτική απεικόνιση δημοφιλέστερων hashtags ανά πόλη στις 13/01/2017


Κατόπιν σχετικής αναζήτησης στο διαδίκτυο, διαπιστώθηκε πως στο Σαν Φρανσίσκο σημειώθηκε σημαντικό ατύχημα σε κεντρικό δρόμο της πόλης, προκαλώντας αυξημένη κίνηση εκείνη την ημέρα, δικαιολογώντας την κυριαρχία του αντίστοιχου hashtag εκείνη την μέρα για τα tweets που δημοσιεύτηκαν από το Σαν Φρανσίσκο στις 13/01/2017 (News, 2017):

abc 7 SECTIONS TRAFFIC WATCH

TRAFFIC

RV fire on Highway 101 causes traffic delays near San Francisco airport

Share G+ Tweet Email



TRAFFIC

- Caltrain asking public to vote on train car redesign
- Tourist killed, 22 hurt after driver plows into crowd in Times Square
- PHOTOS: 1 dead, dozens injured after car hit pedestrians in New York's Times Square
- SF mayor pledges to spend \$90M to re-pave torn up streets
- Friends remember Alameda Co. deputy killed in crash

MORE TRAFFIC ▸

Εικόνα 63: Είδηση τροχαίου στο Σαν Φρανσίσκο στις 13 Ιανουαρίου 2017 (News, 2017)

ΚΕΦ.7: Συμπεράσματα

7.1 Συμπεράσματα έρευνας

Στο πλαίσιο της παρούσας έρευνας συλλέχθηκαν 1 εκ. tweets, τα οποία αποτέλεσαν τη βάση για την ανάπτυξη εφαρμογής ανάλυσης και οπτικής απεικόνισης (visualization) πληροφοριών που περιλαμβάνουν τα αντίστοιχα tweets. Η εφαρμογή που αναπτύχθηκε συνδυάζει δυνατότητες που περιλαμβάνουν άλλες σχετικές εφαρμογές, όπως (i) δυνατότητα απεικόνισης δημοφιλών λέξεων και hashtags που περιλαμβάνει η εφαρμογή TrendsMap, (ii) δυνατότητα απεικόνισης στο χάρτη του συνόλου των διαθέσιμων tweets που περιλαμβάνει η εφαρμογή OneMillionTweetMap και (iii) δυνατότητα απεικόνισης της ροής εξέλιξης διάδοσης δημοφιλών hashtags που περιλαμβάνει η εφαρμογή TweetMap.

Οι δυνατότητες της εφαρμογής που αναπτύχθηκε συνοψίζονται ως εξής:

- χωροταξική απεικόνιση του συνόλου των tweets με τη μορφή συστοιχιών σήμανσης (marker clusters) χωροταξική απεικόνιση του συνόλου των tweets με τη μορφή οπτικού χάρτη εστίασης (heat map)
- χωροχρονική απεικόνιση διάδοσης δημοφιλών hashtags, για κάθε ημέρα ξεχωριστά
- χωροχρονική απεικόνιση διάδοσης δημοφιλών retweets, για κάθε ημέρα ξεχωριστά
- χωροταξική απεικόνιση δημοφιλέστερων hashtags (για πέντε συνολικά πόλεις στις Ηνωμένες Πολιτείες και το Ηνωμένο Βασίλειο), για το σύνολο των tweets
- χωροταξική απεικόνιση δημοφιλέστερων λέξεων (για πέντε συνολικά πόλεις στις Ηνωμένες Πολιτείες και το Ηνωμένο Βασίλειο), για το σύνολο των tweets
- χωροχρονική απεικόνιση δημοφιλέστερων hashtags (για πέντε συνολικά πόλεις στις Ηνωμένες Πολιτείες και το Ηνωμένο Βασίλειο), για κάθε ημέρα ξεχωριστά
- χωροχρονική απεικόνιση δημοφιλέστερων λέξεων (για πέντε συνολικά πόλεις στις Ηνωμένες Πολιτείες και το Ηνωμένο Βασίλειο), για κάθε ημέρα ξεχωριστά

Η εφαρμογή που αναπτύχθηκε ικανοποιεί τους σκοπούς της παρούσας μελέτης σχετικά με την (i) οπτική χωροχρονική απεικόνιση πληροφοριών σχετικά με αναρτήσεις μηνυμάτων – tweets στην πλατφόρμα Twitter, καθώς επίσης και (ii) την οπτική χωροχρονική απεικόνιση της ροής εξέλιξης ή διάδοσης ή εναλλαγής των σχετικών πληροφοριών.

Όπως διαφάνηκε από τη παρουσίαση των δημοφιλέστερων hashtags ανά ημέρα για πόλεις των Ηνωμένων Πολιτειών και του Λονδίνου, η χωροχρονική απεικόνιση των σχετικών πληροφοριών δύναται να διευκολύνει (i) τον άμεσο εντοπισμό σημαντικών τοπικών γεγονότων (κυριαρχία του hashtag #traffic στο Σαν Φρανσίσκο για ημέρα με σημαντικό αυτοκινητιστικό ατύχημα στην πόλη) και (ii) την τάση - ενδιαφέροντα της πλειοψηφίας του πληθυσμού (κυριαρχία του hashtag #survivor στην Αττική κατά το χρόνο προβολής του αντίστοιχου προγράμματος στη τηλεόραση – χρήση εφαρμογής TrendsMaps) σε συγκεκριμένη χρονική στιγμή.

Η απεικόνιση των σχετικών πληροφοριών σε χάρτη επιτρέπει την άμεση διαπίστωση γεγονότων σημαντικού ενδιαφέροντος, καθώς επίσης και της διαφοροποίησης των τάσεων μεταξύ διαφόρων γεωγραφικών περιοχών. Η χρήση χαρτών διευκολύνει την άμεση πρόσβαση - κατανόηση πληροφοριών που φέρουν τα διάφορα tweets που αναρτώνται στο Twitter από το κοινό, χωρίς την ανάγκη ειδικών αναλύσεων – μελετών για την πρόσβαση στην αντίστοιχη πληροφορία.

Συνοψίζοντας η απεικόνιση των tweets και των πληροφοριών που φέρουν τα αντίστοιχα tweets σε διεπαφή χάρτη προσφέρουν:

- Ευκολία πρόσβασης στις σχετικές πληροφορίες
- Δυνατότητα άμεσης πρόσβασης στις σχετικές πληροφορίες
- Δυνατότητα καλύτερης κατανόησης των σχετικών πληροφοριών
- Δυνατότητα εντοπισμού ειδικών πληροφοριών που ενδιαφέρουν το χρήστη, με χρήση των δυνατοτήτων – φίλτρων που προσφέρουν οι εφαρμογές οπτικής απεικόνισης πληροφοριών σε διεπαφή χάρτη
- Δυνατότητα άμεσης διαπίστωσης τάσεων – συμπεριφορών των χρηστών σε διάφορες γεωγραφικές περιοχές

Επίσης, η εφαρμογή που αναπτύχθηκε επιτρέπει εν μέρει την αποτύπωση της συχνότητας χρήσης της εφαρμογής Twitter με ενεργοποιημένη την καταγραφή της τοποθεσίας χρήστη, για τις περιοχές που συλλέχτηκαν τα αντίστοιχα tweets (δείγμα έρευνας) και την απεικόνιση της σχετικής πληροφορίας σε διεπαφή χάρτη, σε αντιστοιχία με σχετική έρευνα των Budak et al. (2014).

Τέλος, η δυνατότητα απεικόνισης σε διεπαφή χάρτη της ροής εξέλιξης – διάδοσης στο Twitter διαφόρων hashtags επιτρέπει την αποτύπωση της ταχύτητας διάδοσης – εξάπλωσης γεγονότων – ειδήσεων με χρήση της πλατφόρμας Twitter. Η αντίστοιχη δυνατότητα που προσφέρει το Twitter διαφαίνεται και από την απεικόνιση διάδοσης των tweets – retweets που περιλαμβάνει η εφαρμογή που αναπτύχθηκε.

7.2 Περιορισμοί έρευνας

Βασικό περιορισμό της έρευνας αποτελεί το γεγονός πως στην ανάλυση που παρουσιάστηκε παραπάνω δεν περιλαμβάνεται το σύνολο των tweets που δημοσιεύτηκαν στις περιοχές των Ηνωμένων Πολιτειών και του Ηνωμένου Βασιλείου που αναφέρθηκαν παραπάνω, αλλά μονάχα τα tweets που δημοσιεύτηκαν από χρήστες με ενεργοποιημένη τη μετάδοση πληροφοριών τοποθεσίας κατά την ανάρτηση μηνυμάτων στο Twitter. Αν όλοι οι χρήστες που πραγματοποίησαν σχετικές αναρτήσεις είχαν ενεργοποιημένη τη σχετική δυνατότητα, ίσως τα αποτελέσματα της ανάλυσης να ήταν διαφορετικά. Το ίδιο ισχύει και στην περίπτωση που ορισμένοι χρήστες που ανάρτησαν μηνύματα στο Twitter με ενεργοποιημένη τη σχετική δυνατότητα, στην πράξη είχαν αποφασίσει την απόκρυψη των σχετικών πληροφοριών. Παρόλα αυτά ο περιορισμός αυτός υπήρξε η πρωταρχική συνθήκη στην κατεύθυνση της οποίας αναπτύχθηκε η μεθοδολογία της ανάλυσης, δηλαδή η διαθεσιμότητα της αντίστοιχης γεωχωρικής πληροφορίας.

7.3 Προτάσεις για μελλοντικές έρευνες

Η εφαρμογή που αναπτύχθηκε θα μπορούσε να βελτιωθεί προσφέροντας στο χρήστη τη δυνατότητα επιλογής εκ μέρους του, των hashtags ή tweets που επιθυμεί να παρακολουθήσει τη ροή εξέλιξης ή διάδοσης και όχι αποκλειστικά των δέκα (10) δημοφιλέστερων hashtags ή λέξεων που περιλαμβάνονται στο κείμενο των tweets που συλλέχτηκαν στο πλαίσιο της παρούσας έρευνας. Ανάλογη δυνατότητα θα ήταν χρήσιμο να προστεθεί σχετικά με την αναζήτηση – προβολή των δημοφιλέστερων hashtags ή λέξεων για κάθε μια εκ των πόλεων των Ηνωμένων Πολιτειών και της Μεγάλης Βρετανίας από τις οποίες προέρχονται τα tweets που συλλέχτηκαν στο πλαίσιο της παρούσας έρευνας. Η παροχή προς το χρήστη εργαλείων επιλογής και περιορισμού των tweets που αναλύονται – προβάλλονται οπτικά από την εφαρμογή, θα ενισχύσει τη λειτουργικότητά της και τα οφέλη που μπορεί να αποκομίσει κάποιος από τη χρήση της.

Σχετικά με ανάλογες μελλοντικές έρευνες, χρήσιμη θα ήταν η υλοποίηση έρευνας με ενεργή διασύνδεση στο Twitter API, με απεικόνιση tweets σε πραγματικό χρόνο και αποτύπωσή τους σε διεπαφή χάρτη σε αντιστοιχία με την εφαρμογή OneMillionTweetMap. Ωστόσο, σκοπός της εν λόγω εφαρμογής δε θα ήταν μονάχα η οπτική απεικόνιση των αντίστοιχων πληροφοριών, αλλά η λειτουργία της ως εφαρμογή παρακολούθησης διάδοσης επιλεγμένων από το χρήστη hashtags (hashtag monitoring). Η αντίστοιχη λειτουργία θα ήταν ιδιαίτερα χρήσιμη για την παρακολούθηση της επιτυχίας εφαρμογής στρατηγικών διαδικτυακού μάρκετινγκ.

Βιβλιογραφία

Ajjan, H., & Hartshorne, R. (2008). Investigating faculty decisions to adopt Web 2.0 technologies: Theory and empirical tests. *The internet and higher education*, 11(2), 71-80.

Asokan, M. (2013). E-commerce websites with jQuery. *International Journal on Computer Science and Engineering*, 5(3), 153-165.

Baik, B., Cao, Q., Choi, S., & Kim, J. M. (2016). Local Twitter Activity and Stock Returns. *Seoul National University* working paper.

BBC News, (2017). Chris Cornell: Soundgarden star hanged himself. Retrieved 2 June 2017, from <http://www.bbc.com/news/entertainment-arts-39960066>

Benevenuto, F., Magno, G., Rodrigues, T., & Almeida, V. (2010, July). Detecting spammers on Twitter. In *Collaboration, electronic messaging, anti-abuse and spam conference (CEAS)* (Vol. 6, p. 12).

Bildirici, I. O., & Ulugtekin, N. N. (2010, November). Web mapping with Google Maps mashups: overlaying geodata. In *A special joint symposium of ISPRS technical commission IV & AutoCarto in conjunction with ASPRS/CaGIS 2010 fall specialty conference, Orlando, FL* (pp. 15-19).

Budak, C., Georgiou, T., Agrawal, D., & El Abbadi, A. (2013). Geoscope: Online detection of geo-correlated information trends in social networks. *Proceedings of the VLDB Endowment*, 7(4), 229-240.

CBS, (2016). Man Caught Stealing Another Man's Sword At Newark Penn Station, Police Say. Retrieved 1 June 2017, from Retrieved 2 June 2017, from Retrieved 2 June 2017, from <http://newyork.cbslocal.com/2016/12/22/penn-station-sword-theft/>

Cossu, J. V., Dugué, N., & Labatut, V. (2015, September). Detecting real-world influence through Twitter. In *Network Intelligence Conference (ENIC), 2015 Second European* (pp. 83-90). IEEE.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1992). Extrinsic and intrinsic motivation to use computers in the workplace¹. *Journal of applied social psychology*, 22(14), 1111-1132.

Deci, E. L. (1975). *Intrinsic motivation*. New York: Plenum Press.

Ebersbach, A., Glaser, M., Heigl, R., & Warta, A. (2008). *Wiki: web collaboration*. Berlin and Heidelberg: Springer-Verlag GmbH & Co. KG.

Ghanem, T. M., Magdy, A., Musleh, M., Ghani, S., & Mokbel, M. F. (2014, November). Viscat: Spatio-temporal visualization and aggregation of categorical attributes in Twitter data. In *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (pp. 537-540). ACM.

Global Web Index, (2015). Top 10 Reasons for Using Social Media. Retrieved 24 January 2017, from <https://www.globalwebindex.net/blog/top-10-reasons-for-using-social-media>

- Goodchild, M. F. (2007). Citizens as voluntary sensors: spatial data infrastructure in the world of Web 2.0. *IJSDIR* 2: 24–32.
- Graham, M., Hale, S. A., & Gaffney, D. (2014). Where in the world are you? Geolocation and language identification in Twitter. *The Professional Geographer*, 66(4), 568-578.
- Harris, A. L., & Rea, A. (2009). Web 2.0 and virtual world technologies: A growing impact on IS education. *Journal of Information Systems Education*, 20(2), 137–144.
- Johnson, M., Egelman, S., & Bellovin, S. M. (2012, July). Facebook and privacy: it's complicated. In *Proceedings of the eighth symposium on usable privacy and security* (pp. 9-23). ACM.
- ITeLearn, (2016). .Net Framework Advantages and Disadvantages. Retrieved 21 April 2017, from <http://itelearn.com/blog/2016/07/20/dot-net-framework-advantages-and-disadvantages/>
- Kane, G. C., Fichman, R. G., Gallagher, J., & Glaser, J. (2009). Community relations 2.0. *Harvard business review*, 87(11), 45-50.
- Kaneko, T., & Yanai, K. (2016, April). Visual Event Mining from the Twitter Stream. In *Proceedings of the 25th International Conference Companion on World Wide Web* (pp. 51-52). International World Wide Web Conferences Steering Committee.
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons*, 53(1), 59-68.
- Kim, H. W., Chan, H. C., & Gupta, S. (2007). Value-based adoption of mobile internet: an empirical investigation. *Decision support systems*, 43(1), 111-126.
- Krasnova, H., Günther, O., Spiekermann, S., & Koroleva, K. (2009). Privacy concerns and identity in online social networks. *Identity in the Information Society*, 2(1), 39-63.
- Lin, K. Y., & Lu, H. P. (2011). Why people use social networking sites: An empirical study integrating network externalities and motivation theory. *Computers in Human Behavior*, 27(3), 1152-1161.
- Lotan, G., Graeff, E., Ananny, M., Gaffney, D., & Pearce, I. (2011). The Arab Spring| the revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions. *International journal of communication*, 5, 1375–1405.
- Marcus, A., Bernstein, M. S., Badar, O., Karger, D. R., Madden, S., & Miller, R. C. (2011, May). Twitinfo: aggregating and visualizing microblogs for event exploration. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 227-236). ACM.
- Mc Mahon, C. (2015). Why do we 'like' social media?. *PSYCHOLOGIST*, 28(9), 724-728.
- MobDevApp, (2015). Developers welcome Google Maps API integration with Java and Python libraries. Retrieved 29 April 2017, from <http://www.mobdevapp.com/developers-welcome-google-maps-api-integration-with-java-and-python-libraries/>

Moeller, N. & Sass, P. (2013). Twitter Visualization. Retrieved 1 May 2017, from <http://pages.cs.wisc.edu/~sass/>

Morstatter, F., Kumar, S., Liu, H., & Maciejewski, R. (2013, August). Understanding Twitter data with tweetexplorer. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1482-1485). ACM.

Murphy, I.B. (2013). *Fast Database Emerges from MIT Class, GPUs and Student's Invention*. Retrieved 19 May 2017, from <http://data-informed.com/fast-database-emerges-from-mit-class-gpus-and-students-invention>

Nakaji, Y., & Yanai, K. (2012, July). Visualization of real-world events with geotagged tweet photos. In *Multimedia and Expo Workshops (ICMEW), 2012 IEEE International Conference on* (pp. 272-277). IEEE.

News, (2017). RV FIRE ON HIGHWAY 101 CAUSES TRAFFIC DELAYS NEAR SAN FRANCISCO AIRPORT. Retrieved 8 May 2017, from <http://abc7news.com/traffic/rv-fire-on-highway-101-causes-traffic-delays-near-san-francisco-airport/1701637/>

Niazi, A. (2014). 10 ADVANTAGES OF C# PROGRAMMING LANGUAGE. Retrieved 27 April 2017, from <http://www.pprogramming.com/2016/04/top-10-advantages-of-c.html>

Obar, J. A., & Wildman, S. (2015). Social media definition and the governance challenge: An introduction to the special issue. *Telecommunications policy*. 39(9). 745–750

Paquet, S. (2003). Personal Knowledge publishing and its uses in research. Retrieved 13 January 2017, from <http://radio.weblogs.com/0110772/stories/2002/10/03/personalKnowledgePublishingAndItsUsesInResearch.html>

Periscope Data, (2013). Count Distinct Compared on Top 4 SQL Databases. Retrieved 22 April 2017, from <https://www.periscopedata.com/blog/count-distinct-in-mysql-postgres-sql-server-and-oracle.html>

Popovic, J. (2016). JSON in SQL Server 2016. Retrieved 23 April 2017, from <https://blogs.technet.microsoft.com/dataplatforminsider/2016/01/05/json-in-sql-server-2016-part-1-of-4/>

PostgreSQL, (2014). About PostgreSQL. Retrieved 28 March 2017, from <https://www.postgresql.org/about/>

Python, (2015). What is Python? Executive Summary. Retrieved 12 March 2017, from <https://www.python.org/doc/essays/blurb/>

Riquelme, F., & González-Cantergiani, P. (2016). Measuring user influence on Twitter: A survey. *Information Processing & Management*, 52(5), 949-975.

Rouse, M. (2013). Web 2.0. What Is. Retrieved 16 January 2017, from <http://whatis.techtarget.com/definition/Web-20-or-Web-2>

Samet, H., Sankaranarayanan, J., Lieberman, M. D., Adelfio, M. D., Fruin, B. C., Lotkowski, J. M., ... & Teitler, B. E. (2014). Reading news with maps by exploiting spatial synonyms. *Communications of the ACM*, 57(10), 64-77.

Sankaranarayanan, J., Samet, H., Teitler, B. E., Lieberman, M. D., & Sperling, J. (2009, November). Twitterstand: news in tweets. In *Proceedings of the 17th acm sigspatial international conference on advances in geographic information systems* (pp. 42-51). ACM.

Sechelea, A., Do Huu, T., Zimos, E., & Deligiannis, N. (2016, May). Twitter data clustering and visualization. In *Telecommunications (ICT), 2016 23rd International Conference on* (pp. 1-5). IEEE.

Schmidt, M., & Weiser, P. (2012). Web mapping services: development and trends. In Peterson, M. P. (2012), *Online maps with APIs and WebServices* (pp. 13-21). Berlin: Springer-Verlag.

Statista, (2016). Number of social network users worldwide from 2010 to 2020 (in billions). Retrieved 11 January 2017, from <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>

Statista, (2017). Leading social networks worldwide as of January 2017, ranked by number of active users (in millions). Retrieved 09 February 2017, from <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>

Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior*, 7(3), 321–326.

Tapscott, D. (2008). *Grown up digital: How the next generation is changing your world*. New York: McGraw-Hill.

Team Gwava, (2016). How Much Data is Created on the Internet Each Day?. Retrieved 18 January 2017, from <https://www.gwava.com/blog/internet-data-created-daily>

Twitter, (2012a). Twitter Developer Documentation - Streaming APIs. Retrieved 11 February 2017, from <https://dev.Twitter.com/streaming/overview>

Twitter, (2012b). Twitter Developer Documentation - Tweets. Retrieved 11 February 2017, from <https://dev.Twitter.com/overview/api/tweets>

Twitter, (2013). Company | About. Retrieved 23 January 2017, from <https://about.Twitter.com/company>

Twitter, (2015). Getting started with Twitter. Retrieved 11 February 2017, from <https://support.Twitter.com/articles/215585>

Von Muhlen, M., & Ohno-Machado, L. (2012). Reviewing social media use by clinicians. *Journal of the American Medical Informatics Association*, 19(5), 777-781.

Vosecky, J., Jiang, D., & Ng, W. (2013, March). Limosa: A system for geographic user interest analysis in Twitter. In *Proceedings of the 16th International Conference on Extending Database Technology* (pp. 709-712). ACM.

Wanichayapong, N., Pruthipunyaskul, W., Pattara-Atikom, W., & Chaovalit, P. (2011, August). Social-based traffic information extraction and classification. In *ITS Telecommunications (ITST), 2011 11th International Conference on* (pp. 107-112). IEEE.

Whiting, A., & Williams, D. (2013). Why people use social media: a uses and gratifications approach. *Qualitative Market Research: An International Journal*, 16(4), 362-369.

Zhao, D., & Rosson, M. B. (2009, May). How and why people Twitter: the role that micro-blogging plays in informal communication at work. In *Proceedings of the ACM 2009 international conference on Supporting group work* (pp. 243-252). ACM.