

ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ



– Πτυχιακή Εργασία –

Σύστημα παρακολούθησης κοινωνικών φαινομένων
σε πραγματικό χρόνο μέσω της χρήσης κοινωνικών δικτύων

Κάρολος Ταλβής

Επιβλέπων: – Κωνσταντίνος Χωριανόπουλος –

19 Σεπτεμβρίου 2013

Επιβλέπων

Κωνσταντίνος Χωριανόπουλος, Λέκτορας,
Ιόνιο Πανεπιστήμιο

Τριμελής Επιτροπή

Κάτια - Λήδα Κέρμανίδου, Λέκτορας,
Ιόνιο Πανεπιστήμιο

Κωνσταντίνος Χωριανόπουλος, Λέκτορας,
Ιόνιο Πανεπιστήμιο

Μάρκος Αυλωνίτης, Λέκτορας,
Ιόνιο Πανεπιστήμιο

Αυτή η προσπάθεια αφιερώνεται στους φοιτητές του Τμήματος Πληροφορικής του
Ιονίου Πανεπιστημίου

Περίληψη

Σκοπός αυτού του κειμένου είναι η παρουσίαση του συστήματος πληροφορικής που αναπτύχθηκε στα πλαίσια της διπλωματικής εργασίας του τμήματος Πληροφορικής του Ιονίου Πανεπιστημίου, Κέρκυρας. Η επιλογή της εργασίας έγινε με βάση τις υποδείξεις του επιβλέποντα καθηγητή, τη χρησιμότητα του έργου, την ακαδημαϊκή προσφορά λόγω της δυσκολίας του και των προσωπικών προτιμήσεων. Για τις ανάγκες της εργασίας, αναπτύχθηκε ένα σύστημα ευρείας χρήσης που εκμεταλλεύεται τα δεδομένα των μέσων κοινωνικής δικτύωσης. Πιο συγκεκριμένα, για την επαλήθευση της ορθής λειτουργίας του συστήματος, αναπτύχθηκε το πρωτότυπο σύστημα Flutrack, του οποίου η λειτουργία είναι ο εντοπισμός και η οπτικοποίησή συμπτωμάτων εποχικής γρίπης. Ο έγκαιρος εντοπισμός και η πρόβλεψη της εποχικής γρίπης μπορούν να οδηγήσουν στην πρόληψη επιδημιών γρίπης και την προστασία κοινωνικά ευαίσθητων ομάδων. Το λειτουργικό πρωτότυπο που αναπτύχθηκε, προτείνει ένα διαφορετικό, σε σχέση με τους υπάρχοντες τρόπους εντοπισμού της γρίπης. Οι στόχοι και οι απαιτήσεις του συστήματος αναλύονται διεξοδικά. Εκτενής ανάλυση γίνεται ακόμη σε ότι αφορά το σχεδιασμό, την αρχιτεκτονική και την υλοποίηση του συστήματος. Επίσης, γίνεται λόγος για τα συμπεράσματα που προκύπτουν από τα αποτελέσματα της εφαρμογής. Ακόμη, δεν παραλείπονται στοιχεία που σχετίζονται με ηθικά ζητήματα.

Πρόλογος και Ευχαριστίες

Η ραγδαία εξάπλωση των κοινωνικών δικτύων μέσω του διαδικτύου και η ευκολία με την οποία μας επιτρέπεται να προσπελαύνουμε τα δεδομένα που προσφέρουν, αποτελεί αντικείμενο μελέτης και ίσως προβληματισμού. Ταυτόχρονα όμως, αποτέλεσε πηγή έμπνευσης για την εκπόνηση της συγκεκριμένης διπλωματικής εργασίας. Τα ανοιχτά δεδομένα, η ανάπτυξη εξειδικευμένων γλωσσών προγραμματισμού για τον παγκόσμιο Ιστό και οι σύγχρονες τεχνολογίες ανοιχτού κώδικα δημιούργησαν την πρόκληση της ανάπτυξης μιας δικτυακής εφαρμογής που επεξεργάζεται και οπτικοποιεί σε πραγματικό χρόνο δεδομένα που προέρχονται από κοινωνικά δίκτυα.

Ο προγραμματισμός και η ανάπτυξη εφαρμογών είναι εργασίες πολύπλοκες και γεμάτες λεπτομέρειες. Η παρακολούθηση της λογικής βάσης με την οποία δημιουργήθηκε ένα πρόγραμμα από κάποιον τρίτο, όπως και των λεπτομερειών του προγράμματος αυτού, είναι διαδικασία πολύπλοκη και απαιτεί προσοχή. Για το λόγο αυτό, στο παρόν έγγραφο περιγράφονται με τρόπο αναλυτικό, οι ιδέες που οδήγησαν στην ανάπτυξη της εφαρμογής, καθώς και οι τεχνολογίες που χρησιμοποιήθηκαν. Έχει δοθεί ιδιαίτερη έμφαση στην παρουσίαση του εγχειρήματος από μια ευρύτερη οπτική γωνία σε σχέση με αυτήν ενός απλού εγχειριδίου λειτουργίας της εφαρμογής-συστήματος. Οι τεχνολογίες αλλάζουν και πρέπει καθημερινά να γίνεται έλεγχος από διαφορετική σκοπιά, ακόμα και για αυτά που στοιχειοθετούν βασικές αρχές και αναλλοίωτες έννοιες. Έγινε προσπάθεια περιγραφής όχι μόνο του πρωτοτύπου που αναπτύχθηκε, αλλά της γενικότερης ιδέας, έννοιας και χρησιμότητας της διπλωματικής αυτής εργασίας. Η σύλληψη δυσδιάκριτων απαιτήσεων, η υλοποίηση αυτών και ο τρόπος έκφρασης τους, οδήγησαν στη λήψη συγκεκριμένων αποφάσεων που ίσως να μην είναι οι βέλτιστες, μεθοδεύουν όμως την εργασία με τρόπο τέτοιο, ώστε να χρησιμεύσει ως βάση για άλλα έργα, πιο σύνθετα. Εξάλλου, δεν υπάρχει βέλτιστη λύση, παρά μόνο συστήματα τα οποία είναι καταλληλότερα κάτω από ορισμένες συνθήκες.

Θα ήθελα να ευχαριστήσω τους καθηγητές του τμήματος Πληροφορικής του Ιονίου Πανεπι-

στημίου, Κέρκυρας που τα τελευταία χρόνια πρόσφεραν τις γνώσεις τους και συνέβαλαν στην προσωπική μου εξέλιξη. Πιο συγκεκριμένα, θα ήθελα να εκφράσω τις ευχαριστίες μου στον επιβλέπων καθηγητή της διπλωματικής μου εργασίας, κ.Κωσταντίνο Χωριανόπουλο διότι χωρίς τις καινοτόμες ιδέες, την καθημερινή καθοδήγηση, την προσωπική υποστήριξη για πολύ μεγάλο χρονικό διάστημα και τις συμβουλές του δεν θα είχα ολοκληρώσει την εκπόνηση του συγκεκριμένου έργου. Τέλος θα ήθελα να ευχαριστήσω όλους όσους στάθηκαν δίπλα μου και βοήθησαν σε διάφορα στάδια της εργασίας, σε όλα όσα στερήθηκα και μου έλειψαν σε αυτό το χρονικό διάστημα.

Κέρκυρα, 19 Σεπτεμβρίου 2013

Κάρολος Ταλβής

Περιεχόμενα

A´	Εισαγωγή	1
A´.1	ΑΝΤΙΚΕΙΜΕΝΟ ΤΗΣ ΕΡΓΑΣΙΑΣ	1
A´.2	ΣΚΟΠΟΣ ΤΗΣ ΕΡΓΑΣΙΑΣ	2
A´.3	ΔΙΑΡΘΡΩΣΗ ΤΗΣ ΕΡΓΑΣΙΑΣ	2
A´.4	ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ	3
A´.4.1	Εποχική γρίπη	3
A´.4.2	Πανδημίες γρίπης	5
A´.4.3	Ανάγκη για παρακολούθηση	5
A´.4.4	Διεπαφή Προγραμματισμού Εφαρμογών (API)	6
A´.4.5	Υπηρεσίες Κοινωνικών δικτύων	6
A´.4.6	Twitter	7
B´	Σχετικές Εργασίες	8
B´.1	ΠΑΡΑΔΟΣΙΑΚΕΣ ΜΕΘΟΔΟΙ ΠΑΡΑΚΟΛΟΥΘΗΣΗΣ ΤΗΣ ΓΡΙΠΗΣ	8
B´.1.1	Δίκτυο Εθνικών Εργαστηρίων Αναφοράς Γρίπης	8
B´.1.2	Κέντρα ελέγχου και πρόληψης νοσημάτων	9
B´.2	ΕΝΑΛΛΑΚΤΙΚΕΣ ΠΡΟΣΕΓΓΙΣΕΙΣ ΠΑΡΑΚΟΛΟΥΘΗΣΗΣ ΤΗΣ ΓΡΙΠΗΣ	9
B´.2.1	Διαδικτυακές εφαρμογές	10
B´.2.2	Άρθρα βασισμένα στις ιστοσελίδες κοινωνικής δικτύωσης	12
Γ´	Κύριο Μέρος	14
Γ´.1	ΣΧΕΔΙΑΣΜΟΣ	14
Γ´.1.1	Καθορισμός απαιτήσεων	14
Γ´.1.2	Καθορισμός των αναγκών του χρήστη	15

Γ'.1.3	Μεθοδολογία	16
Γ'.1.4	Model View Controller (MVC)	16
Γ'.1.5	Δομική ανάλυση του έργου	17
Γ'.1.6	Εννοιολογικό Μοντέλο (Conceptual Model)	18
Γ'.1.7	Απαιτήσεις σχεδιασμού	20
Γ'.2	ΥΛΟΠΟΙΗΣΗ	20
Γ'.2.1	Το πρωτότυπο Flutrack	20
Γ'.2.2	Ετικέτες (Tags) παρακολούθησης	21
Γ'.2.3	Αρχιτεκτονική συστήματος Flutrack	22
Γ'.2.4	Γλώσσες προγραμματισμού και εργαλεία	22
Γ'.2.4.1	PHP	22
Γ'.2.4.2	CURL	23
Γ'.2.4.3	SQL	23
Γ'.2.4.4	Javascript	23
Γ'.2.4.5	HTML	24
Γ'.2.4.6	CSS	25
Γ'.2.4.7	Google Maps	25
Γ'.2.5	Back-end τμήμα	26
Γ'.2.5.1	Αίτηση στη Διεπαφή Προγραμματισμού Εφαρμογών	26
Γ'.2.5.2	Ρυθμός παρακολούθησης	27
Γ'.2.5.3	Μεταδεδομένα	28
Γ'.2.5.4	Επιδείνωση κλινικής κατάστασης χρήστη	28
Γ'.2.5.5	Φιλτράρισμα	29
Γ'.2.5.6	Βάση δεδομένων	30
Γ'.2.5.7	Διεπαφή Προγραμματισμού Εφαρμογών (Flutrack API)	31
Γ'.2.6	Front-end τμήμα	32
Γ'.2.6.1	Οπτικοποίηση	32
Γ'.2.6.2	Ιστοσελίδα εφαρμογής	33
Δ'	Αποτελέσματα και συμπεράσματα	35
Δ'.1	ΑΠΟΤΕΛΕΣΜΑΤΑ	35
Δ'.2	ΜΕΛΛΟΝΤΙΚΗ ΕΡΓΑΣΙΑ	37
Δ'.3	ΗΘΙΚΑ ΖΗΤΗΜΑΤΑ	38

Περιεχόμενα	vi
Ε΄ Επίλογος	40
Ε΄.1 ΣΥΝΟΨΗ	40
Παράρτημα Α΄	41
Βιβλιογραφία	42
Συντμήσεις	44
Γλωσσάρι Ξενικών Όρων	46

Κατάλογος Σχημάτων

Γ'.1	<i>Model View Controller</i> διάγραμμα	17
Γ'.2	<i>Work Breakdown Structure</i> διάγραμμα	18
Γ'.3	<i>Conceptual Model</i> διάγραμμα	19
Γ'.4	διάγραμμα διαδικασίας <i>Call Request</i>	26
Γ'.5	διάγραμμα διαδικασίας <i>Streaming</i>	27
Γ'.6	διάγραμμα μεταδεδομένων	29
Γ'.7	Παράδειγμα οπτικοποίησης σε σε συγκεκριμένες ημερομηνίες	33
Γ'.8	Αρχική σελίδα του ιστοτόπου http://flutrack.org	34
Δ'.9	Συγκριτικοί πίνακες δεδομένων των υπηρεσιών <i>Google Flu Trends</i> και <i>Flutrack</i> ανά εβδομάδα, ενδεικτικά από τις 2/12/12 έως 13/1/2013.	36
Δ'.10	διάγραμμα διασποράς δεδομένων των υπηρεσιών <i>Google Flu Trends</i> και <i>Flutrack</i>	37

Κεφάλαιο Α΄

Εισαγωγή

Α΄.1 Αντικείμενο της εργασίας

Αναντίρρητα, το διαδίκτυο έχει γίνει μέρος της καθημερινότητας των περισσότερων ανθρώπων και αυτό είχε ως αποτέλεσμα τη χρήση ιστοσελίδων κοινωνική δικτύωσης από εκατομύρια χρήστες ανά τον κόσμο. Οι ιστοσελίδες κοινωνικής δικτύωσης αποτελούν εικονικές κοινότητες, όπου οι χρήστες του διαδικτύου έχουν τη δυνατότητα να δημιουργήσουν εικονικά προφίλ και να αναπτύξουν ένα δίκτυο επαφών, με τις οποίες μπορούν να επικοινωνούν μέσω της ιστοσελίδας. Η εύκολη πρόσβαση στα μέσα κοινωνικής δικτύωσης από κάθε λόγής συσκευές που υποστηρίζουν πρόσβαση στο διαδίκτυο είχε ως αποτέλεσμα την διαρκώς αυξανόμενη τάση χρήσης των δικτύων αυτών. Εκατομμύρια χρήστες κάθε δευτερόλεπτο μοιράζονται από σκέψεις και συναισθήματα μέχρι τις αλλαγές της κατάστασης των προσωπικών και οικονομικών τους στοιχείων, αλλά και της υγείας τους. Έτσι, δημιουργήθηκαν σκέψεις για το πως αυτή η τάση των χρηστών θα μπορούσε να φανεί χρήσιμη και άξια αναφοράς σε ευρύτερα πλαίσια, εκτός από αυτά του κλαδού της επιστήμης της πληροφορικής. Τα τελευταία χρόνια γίνεται σημαντική προσπάθεια στο να χρησιμοποιηθούν τα δεδομένα των χρηστών των κοινωνικών δικτύων, ούτως ώστε, έπειτα από επεξεργασία να βοηθήσουν στην εξαγωγή συμπερασμάτων και στατιστικών, αλλά και να συμβάλλουν στην επίλυση προβλημάτων με εναλλακτικές μεθόδους. Συνεπώς, ως αντικείμενο της συγκεκριμένης εργασίας θεμελιώθηκε, να είναι η ανάπτυξη μιας διαδραστικής εφαρμογής που συλλέγει και επεξεργάζεται δεδομένα σε πραγματικό χρόνο, χρησιμοποιώντας τις διεπαφές προγραμματισμού εφαρμογών (API) από ιστοσελίδες κοινωνικής δικτύωσης. Μετά από μελέτη αποφασίστηκε επίσης ότι τα δεδομένα που θα συλλέγονται θα

αφορούν στο θέμα της παρακολούθησης των συμπτωμάτων της εποχικής γρίπης μέσω του κοινωνικού δικτύου Twitter.

Η παρακολούθηση των συμπτωμάτων της εποχικής γρίπης επιλέχτηκε αρχικά με γνώμονα το ότι είναι σχετικά εύκολο να αυτοδιαγνωστεί από τον ίδιο τον ασθενή. Με την παρακολούθηση, την επεξεργασία και την ανάλυση των σχετικών δεδομένων καθώς και τη γεωγραφική θέση των χρηστών-ασθενών στο χάρτη, μπορούν να εξαχθούν πληροφορίες που είναι ικανές να οδηγήσουν στον εντοπισμό της εποχικής γρίπης ανά την υφήλιο σε πραγματικό χρόνο. Επιπλέον, ο εντοπισμός της γρίπης σε πραγματικό χρόνο για μεγάλο χρονικό διάστημα, υπό κατάλληλη ανάλυση μπορεί να οδηγήσει στην πρόβλεψη και την πρόληψη της εξάπλωσης της γρίπης, όπως και των επιδημιών που είναι πιθανό να ξεσπάσουν. Για τους λόγους που προαναφέρθηκαν, αλλά και για άλλους που θα παρουσιαστούν παρακάτω, αναπτύχθηκε η διαδικτυακή πλατφόρμα Flutrack. Η πλατφόρμα-εφαρμογή Flutrack θα παρουσιαστεί αναλυτικά στο κύριο μέρος της εργασίας αυτής.

Α΄.2 Σκοπός της εργασίας

Σκοπός της εργασίας ήταν ο σχεδιασμός και η ανάπτυξη ενός σύνθετου συστήματος που εξυπηρετεί την διεύρυνση των ακαδημαϊκών γνώσεων σε συνδυασμό με τη μελέτη ενός υπαρκτού φαινομένου, η οποία προσδίδει ένα λόγο ύπαρξης στο έργο. Επίσης, η ανάπτυξη και διαχείριση του έργου αυτού, αποτελεί μια οργανωμένη προσέγγιση που μπορεί να χρησιμοποιηθεί με διαφορετικούς τρόπους και η συγκεκριμένη υλοποίηση αποτελεί μονάχα ένα πρωτότυπο, το οποίο αποδεικνύει τη χρησιμότητα και την ευκολία χρήσης του. Γενικότερος σκοπός της εργασίας δηλαδή, ήταν η ανάπτυξη ενός ολοκληρωμένου συστήματος πληροφόρησης και ελέγχου που μπορεί να χρησιμοποιηθεί για ποικίλα προβλήματα και λήψη αποφάσεων. Ο ειδικός στόχος του πρωτοτύπου που αναπτύχθηκε, δηλαδή η μελέτη του φαινομένου της εποχικής γρίπης μέσω των ιστοσελίδων κοινωνικής δικτύωσης, αποτελεί την απόδειξη ότι το σύστημα είναι ικανό να λειτουργήσει αποτελεσματικά με τις συγκεκριμένες απαιτήσεις ενός πρωτοτύπου.

Α΄.3 Διάρθρωση της εργασίας

Η εργασία ξεκινάει με την παρουσίαση του απαραίτητου θεωρητικού υποβάθρου που σχετίζεται με την εποχική γρίπη και όλα όσα θεωρούνται απαραίτητα ώστε να μπορεί ο αναγνώστης, ανεξάρτητα από τις γνώσεις του στην πληροφορική, να παρακολουθήσει τη ροή της ανάλυσης

της εργασίας. Στη συνέχεια ακολουθεί η παρουσίαση σύγχρονων μεθόδων εντοπισμού της γρίπης. Αναλύονται διαδικτυακές και μη, εργασίες κλάδων της Πληροφορικής που σχετίζονται με την ανίχνευση της εξάπλωσης της εποχικής γρίπης και των επιδημιών αυτής. Στα επόμενα κεφάλαια γίνεται αναλυτική περιγραφή της μεθοδολογίας, της σχεδίασης και της υλοποίησης της πλατφόρμας Flutrack. Ακολουθεί η εξαγωγή συμπερασμάτων, σχέψεις για περαιτέρω μελλοντική εργασία, ηθικά ζητήματα και προβληματισμοί και η εργασία κλείνει με τον επίλογο.

Α΄.4 Θεωρητικό υπόβαθρο

Α΄.4.1 Εποχική γρίπη

Ο ιός της γρίπης ευθύνεται για την εποχική γρίπη που παρουσιάζεται κάθε χρόνο στις χώρες του νοτίου και του βορείου ημισφαιρίου, εναλλάξ, κυρίως κατά την χειμερινή περίοδο. Η εποχική γρίπη οφείλεται σε μικρές συνήθως αντιγονικές μεταβολές των ιών γρίπης τύπου Α και τύπου Β που συμβάλλουν στη μολυσματικότητά του. Τα νέα στελέχη που προκύπτουν μολύνουν ακόμη και ανθρώπους οι οποίοι είχαν ανοσία λόγω προηγούμενης μόλυνσης ή εμβολιασμού στα στελέχη ιού γρίπης από τα οποία προήλθαν και έτσι προκαλούν τις επιδημίες της εποχικής γρίπης. Η γρίπη είναι μια ιογενής λοίμωξη που προσβάλλει κυρίως το λαιμό, τη μύτη, τους βρόγχους και τους πνεύμονες περιστασιακά. Θεωρείται ως ένα από τα πιο κοινά σύνδρομα μόλυνσης των ανθρώπων και τα συμπτώματά της είναι τόσο κοινά ώστε μπορεί να γίνει αυτοδιάγνωση από τους ίδιους τους ασθενείς χωρίς κλινική διάγνωση από θεράποντες ιατρούς.

Πολλές φορές η γρίπη συγχέεται με το κοινό κρύωμα, διαφέρει όμως από αυτό καθώς προκαλείται από διαφορετικές ομάδες ιών και τα συμπτώματά είναι πιο έντονα και διαρκούν περισσότερο. Η λοίμωξη διαρκεί περίπου μια εβδομάδα και χαρακτηρίζεται από αιφνίδιο υψηλό πυρετό, πόνος στους μύες, πονοκέφαλο και σοβαρή αδιαθεσία, μη παραγωγικό βήχα, πονόλαιμο και ρινίτιδα. Τα συμπτώματα συνήθως κορυφώνονται μετά από δύο ή τρεις ημέρες και έπειτα αρχίζουν να υποχωρούν. Πιο σπάνια συμπτώματα όπως ο εμετός και η διάρροια οφείλονται στη γρίπη, διαχωρισμός που είναι δυσδιάκριτος διότι τα συμπτώματα αυτά είναι πιο πιθανό να προκληθούν από άλλου είδους λοιμώξεις, όπως για παράδειγμα η ιογενής γαστρεντερίτιδα [1].

Οι περισσότεροι άνθρωποι που θα νοσήσουν από γρίπη θα έχουν ήπια συμπτώματα, δε θα χρειαστούν ιατρική φροντίδα ή αντιικά φάρμακα και θα αναρρώσουν μέσα σε λιγότερο από δύο εβδομάδες. Κάποιοι όμως, είναι πιθανό να έχουν επιπλοκές οι οποίες να οδηγήσουν σε νοση-

λεία ή ακόμα και σε θάνατο. Η Πνευμονία η βρογχίτιδα, οι λοιμώξεις κόλπων και των αυτιών είναι κάποια παραδείγματα των επιπλοκών που σχετίζονται με τη γρίπη. Οι ευπαθείς κοινωνικές ομάδες, δηλαδή οι ηλικιωμένοι, τα παιδιά, οι έγκυες γυναίκες και άτομα που πάσχουν από ορισμένα χρόνια νοσήματα κινδυνεύουν περισσότερο από σοβαρές επιπλοκές της γρίπης. Η εποχική γρίπη ευθύνεται για 250.000 με 500.000 θανάτους κάθε χρόνο σε όλο τον κόσμο. Σύμφωνα με στοιχεία από τις ΗΠΑ, η εποχική γρίπη, είναι υπεύθυνη ετησίως για 25.000.000 επισκέψεις σε ιατρούς, για περίπου 150.000 εισαγωγές σε νοσοκομεία, 40.000 θανάτους και αποτελεί τη συχνότερη αιτία κατανάλωσης αντιβιοτικών που δεν χρειάζονται [2][3].

Άτομα σε κίνδυνο επιπλοκών σχετικών με τη γρίπη είναι:

- Παιδιά μικρότερα των 5 ετών , ειδικά παιδιά κάτω των 2 ετών.
- Ενήλικες 65 ετών και άνω.
- Έγκυες γυναίκες.

Άτομα με χρόνια προβλήματα υγείας όπως:

- Άσθμα
- Νευρολογικές διαταραχές.
- Χρόνιες νόσοι των πνευμόνων, αποφρακτική πνευμονοπάθεια και κυστική ίνωση.
- Καρδιοπάθειες όπως συμφορητική καρδιακή ανεπάρκεια και στεφανιαία νόσος.
- Διαταραχές αίματος (δρεπανοκυτταρική αναιμία).
- Ενδοκρινικές νόσοι (όπως σακχαρώδης διαβήτης).
- Νεφρικές νόσοι.
- Νόσοι του ήπατος.
- Μεταβολικές διαταραχές.
- Εξασθένηση ανοσοποιητικού συστήματος λόγω νόσου ή αγωγής (ΗIV ή ΑΙΔΣ, καρκίνος, χρήση στεροειδών).
- Άτομα μικρότερα των 19 ετών που λαμβάνουν χρόνια θεραπεία με ασπιρίνη.
- Νοσηρή παχυσαρκία (δείκτης μάζας σώματος 40 άνω).

Α΄.4.2 Πανδημίες γρίπης

Κατά τον 20ο αιώνα συνέβησαν τρεις μεγάλες πανδημίες γρίπης:

- Η Ισπανική γρίπη' την περίοδο 1918-1920, στην οποία έχασαν τη ζωή τους εκατομμύρια άνθρωποι από τον ιό της γρίπης A(H1N1).
- Η Άσιατική γρίπη' του 1957 από τον ιό γρίπης A(H2N2).
- Η γρίπη του 'Hong Kong' το 1968 από τον ιό γρίπης τύπου A(H3N2).

Πανδημίες γρίπης συμβαίνουν είτε όταν ένας ιός της γρίπης εξαπλώνεται στον ανθρώπινο οργανισμό από κάποιο είδος ζώου (π.χ. πτηνά ή θηλαστικά), είτε όταν ένας ιός γρίπης του ανθρώπου αποκτήσει νέα γονίδια από έναν ιό που συνήθως προσβάλλει πτηνά ή χοίρους [4]. Στις περιπτώσεις αυτές, ο ανθρώπινος πληθυσμός δεν έχει ανοσία έναντι του μεταλλαγμένου ιού, με συνέπεια να διευκολύνεται η ταχύτατη εξάπλωσή του. Η πρόσφατη πανδημία από τον ιό γρίπης A(H1N1) που ξεκίνησε την άνοιξη του 2009 στο Μεξικό για να εξαπλωθεί ταχύτατα σε όλο τον κόσμο αποτελεί το καλύτερο παράδειγμα για τη δυνατότητα του ιού της γρίπης να προκαλεί πανδημίες. Ο ιός αυτός περιείχε συνδυασμό γονιδίων από ιούς γρίπης που προσβάλλουν τον άνθρωπο, το χοίρο και τα πτηνά. Η δραματική εμπειρία της ανθρωπότητας από τις τρεις πιο πρόσφατες πανδημίες γρίπης του εικοστού αιώνα που έσπειραν το θάνατο σε εκατομμύρια ανθρώπους σε όλο τον κόσμο, δίδαξε την αναγκαιότητα της συνεχούς επαγρύπνησης και επιτήρησης για την πιθανότητα εμφάνισης νεοεμφανιζόμενων στελεχών γρίπης με δυνατότητα πρόκλησης πανδημίας [5][6] .

Α΄.4.3 Ανάγκη για παρακολούθηση

Η υψηλή μεταλλακτικότητα του ιού της γρίπης που έχει σαν αποτέλεσμα τη ανάδυση εντελώς νέων στελεχών στα οποία ο πληθυσμός δεν είναι άνοσος, πιθανώς με μεγάλη λοιμογόνο οξύτητα, καθιστά τον ιό πραγματική απειλή για την ανθρωπότητα. Επίσης, οι σύγχρονες συνθήκες ζωής με την αστικοποίηση και την αύξηση των παγκόσμιων μεταφορών και επικοινωνίας αποτελούν το ιδανικό υπόστρωμα για την ταχύτατη εξάπλωση ενός δυνητικά πανδημικού στελέχους. Συνεπώς, γίνεται αντιληπτή η άμεση ανάγκη για συνεχή πρόληψη (μέσω εμβολίων) και παρακολούθηση της εποχικής γρίπης με οποιοδήποτε τρόπο είναι κάτι τέτοιο εφικτό [6].

Α΄.4.4 Διεπαφή Προγραμματισμού Εφαρμογών (API)

Διεπαφή Προγραμματισμού Εφαρμογών ονομάζεται η διεπαφή των προγραμματιστικών διαδικασιών που ένα λειτουργικό σύστημα, βιβλιοθήκη ή εφαρμογή που επιτρέπει να γίνονται προς αυτό αιτήσεις από άλλα προγράμματα και προσφέρει αποστολή ή/και ανταλλαγή δεδομένων. Σε απλούστευση, μια Διεπαφή Προγραμματισμού Εφαρμογών είναι ένα σύνολο προγραμματιστικών εντολών και προτύπων, το οποίο δίνει πρόσβαση σε μια διαδικτυακή εφαρμογή ή εργαλείο. Πρόκειται για ένα σύνολο από ρουτίνες (Routines), δομές δεδομένων (Data Structures), κλάσσεις αντικειμένων (Object Classes) και/ή πρωτόκολλα (Protocols) που υποστηρίζουν την ανάπτυξη λογισμικού και υπηρεσιών βασισμένων σε μια πρωτογενή πηγή (υπολογιστικό σύστημα, βιβλιοθήκη, εφαρμογή).

Μια εταιρεία λογισμικού απελευθερώνει μια Διεπαφή Προγραμματισμού Εφαρμογών στο κοινό, έτσι ώστε να επιτρέπει σε άλλους προγραμματιστές λογισμικού να αναπτύσσουν προϊόντα και υπηρεσίες που τροφοδοτούνται από τα δεδομένα που προσφέρει. Το βασικό προαπαιτούμενο για μια Διεπαφή Προγραμματισμού Εφαρμογών είναι η δυνατότητα άντλησης του συνόλου των δεδομένων σε ένα πρότυπο που είναι ευδιαχείριστο. Έχοντας τα δεδομένα πάντα διαθέσιμα σε όλους τους ενδιαφερόμενους είναι δυνατή η δημιουργία νέων Διεπαφών Προγραμματισμού Εφαρμογών που θα τηρούν τις γενικές αρχές υλοποίησής τους αλλά θα προσφέρουν νέα παροχή δεδομένων σύμφωνα με τις ανάγκες των χρηστών. Θα πρέπει να τονιστεί ότι μία Διεπαφή Προγραμματισμού Εφαρμογών είναι μία λογισμικό-σε-λογισμικό διεπαφή και όχι μια διεπαφή χρήστη. Αποτελεί ουσιαστικά ένα τρόπο σύνδεσης εφαρμογών χωρίς μεσολάβηση από τους χρήστες.

Α΄.4.5 Υπηρεσίες Κοινωνικών δικτύων

Με τον όρο μέσο μέσο κοινωνικής δικτύωσης εννοείται κάθε υπηρεσία, η οποία δια μέσου ενός απομακρυσμένου δικτύου υπολογιστών, δηλαδή του διαδικτύου, επιτέπει τη δημιουργία ενός δικτύου ανθρώπων. Ένα κοινωνικό δίκτυο επομένως, αναφέρεται στα μέσα αλληλεπίδρασης ομάδων ανθρώπων μέσω διαδικτυακών κοινοτήτων. Τα τελευταία χρόνια εμφανίζονται υπηρεσίες κοινωνικών δικτύων με διαφορετικές μορφές, όπως αυτές ιστολογίων, ιστοσελίδων, εικονικών δωματίων δημοσίων συζητήσεων (Forums) κτλ.

Α΄.4.6 Twitter

Το Twitter δημιουργήθηκε το 2006 από τον Jack Dorsey και αποτελεί έναν ιστοχώρο - υπηρεσία κοινωνικής δικτύωσης που επιτρέπει στους χρήστες του να στέλνουν και να διαβάζουν σύντομα μηνύματα (μέχρι 140 χαρακτήρες), τα οποία ονομάζονται Tweets.

Η εφαρμογή είναι μία από τις 10 δημοφιλέστερες ιστοσελίδες του διαδικτύου και αριθμεί μια κοινότητα 500 εκατομμυρίων ενεργών χρηστών, οι οποίοι δημιουργούν παραπάνω από 340 εκατομμύρια Tweets κάθε ημέρα. Παρά τα υψηλά επίπεδα "γλωσσολογικού θορύβου", το Twitter περιέχει τεράστιο όγκο δεδομένων από τον οποίο, έπειτα από κατάλληλη επεξεργασία μπορούν να εξαχθούν χρήσιμες πληροφορίες. Σημαντικό ποσοστό των δεδομένων απο αποθηκεύονται στην υπηρεσία, αναδιανέμονται μέσω της ανοιχτής Διεπαφής Προγραμματισμού Εφαρμογών που προσφέρει.

Η Διεπαφή Προγραμματισμού Εφαρμογών του Twitter αναδιανέμει τα δεδομένα που ζητούνται από ενδιαφερόμενους προγραμματιστές μέσω ποικίλων βιβλιοθηκών. Η υπηρεσία παρέχει βιβλιοθήκες για πολλά είδη γλωσσών προγραμματισμού, όπως αντικειμενοστρεφείς γλώσσες, γλώσσες βασισμένες σε διακομιστές που λειτουργούν στο διαδίκτυο κτλ. Ή διεπαφή προσφέρει δεδομένα για όλες τις πληροφορίες του προφίλ κάθε χρήστη καθώς και των Tweets που αυτός δημοσιεύει.

Κεφάλαιο Β΄

Σχετικές Εργασίες

Β΄.1 Παραδοσιακές μέθοδοι παρακολούθησης της γρίπης

Β΄.1.1 Δίκτυο Εθνικών Εργαστηρίων Αναφοράς Γρίπης

Ο Παγκόσμιος Οργανισμός Υγείας - WHO έχει δημιουργήσει δίκτυο εθνικών εργαστηρίων για την παρακολούθηση των μεταβολών του ιού της γρίπης (GISRS). Αντικείμενο του οργανισμού είναι:

- Η έγκαιρη αναγνώριση και ο χαρακτηρισμός νέων υποτύπων ιού γρίπης Α που έχουν τη δυνατότητα να προκαλέσουν πανδημίες.
- Η αποστολή αντιπροσωπευτικών στελεχών ιού γρίπης στο διεθνές σύστημα παρακολούθησης της εξέλιξης του ιού της γρίπης.
- Η διάγνωση κρουσμάτων γρίπης και η παρακολούθηση της εξέλιξης του ιού σε περίπτωση πανδημίας.
- Η μελέτη της αντοχής των στελεχών των ιών γρίπης στα αντι-ιικά φάρμακα.

Β'.1.2 Κέντρα ελέγχου και πρόληψης νοσημάτων

Τα κέντρα ελέγχου και πρόληψης νοσημάτων συλλέγουν στοιχεία για τη γρίπη από συμβεβλημένα νοσοκομεία και ιατρούς. Έπειτα από αναλύσεις, δημοσιοποιούν εβδομαδιαία και ετήσια εκθέσεις που αφορούν την παρακολούθηση και τις εξάρσεις της εποχιακής γρίπης.

- Στην Ελλάδα, η γρίπη επιηρεείται από το Τμήμα Επιδημιολογικής Επιτήρησης και Παρέμβασης του ΚΕ.ΕΛ.Π.ΝΟ τόσο κλινικά όσο και εργαστηριακά. Το σύστημα παρατηρητών νοσηρότητας της πρωτοβάθμιας φροντίδας υγείας αποτελείται από ένα δίκτυο ιδιωτών ιατρών, Μονάδων υγείας ΙΚΑ, ιατρών Κέντρων Υγείας και περιφερειακών ιατρών. Κάθε εβδομάδα συλλέγονται στοιχεία που αφορούν στον αριθμό επισκέψεων για κάθε αιτία και τον αριθμό επισκέψεων λόγω "γριπώδους συνδρομής". Επίσης, στο πλαίσιο επιτήρησης της γρίπης γίνεται καταγραφή και διερεύνηση των σοβαρών κρουσμάτων γρίπης με ή χωρίς νοσηλεία σε Μονάδες Εντατικής Θεραπείας (Μ.Ε.Θ), καθώς και των θανάτων από εργαστηριακά επιβεβαιωμένη γρίπη. Τα παραπάνω στοιχεία αναλύονται με στόχους ως προς τον μόνιμο πληθυσμό της χώρας κατά γεωγραφικό διαμέρισμα και αστικότητα και συνεπώς προκύπτει ο αριθμός κρουσμάτων "γριπώδους συνδρομής" ανά 1000 επισκέψεις κατά εβδομάδα.
- Στις ΗΠΑ, το αντίστοιχο κέντρο ελέγχου και πρόληψης νοσημάτων - CDC συνεργάζεται με τον Παγκόσμιο Οργανισμό Υγείας - WHO χρησιμοποιώντας 50 διαφορετικά εργαστήρια στην επικράτεια της χώρας. Εκτός από τις εκθέσεις για την επιτήρηση της γρίπης, ο οργανισμός CDC δημοσιεύει αναλυτικές αναφορές για κάθε τύπο γρίπης, αλλά και χάρτες, διαγράμματα και αναλυτικά στατιστικά στοιχεία που αφορούν στην εποχική γρίπη στις ΗΠΑ.

Β'.2 Εναλλακτικές προσεγγίσεις παρακολούθησης της γρίπης

Ακολουθούν οι πιο πρόσφατες και άξιες αναφοράς μέθοδοι επιτήρησης και πρόβλεψης της εξάπλωσης της γρίπης που χρησιμοποιούν δεδομένα από το διαδίκτυο. Επιλέχτηκε να διαχωριστούν σε δύο επιμέρους κατηγορίες, σε αυτήν των διαδικτυακών εφαρμογών και στη δεύτερη των ερευνών που κατέληξαν να δημοσιευτούν σε κάποιο διεθνές συνέδριο μέσω κάποιου άρθρου. Θα πρέπει σε αυτό το σημείο να τονιστεί ότι ο διαχωρισμός αυτός δεν σημαίνει πως οι

διαδικτυακές εφαρμογές που παρουσιάζονται δεν έχουν δημοσιευτεί δημοσιευτεί σε συνέδρια ή έχουν μικρότερη αξία.

Β'.2.1 Διαδικτυακές εφαρμογές

Κάποιες αξιοσημείωτες προσεγγίσεις είναι οι εξής:

- Google Flu Trends (<http://www.google.org/flutrends>)

Η εταιρία Google, μέσω της μηχανής αναζήτησης (διαδικτυακής εφαρμογής που επιτρέπει την αναζήτηση ιστοσελίδων και αρχείων στο διαδίκτυο) που έχει η ίδια αναπτύξει, χρησιμοποίησε τις αναζητήσεις των χρηστών που είναι σχετικές με τη γρίπη και το 2008 λάνσаре την εφαρμογή Google Flu Trends. Όπως είναι αναμενόμενο, περισσότερες αναζητήσεις σχετικές με τη γρίπη, εμφανίζονται κατά τη διάρκεια κάποιας περιόδου έξαρσης της γρίπης. Μετρώντας τη συχνότητα εμφάνισης αυτών των ερωτημάτων αναζήτησης και υπολογίζοντας τον τόπο αναζήτησης μέσω της IP διεύθυνσης των χρηστών, το Google Flu Trends υπολογίζει την έκταση της γρίπης σε διαφορετικές χώρες και περιοχές σε όλο τον κόσμο.

Η ανάλυση των δεδομένων αποδίδει μια στενή σχέση ανάμεσα στον αριθμό των ατόμων που εκτελούν αναζήτηση για θέματα σχετικά με τη γρίπη και στον αριθμό των ατόμων που παρουσιάζουν στην πραγματικότητα συμπτώματα γρίπης. Οι εκτιμήσεις που βασίστηκαν στην Αναζήτηση Google σχετικά με τη γρίπη βρίσκονται πολύ κοντά στους παραδοσιακούς δείκτες δραστηριότητας της γρίπης που προέρχονται από τα κέντρα ελέγχου και πρόληψης νοσημάτων (CDC) των ΗΠΑ.

Τα ερωτήματα αναζητήσεων που αποθηκεύονται και επεξεργάζονται από τη βάση δεδομένων του Google Flu Trends, αποτελούνται από ακριβείς ακολουθίες όρων σχετικών με τη γρίπη, χωρίς να λαμβάνονται υπ' όψη γλωσσικές παραλλαγές, συνώνυμα, διαγλωσσικές μεταφράσεις και ορθογραφικά λάθη. Η βάση δεδομένων αναζητήσεων περιέχει περισσότερα από 50 εκατομμύρια αναζητήσεις οι οποίες πρώτα αποθηκεύονται και στη συνέχεια φιλτράρονται για να αποδώσουν τα τελικά αποτελέσματα.

- Mappyhealth: (<http://mappyhealth.com>)

Το Mappyhealth είναι μια εφαρμογή που εξορύσσει δεδομένα από το κοινωνικό δίκτυο τвиттер και μελετά τις τάσεις των χρηστών να κάνουν αναρτήσεις σχετικές με θέματα υγείας. Βασισμένη στην υπόθεση ότι τα δεδομένα των ιστοσελίδων κοινωνικής δικτύωσης

μπορούν να προβλέψουν την εξάπλωση ασθενειών, η εφαρμογή παρακολουθεί Tweets χρηστών που έχουν σχέση με όρους ασθένειας. Τα Tweets υπόκεινται σε επεξεργασία και έπειτα οπτικοποιούνται με τεχνικές ψηφιακής χαρτογράφησης (Mapping) και αναλυτικών προβολών.

Η εφαρμογή Mappyhealth εξορύσσει σε πραγματικό χρόνο από το Twitter (μέσω του API ροής του Twitter) που περιέχουν 234 διαφορετικούς όρους που σχετίζονται με διάφορες ασθένειες. Κάθε Tweet που περιέχει κάποιον/ους από τους όρους αυτούς, επεξεργάζεται ώστε να αποφασιστεί αν συνδέεται με κάποια από τις 29 ασθένειες που επιτηρούνται από την εφαρμογή. Όταν η ανάλυση ολοκληρώνεται, τα δεδομένα δημοσιεύονται στην ιστοσελίδα που φιλοξενεί την εφαρμογή. Τα Tweets που συλλέγονται ανήκουν μόνο στην αγγλική γλώσσα. Για να γίνει η οπτικοποίηση των δεδομένων σε ψηφιακό χάρτη λαμβάνεται υπ' όψη η τοποθεσία που έχει δηλώσει ο χρήστης στο προφίλ του στην ιστοσελίδα του Twitter (User Profile Location).

Ανάμεσα στις ασθένειες που παρακολουθεί το Mappyhealth είναι και η εποχιακή γρίπη. Τα στοιχεία της εφαρμογής διατίθενται δωρεάν προς όλους τους χρήστες της ιστοσελίδας.

- Flutracking: (<http://www.flutracking.net>)

Το Flutracking είναι ένα Online σύστημα παρακολούθησης υγείας ασθενών και πιο συγκεκριμένα ανίχνευσης επιδημιών της γρίπης. Το Flutracking αναζητά ανθρώπους που ζουν στην ήπειρο της Αυστραλίας και έχουν πρόσβαση σε λογαριασμό ηλεκτρονικού ταχυδρομείου σε εβδομαδιαία βάση.

Το σύστημα ζητά από τους χρήστες να συμπληρώνουν μια διαδικτυακή φόρμα. Εβδομαδιαία, οι χρήστες δίνουν πληροφορίες σχετικά με συμπτώματα γρίπης που εμφάνισαν οι ίδιοι ή κάποιος από το οικογενειακό τους περιβάλλον. Οι πληροφορίες αυτές, αποστέλλονται μέσω ηλεκτρονικού ταχυδρομείου στο σύστημα Flutracking, το οποίο έπειτα από ανάλυση προσφέρει ένα τρόπο ανίχνευσης τόσο για την εποχική, όσο και για την πανδημική γρίπη αλλά και άλλες ασθένειες. Η φόρμα που συμπληρώνει ο χρήστης περιέχει πληροφορίες όπως την ημερομηνία γέννησης του ασθενούς, τα συμπτώματα που έχει παρουσιάσει, την φαρμακευτική αγωγή που έχει ακολουθήσει, τον ταχυδρομικό κώδικα της κατοικίας του και άλλα.

Σε 6 χρόνια λειτουργίας, το Flutracking καταμετράει τουλάχιστον 875.500 εβδομαδιαίες αναφορές προερχόμενες από το λιγότερο 10.000 χρήστες. Η ιστοσελίδα της έρευνας

παρουσιάζει διαγράμματα, στατιστικά στοιχεία και συμπεράσματα που προκύπτουν από τα προαναφερθέντα δεδομένα.

- FluDetector (<http://geopatterns.enm.bris.ac.uk/epidemics>)

Το σύστημα FluDetector αναπτύχθηκε στα πλαίσια της εκπόνησης του διδακτορικού διπλώματος του κ.Βασίλειου Λάμπρου στο Πανεπιστήμιο του Bristol του Ηνωμένου Βασιλείου. Έχει ως αντικείμενο την παρακολούθηση επιδημιών γρίπης παίρνοντας δεδομένα από το κοινωνικό δίκτυο Twitter. Στόχος της εφαρμογής είναι η αυτόματη ανάλυση των αδόμητων κειμένων των Tweets των χρηστών που έχουν σχέση με τη γρίπη, χρησιμοποιώντας μεθόδους στατιστικής ανάλυσης και μηχανικής μάθησης. Το σύστημα εξάγει ένα δείκτη εκτίμησης της γρίπης τον οποίο οπτικοποιεί για τους 6 πιο πρόσφατους μήνες σε ένα στατιστικό διάγραμμα. Η εφαρμογή παρακολουθεί Tweets που προέρχονται μόνο από περιοχές του Ηνωμένου Βασιλείου.

- Germtracker: (<http://fount.in/m>)

Η εφαρμογή Germtracker αναπτύχθηκε από ερευνητική ομάδα του πανεπιστημίου το Rochester των ΗΠΑ. Όπως στις προαναφερθείσες εφαρμογές παρακολουθεί τη γρίπη χρησιμοποιώντας Tweets που περιέχουν λέξεις σχετικές με τη γρίπη και ταυτόχρονα παρέχουν πληροφορίες για την γεωγραφική τοποθεσία απ' όπου δημιουργήθηκαν. Το σύστημα χρησιμοποιεί τεχνικές μηχανικής μάθησης και μοντελοποίησης για την εξαγωγή αποτελεσμάτων. Επίσης ερευνά το αν υπάρχει συσχέτιση της γρίπης με τα επίπεδα μόλυνσης των περιοχών από τις οποίες προέρχονται τα Tweets. Διαφέρει από τα προηγούμενα συστήματα στο ότι τα δεδομένα που οπτικοποιεί δεν παρέχουν ανωνυμία, καθώς ο χρήστης της εφαρμογής μπορεί να δει σε κάθε μεμονωμένο Tweet που οπτικοποιείται, το όνομα και τη φωτογραφία του προφίλ του χρήστη που το έγραψε.

Β'.2.2 Άρθρα βασισμένα στις ιστοσελίδες κοινωνικής δικτύωσης

- Το 2010, ο Aron Culotta, στο δημοσιευμένο άρθρο του με τίτλο "Towards detecting influenza epidemics by analyzing Twitter messages", ερευνά διάφορα 10 διαφορετικά μοντέλα ανάλυσης μηνυμάτων του Twitter με σκοπό την πρόβλεψη των ποσοστών γρίπης σε συγκεκριμένους πληθυσμούς ανθρώπων. Στην έρευνά αναλύθηκαν 574.643 Tweets προερχόμενα από μια χρονική περίοδο 8 μηνών, οδηγώντας στο ότι η προτεινόμενη εργασία, επεξεργάζοντας ένα μικρό αριθμό Tweets μπορεί να προβλέψει με μεγάλη

ακρίβεια την εξάπλωση της γρίπης. Τα ποσοστά συσχέτισης του προτεινόμενου συστήματος και των εθνικών στατιστικών υγείας αγγίζουν το 95

- Το 2011, οι E.Aramaki, S.Maskawa, και M.Morita ασχολήθηκαν με την ανίχνευση επιδημιών γρίπης σε μεγάλη κλίμακα και πραγματικό χρόνο. Τα δεδομένα που χρησιμοποιήσαν ήταν Tweets τα οποία περιείχαν τη λέξη influenza/flu. Ένα σύνολο 5000 δεδομένων επεξεργάστηκε με αλγορίθμους μηχανικής μάθησης.

Το κάθε Tweet αναλυθηκε από έναν ταξινομητή ώστε να του δοθεί μια θετική ή αρνητική ετικέτα. Τα Tweets που περιείχαν τον όρο influenza/flu αλλά δεν χρησιμοποιούνταν για να δηλώσουν εκδήλωση ασθένειας χαρακτηρίζονταν ως αρνητικά. Για τα αρνητικά αυτά Tweets δημιουργήθηκε εκ νέου ένας ταξινομητής που βασιζόμενος στη μηχανική μάθηση έκανε φιλτράρισμα.

Η παραπάνω έρευνα πέτυχε μεγαλύτερα ποσοστά συσχέτισης με τις εθνικές υπηρεσίες ελέγχου και πρόληψης νοσημάτων σε σχέση με το Google Flu Trends για το χρονικό διάστημα που έγινε η έρευνα.

Κεφάλαιο Γ΄

Κύριο Μέρος

Γ΄.1 Σχεδιασμός

Γ΄.1.1 Καθορισμός απαιτήσεων

Πριν αρχίσει η εκπόνηση της παρούσας πτυχιακής εργασίας, τέθηκαν κάποιες βασικές απαιτήσεις, τις οποίες έπρεπε να καλύπτει η εφαρμογή που θα αναπτυσσόταν. Αναφορικά, οι αρχικοί αντικειμενικοί στόχοι προς επίτευξη είναι οι εξής:

Δημιουργία διαδικτυακής εφαρμογής η οποία:

- Εξορύσσει δεδομένα από ένα ή περισσότερα κοινωνικά δίκτυα που προσφέρουν κάποια διαπεαφή προγραμματισμού εφαρμογών.
- Εξορύσσει, επεξεργάζεται και οπτικοποιεί δεδομένα αυτόματοποιημένα.
- Εξορύσσει, επεξεργάζεται και οπτικοποιεί δεδομένα σε πραγματικό χρόνο.
- Αποθηκεύει τα επεξεργασμένα δεδομένα και τα κρατάει ως ιστορικό σε βάση δεδομένων.
- Παρέχει ένα είδος υπηρεσίας παροχής επεξεργασμένων (μια απλοποιημένη μορφή διεπαφής προγραμματισμού εφαρμογών).
- Περιέχει ένα είδος διεπαφής με την οποία μπορεί να διαδράσει ο χρήστης.
- Περιέχει μεθόδους αυτόματης γλωσσικής επεξεργασίας-ανάλυσης κειμένου.

Το σύνολο των παραπάνω απαιτήσεων αποτέλεσε τη βάση πάνω στην οποία στηρίχτηκε ο σχεδιασμός του συστήματος που αναπτύχθηκε.

Γ'.1.2 Καθορισμός των αναγκών του χρήστη

Η πρόκληση στην οποία καλείται να ανταποκριθεί ο διευθυντής του έργου είναι να μεταβάλλει την ανάγκη του χρήστη από κάτι αόριστο σε κάτι συγκεκριμένο, πάνω στο οποίο θα μπορεί να στηριχθεί ο σχεδιασμός του έργου. Ο καθορισμός απαιτήσεων μπορεί να ταυτίζεται με τον καθορισμό των αναγκών του χρήστη, μπορεί όμως και να διαφέρει από τη στιγμή που ο τελικός χρήστης δεν είναι αυτός που έθεσε τις αρχικές απαιτήσεις. Διαχωρίζοντας λοιπόν τις ανάγκες του χρήστη από τις απαιτήσεις της υπό ανάπτυξης εφαρμογής, δημιουργούνται κάποιοι επιπλέον αντικειμενικοί στόχοι όπως οι ακόλουθοι:

- Να σχεδιαστεί μια υπηρεσία, η οποία θα είναι σε θέση να εκτελεί κάποια συγκεκριμένη λειτουργία με προκαθορισμένο ρυθμό.
- Η υπηρεσία πρέπει να είναι σε θέση να λειτουργεί σε συγκεκριμένο περιβάλλον.
- Το σύστημα πρέπει να έχει προκαθορισμένη διάρκεια ζωής.
- Το έργο πρέπει να ανταποκρίνεται στα σύγχρονα πρότυπα τεχνολογίας.
- Η εργονομία θα πρέπει να είναι συμβατή με τις πλέον πρόσφατες κοινές πρακτικές.
- Οι απαιτήσεις λειτουργίας θα πρέπει να ικανοποιούνται με ένα προκαθορισμένο επίπεδο αυτοματοποίησης.
- Το σύστημα θα πρέπει να έχει προδιαγραφές που να επηρεάζουν τη μελλοντική του επέκταση.

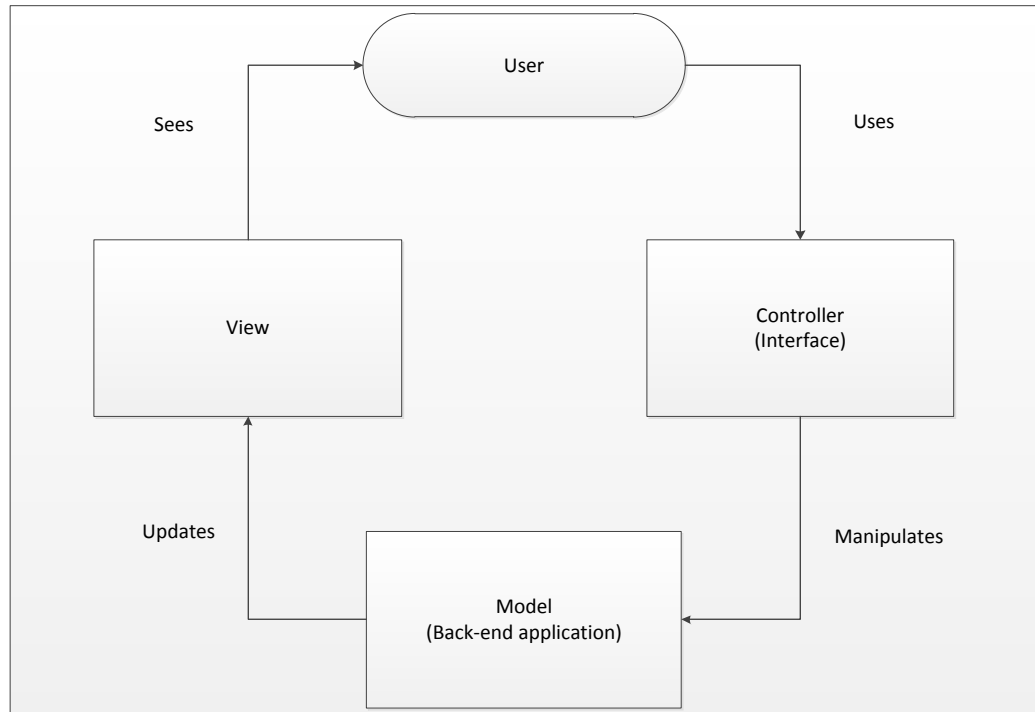
Γ'.1.3 Μεθοδολογία

Για την ανάπτυξη του παρόντος συστήματος, η μεθοδολογία της στρατηγικής σχεδιασμού του έργου ήταν να μετουσιωθούν οι απαιτήσεις που τέθηκαν κατά τη διαδικασίας καθορισμού απαιτήσεων σε ένα ολοκλήρωμένο έργο πληροφορικής. Τουτέστιν, η γενική μεθοδολογία σχεδιασμού του έργου ήταν το να αναλυθεί μια μελέτη περίπτωσης και να αναπτυχθεί ένα πρωτότυπο ώστε να αποδεικνύεται η ορθή λειτουργία και τη χρησιμότητά του.

Ιδέες, ανάγκες και προβλήματα αποκρυσταλλώνονται σε έργα με διαφορετικούς τρόπους. Η διαδικασία σχεδίασης, διαμόρφωσης και τελικά υλοποίησης διαφέρει ανάλογα με τον τύπο του έργου. Σύμφωνα με την προσέγγιση της διαχείρισης έργου, το αντικείμενο εργασιών υποδιαιρείται σε μικρότερες μονάδες, γεγονός που αυξάνει το επίπεδο ανάλυσης και τη δυνατότητα ελέγχου του έργου. Η μεθοδολογία αυτή εφαρμόστηκε από τη σύλληψη του έργου και τη σχεδίασή της δομής του, μέχρι την υλοποίηση και τη συντήρησή του.

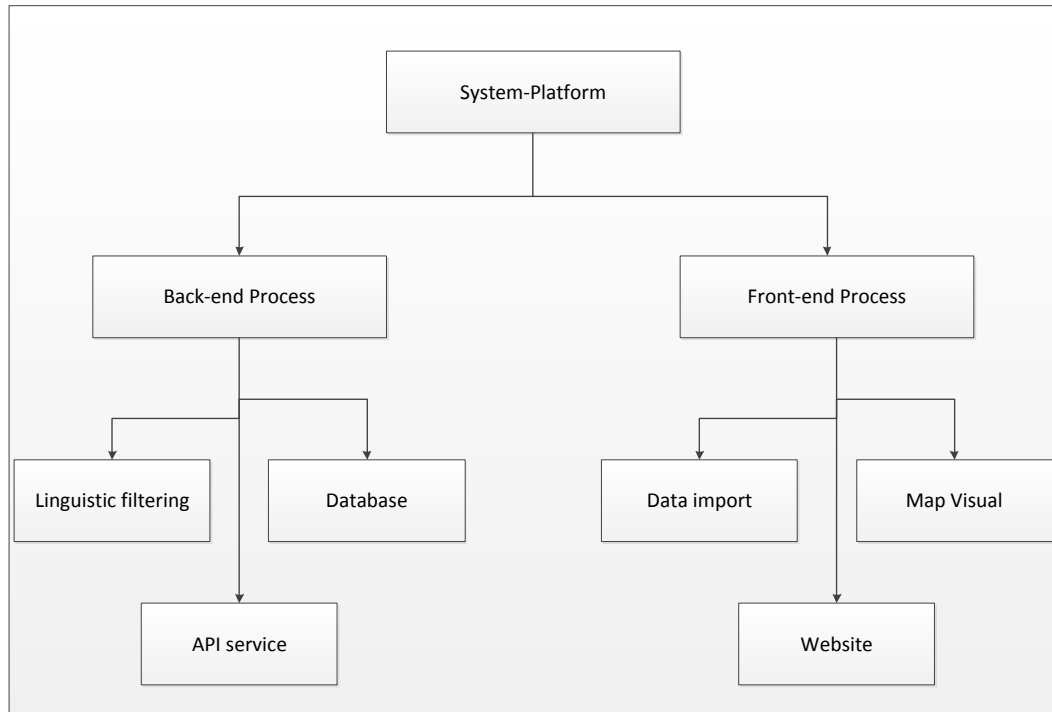
Γ'.1.4 Model View Controller (MVC)

Το MVC είναι ένα πρότυπο αρχιτεκτονικής λογισμικού που χωρίζει την αναπαράσταση των πληροφοριών από την αλληλεπίδραση του χρήστη με αυτή. Το μοντέλο αποτελείται από τα δεδομένα της εφαρμογής, τους κανόνες, τη λογική και τις λειτουργίες. Μια προβολή του MVC μπορεί να είναι οποιαδήποτε αναπαράσταση εξόδου δεδομένων, όπως ένα γράφημα ή ένα διάγραμμα. Αν και αρχικά αναπτύχθηκε για τη χρήση σε προσωπικούς υπολογιστές, το Μοδελ Ίεω Θντρολλερ έχει πλέον υιοθετηθεί ως μια αρχιτεκτονική για τον Παγκόσμιο Ιστό σε όλες τις βασικές γλώσσες προγραμματισμού. Αρκετές εμπορικές και μη εφαρμογές έχουν δημιουργηθεί με βάση αυτό το πρότυπο. Στο βρόχο αυτό, όπως απεικονίζεται παραστατικά στο σχήμα Γ'.1 το πως ο χρήστης της εφαρμογής χρησιμοποιεί τη διεπαφή για να επικοινωνήσει με το Back-end κομμάτι του συστήματος. Επιπλέον, φαίνεται ξεκάθαρα ότι στη συνέχεια ανανεώνεται το Front-end κομμάτι, το οποίο και τελικά φτάνει στο μάτι του χρήστη. Η διαδικασία ξεκινάει από την αρχή από τη στιγμή που ο χρήστης ζητήσει κάτι καινούριο από τη διεπαφή.

Σχήμα Γ'.1: *Model View Controller* διάγραμμα

Γ'.1.5 Δομική ανάλυση του έργου

Ο ρόλος της δομικής ανάλυσης έργου (WBS, Work Breakdown Structure) είναι να υποδιαιρεί το αντικείμενο εργασιών σε πακέτα εργασιών. Η δομική ανάλυση του έργου αποτελεί εξαιρετικό εργαλείο για την ποσοτική αποτίμηση του αντικειμένου των εργασιών. Μπορεί επίσης να θεωρηθεί, ως ένας ιεραρχημένος νοητικός χάρτης που επιτρέπει την ανάλυση της πολυπλοκότητας του έργου. Με άλλα λόγια, ο αντικειμενικός στόχος του έργου αναλύεται σε σύνολα μερών και συστατικά μέρη, κάτι πολύ σημαντικό για τον αρχικό σχεδιασμό του συστήματος. Διαγραμματικά, η δομική ανάλυση του προς ανάπτυξη έργου είναι η εξής:



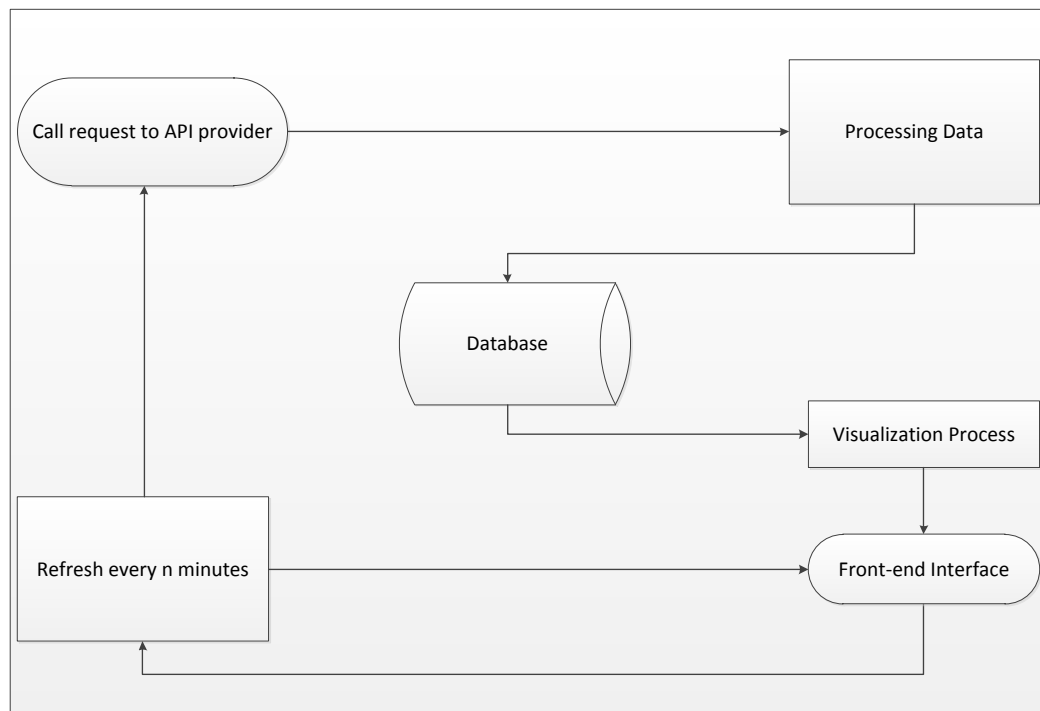
Σχήμα Γ'.2: *Work Breakdown Structure* διάγραμμα

Σύμφωνα με το παραπάνω διάγραμμα Γ'.2, φαίνεται ότι ο σχεδιασμός υποδιαιρείται αρχικά στο σχεδιασμό του Back-end και του Front-end μέρους της εφαρμογής. Ο Back-end σχεδιασμός διαιρείται στη συνέχεια στις υποδιεργασίες της γλωσσικής ανάλυσης (Linguistic filtering), της δημιουργίας και διαχείρισης της βάσης δεδομένων και της ανάπτυξης μιας διεπαφής προγραμματισμού εφαρμογών (API). Το Front-end υπομέρος, υποδιαιρείται με τη σειρά του στο κομμάτι της εισαγωγής δεδομένων από τη βάση δεδομένων της Back-end διεργασίας, στην οπτικοποίηση των δεδομένων αυτών σε χάρτη και τέλος στην ανάπτυξη της ιστοσελίδας όπου θα μπορεί να τα παρακολουθήσει ο τελικός χρήστης.

Γ'.1.6 Εννοιολογικό Μοντέλο (Conceptual Model)

Ένα εννοιολογικό μοντέλο αντιπροσωπεύει τις έννοιες των οντοτήτων και τις μεταξύ τους σχέσεις. Το εννοιολογικό μοντέλο δεν συγχέεται με άλλους τομείς όπως η μοντελοποίηση

δεδομένων, η λογική και η φυσική μοντελοποίηση. Ο στόχος του εννοιολογικού μοντέλου είναι να εκφράσει την έννοια των όρων και των εννοιών που χρησιμοποιούνται στο υπό ανάπτυξη σύστημα και να παρουσιάσει με ακρίβεια τις σχέσεις μεταξύ διαφορετικών εννοιών. Ένα τέτοιο σύνολο εννοιών μπορεί να αναπαρασταθεί με ποικίλα είδη διαγραμμάτων. Μια προσέγγιση είναι η παρακάτω:



Σχήμα Γ'.3: *Conceptual Model* διάγραμμα

Το διάγραμμα Γ'.3 στηρίζεται στην υλοποίηση των απαιτήσεων που καθορίστηκαν και συνεπώς ξεκινά με την κλήση στη Διεπαφή Προγραμματισμού Εφαρμογών (API) κάποιου παρόχου δεδομένων. Από τη στιγμή που ο πάροχος επιστρέφει ένα σύνολο δεδομένων, το σύνολο αυτό υποβάλλεται σε αυτόματη γλωσσική ανάλυση και φιλτράρεται ανάλογα με τις εκάστοτε συγκεκριμένες ανάγκες του συστήματος. Τα φιλτραρισμένα δεδομένα αποθηκεύονται στη βάση δεδομένων του συστήματος και στη συνέχεια εξάγονται, για να οπτικοποιηθούν στο Front-end μέρος του συστήματος (ιστοσελίδα ή εφαρμογή για 'έξυπνο' τηλέφωνο), ώστε να είναι ορατά από τον τελικό χρήστη. Η διαδικασία επαναλαμβάνεται ατέρμονα ανά n λεπτά, όπου

το ν ορίζεται ανάλογα με το πόσο συχνά προσφέρει δεδομένα η Διεπαφή Προγραμματισμού Εφαρμογών του παρόχου.

Γ'.1.7 Απαιτήσεις σχεδιασμού

Σύμφωνα με το εννοιολογικό μοντέλο που παρουσιάστηκε, θέτονται αυτόματα νέες απαιτήσεις για τις τεχνολογίες και τα εργαλεία που πρέπει να χρησιμοποιηθούν για την ανάπτυξη του συστήματος. Στο επίπεδο του σχεδιασμού δεν μελετάται η επιλογή συγκεκριμένων εργαλείων και γλωσσών προγραμματισμού. Αναλύονται οι ανάγκες του έργου και αποφασίζεται η στρόφη σε λύσεις που ικανοποιούν τις ανάγκες αυτές σε μεγαλύτερο βαθμό. Οι απαιτήσεις για την υλοποίηση του εννοιολογικού διαγράμματος είναι οι εξής:

- Επιλογή γλώσσας προγραμματισμού η οποία υποστηρίζεται από τη Διεπαφή Προγραμματισμού Εφαρμογών (API) και επιτρέπει κλήσεις προς αυτή.
- Ανάπτυξη του Back-end συστήματος σε γλώσσα προγραμματισμού βασισμένη σε διακομιστή καθώς η εφαρμογή θα είναι διαδικτυακή. Η επιλεγμένη γλώσσα προγραμματισμού θα πρέπει επίσης να χρησιμοποιείται για σκοπούς γλωσσικής ανάλυσης.
- Αν είναι δυνατό, επιλογή παρόμοιας γλώσσας προγραμματισμού για τη Διεπαφή Προγραμματισμού Εφαρμογών και το συνολικό Back-end σύστημα.
- Επιλογή και δημιουργία συστήματος διαχείρισης σχεσιακών βάσεων δεδομένων που να ευνοεί την αποθήκευση συγκεκριμένου τύπου δεδομένων.
- Επιλογή κατάλληλης γλώσσας προγραμματισμού για οπτικοποίηση δεδομένων.
- Δημιουργία ιστοσελίδας που θα φιλοξενεί το σύστημα και θα προβάλλει τις απαραίτητες πληροφορίες.

Γ'.2 Υλοποίηση

Γ'.2.1 Το πρωτότυπο Flutrack

Το πρωτότυπο σύστημα που αναπτύχθηκε έχει την ονομασία Flutrack και ψάχνει, φιλτράρει και οπτικοποιεί συμπτώματα εποχικής γρίπης σε παγκόσμια κλίμακα.. Ακολουθώντας τις απαιτήσεις που καθορίστηκαν, η πλατφόρμα Flutrack εξάγει δεδομένα από το κοινωνικό δίκτυο

Twitter, τα οποία φιλτράρει χρησιμοποιώντας μεθόδους γλωσσικής ανάλυσης και στη συνέχεια αποθηκεύει σε μια βάση δεδομένων. Τα δεδομένα οπτικοποιούνται σε ψηφιακό χάρτη, ο οποίος φιλοξενείται στην ιστοσελίδα (<http://flutrack.org>) μαζί με πληροφορίες σχετικές για το σύστημα. Το σύστημα Flutrack διαφέρει από τα υπόλοιπα παρόμοια συστήματα στο ότι ψάχνει, φιλτράρει, αποθηκεύει, οπτικοποιεί και επαναδιαθέτει δεδομένα σε πραγματικό χρόνο και επίσης παρακολουθεί τυχούσα επιδείνωση της κλινικής κατάστασης του κάθε χρήστη. Η επαναδιάθεση των δεδομένων γίνεται μέσω από τη Διεπαφή Προγραμματισμού Εφαρμογών (API) του συστήματος.

Γ'.2.2 Ετικέτες (Tags) παρακολούθησης

Οι λέξεις κλειδιά για τις οποίες αναζητούνται Tweets είναι οι εξής:

- flu
- influenza
- chills
- headache
- sore throat
- runny nose
- sneezing
- fever
- dry cough

Οι λέξεις αυτές αναζητούνται συνολικά σε ένα ερώτημα (Query) ανά κλήση χρησιμοποιώντας το λογικό τελεστή OR. Έτσι, σύμφωνα με τις πράξεις τελεστών, αν κάποιο Tweet περιέχει έστω και μια από τις παραπάνω ετικέτες επιστρέφεται από τη Διεπαφή Προγραμματισμού Εφαρμογών (API) του Twitter. Λέξεις όπως diarrhea (διάρροια) και vomit (εμετός), θεωρήθηκε ορθό να μην χρησιμοποιούνται στο ερώτημα, καθώς είναι πολύ σύνηθες, να εμφανίζονται τα συμπτώματα αυτά, σε ασθένειες που δεν οφείλονται στην εποχική γρίπη, αλλά σε τροφικές δηλητηριάσεις, γαστρεντερίτιδα κτλ.

Γ'.2.3 Αρχιτεκτονική συστήματος Flutrack

Η αρχιτεκτονική της εφαρμογής Flutrack επιλέχθηκε με κριτήριο τον σχεδιασμό και τις απαιτήσεις του. Για την ανάλυση της αρχιτεκτονικής του συστήματος, η παρουσίαση διαχωρίζεται σε 2 μέρη, την αρχιτεκτονική του Back-end και την αρχιτεκτονική του Front-end τμήματος. Τα δύο αυτά τμήματα είναι άρρηκτα συνδεδεμένα μεταξύ τους και υλοποιήθηκαν ταυτόχρονα. Αρχικά παρουσιάζεται το Back-end μέρος που αποτελεί τον πυρήνα της εφαρμογής και έπειτα το Front-end που ουσιαστικά οδηγεί στην προβολή των δεδομένων στον τελικό χρήστη. Πριν την ανάλυση των τμημάτων, παρουσιάζονται τα εργαλεία και οι υπηρεσίες που χρησιμοποιήθηκαν για την υλοποίηση του συστήματος.

Γ'.2.4 Γλώσσες προγραμματισμού και εργαλεία

Γ'.2.4.1 PHP

Για την υλοποίηση του Back-end τμήματος επιλέχθηκε η γλώσσα προγραμματισμού PHP. Η PHP είναι μια γλώσσα προγραμματισμού για τη δημιουργία σελίδων Web με δυναμικό περιεχόμενο. Αποτελεί επίσης μια ισχυρή γλώσσα για τη συγγραφή σεναρίων και συνεργάζεται πολύ καλά με τη γλώσσα HTML. Ένα από τα σημαντικότερα χαρακτηριστικά της PHP είναι ότι είναι μια γλώσσα ανοιχτού κώδικα και διανέμεται ελεύθερα, το οποίο σημαίνει ότι κάθε υλικό αναφοράς της μπορεί να εντοπισθεί εύκολα και να χρησιμοποιηθεί από τον καθένα. Επίσης μπορεί να χρησιμοποιηθεί με ευκολία σε διάφορα λειτουργικά συστήματα. Για τον έλεγχο των προγραμμάτων PHP, απαιτείται ένας εξυπηρετητής που υποστηρίζει τη γλώσσα.

Η PHP μπορεί να εγκατασταθεί σε έναν εξυπηρετητή Ιστού σαν ενσωματωμένο επιπλέον λογισμικό (module), είτε σαν αυτόνομος εκτελέσιμος κώδικας. Προτιμάται η πρώτη περίπτωση, αφού έτσι είναι γρηγορότερη η εκτέλεση και δεν επιβαρύνεται ο εξυπηρετητής με επιπλέον απαιτήσεις σε μνήμη και χώρο στο δίσκο. Η γλώσσα PHP είναι μια από τις γλώσσες προγραμματισμού που υποστηρίζει το API του Twitter και καλύπτει τις σχεδιαστικές αρχές που τέθηκαν στο προηγούμενο κεφάλαιο. Επιπλέον, εκτός από τις υπόλοιπες δυνατότητες που έχει η γλώσσα PHP, προσφέρει εύκολη δημιουργία και επεξεργασία κανονικών εκφράσεων (Regular Expressions), γεγονός που επιτρέπει ευκολότερο και γρήγορο φιλτράρισμα δεδομένων κειμένου. Στο σύστημα Flutrack χρησιμοποιείται για να καταστεί δυνατή η κλήση στο (API) του Twitter, βοηθάει στη σύνδεση του Back-end τμήματος με τη βάση δεδομένων του Φλυτραςχ και στο φιλτράρισμα των δεδομένων πριν την αποθήκευσή του.

Γ'.2.4.2 CURL

Το CURL είναι ένα εργαλείο γραμμής εντολών για μεταφορά δεδομένων χρησιμοποιώντας διάφορα πρωτόκολλα επικοινωνίας. Το CURL κυκλοφόρησε με το libCURL, το οποίο είναι μια βιβλιοθήκη για το CURL, η οποία μπορεί να χρησιμοποιηθεί σε πολλές γλώσσες προγραμματισμού. Η πιο διαδεδομένη της μορφή είναι αυτή στην PHP, με τη βιβλιοθήκη ΠΗΠ CURL. Τόσο το CURL όσο και η επέκταση της PHP, libCURL, είναι εργαλεία που μπορούν να χρησιμοποιηθούν για την προσομοίωση ενός φυλλομετρητή διαδικτύου. Με το CURL μπορεί ακόμη να γίνει απόσπαση πληροφοριών από ιστοσελίδες. Μπορεί να ανακτηθεί κώδικας HTML, να φιλτραριστούν και να ανακτηθούν τμήματα μιας ιστοσελίδας ή και πληροφορίες. Για παράδειγμα μπορούν να ανακτηθούν πληροφορίες από τα αποτελέσματα αναζήτησης μηχανών αναζήτησης όπως είναι το Google, Yahoo, Bing και άλλες. Στο σύστημα Flutrack, χρησιμοποιούνται εντολές της βιβλιοθήκης libCURL για τη μεταφορά δεδομένων από το διακομιστή της υπηρεσίας Twitter.

Γ'.2.4.3 SQL

Η SQL (Structured Query Language) είναι μία γλώσσα υπολογιστών στις βάσεις δεδομένων, που σχεδιάστηκε για τη διαχείριση δεδομένων, σε ένα σύστημα διαχείρισης σχεσιακών βάσεων δεδομένων (Relational Database Management System, RDBMS) και η οποία, αρχικά, βασίστηκε στη σχεσιακή άλγεβρα. Η γλώσσα περιλαμβάνει δυνατότητες ανάκτησης και ενημέρωσης δεδομένων, δημιουργίας και τροποποίησης σχημάτων και σχεσιακών πινάκων, αλλά και ελέγχου πρόσβασης στα δεδομένα. Συνδυάζοντας τη γλώσσα προγραμματισμού PHP και την SQL φιλτράρονται και αποθηκεύονται τα δεδομένα στη βάση δεδομένων του Flutrack. Επίσης, μετά την αποθήκευση των δεδομένων, η PHP συνδυάζεται ξανά με την SQL για να εξάγουν εκ νέου τα δεδομένα ώστε να οπτικοποιηθούν.

Γ'.2.4.4 Javascript

Η JavaScript (JS) είναι διερμηνευμένη γλώσσα προγραμματισμού για ηλεκτρονικούς υπολογιστές. Αρχικά αποτέλεσε μέρος της υλοποίησης των φυλλομετρητών Ιστού, ώστε τα σενάρια από την πλευρά του πελάτη (Client-side scripts) να μπορούν να επικοινωνούν με τον χρήστη, να ανταλλάσσουν δεδομένα ασύγχρονα και να αλλάζουν δυναμικά το περιεχόμενο του εγγράφου που εμφανίζεται. Η Javascript είναι ουσιαστικά μια γλώσσα σεναρίων που βασίζεται στα

πρωτότυπα (prototype-based), είναι δυναμική, με ασθενείς τύπους και έχει συναρτήσεις ως αντικείμενα πρώτης τάξης. Τα Scripts εκτελούνται χωρίς να έχει προηγηθεί μεταγλώττιση του κώδικα. Χρησιμοποιείται ακόμη για να δημιουργήσουμε διαδραστικές (Interactive) ιστοσελίδες. Ο κώδικας Javascript συνήθως ενσωματώνεται στον HTML κώδικα. Στην περίπτωση του συστήματος Flutrack η γλώσσα χρησιμοποιείται στο Front-end τμήμα για την οπτικοποίηση των δεδομένων στην ιστοσελίδα της εφαρμογής. Πιο συγκεκριμένα, χρησιμοποιείται η jQuery, μια ελαφριά βιβλιοθήκη της γλώσσας Javascript, συμβατή με όλους του φυλλομετρητές που κυκλοφορούν. Η jQuery χρησιμοποιείται για τη δημιουργία του ψηφιακού χάρτη της εφαρμογής. Προσθέτει κίνηση, αυξάνει τη διαδραστικότητα του χρήστη και αλλάζει το περιεχόμενο της σελίδας χωρίς ο χρήστης να πρέπει να μεταφερθεί σε νέα σελίδα. Η έκδοση jQuery που εισάγεται για την εκτέλεση αυτών των λειτουργιών είναι η: jquery-1.9.1.min. Η έκδοση αυτή δεν είναι δεσμευτική και είναι πιθανό να αλλάξει κατά τη συντήρηση του συστήματος.

Γ'.2.4.5 HTML

Η HTML (ακρωνύμιο του αγγλικού HyperText Markup Language (Γλώσσα Σήμανσης Υπερκειμένου)) είναι η κύρια γλώσσα σήμανσης για τις ιστοσελίδες, και τα στοιχεία της είναι τα βασικά δομικά στοιχεία των ιστοσελίδων. Η HTML γράφεται υπό μορφή στοιχείων HTML τα οποία αποτελούνται από ετικέτες. Ανάμεσα στις ετικέτες, οι σχεδιαστές ιστοσελίδων μπορούν να τοποθετήσουν κείμενο, πίνακες, εικόνες κλπ. Ο σκοπός ενός Web browser είναι να διαβάσει τα έγγραφα HTML και τα συνθέτει σε σελίδες που μπορεί κανείς να διαβάσει ή να ακούσει. Ο Web browser δεν εμφανίζει τις ετικέτες HTML, αλλά τις χρησιμοποιεί για να ερμηνεύσει το περιεχόμενο της σελίδας. Τα στοιχεία της HTML χρησιμοποιούνται για να κτίσουν όλους του ιστότοπους. Η HTML επιτρέπει την ενσωμάτωση εικόνων και άλλων αντικειμένων μέσα στη σελίδα, και μπορεί να χρησιμοποιηθεί για να εμφανίσει διαδραστικές φόρμες. Παρέχει τις μεθόδους δημιουργίας δομημένων εγγράφων (δηλαδή εγγράφων που αποτελούνται από το περιεχόμενο που μεταφέρουν και από τον κώδικα μορφοποίησης του περιεχομένου) καθορίζοντας δομικά σημαντικά στοιχεία για το κείμενο, όπως κεφαλίδες, παραγράφους, λίστες, συνδέσμους, παραθέσεις και άλλα. Μπορούν επίσης να ενσωματώνονται σενάρια εντολών σε γλώσσες όπως η JavaScript, τα οποία επηρεάζουν τη συμπεριφορά των ιστοσελίδων HTML. Οι Web browsers μπορούν επίσης να αναφέρονται σε στυλ μορφοποίησης CSS για να ορίζουν την εμφάνιση και τη διάταξη του κειμένου και του υπόλοιπου υλικού. Ο οργανισμός W3C, ο οποίος δημιουργεί και συντηρεί τα πρότυπα για την HTML και τα CSS, ενθαρρύνει τη χρήση των CSS αντί διαφόρων στοιχείων της HTML για σκοπούς παρουσίασης του περιεχομένου.

Το σύστημα Flutrack αναπτύχθηκε με HTML5

Γ'.2.4.6 CSS

Η CSS είναι μια απλή γλώσσα προγραμματισμού που μας βοηθάει να ορίσουμε με σαφήνεια και ιδιαίτερη ευελιξία τον τρόπο με τον οποίο θα εμφανίζονται τα διάφορα στοιχεία σε μια ιστοσελίδα. Το CSS κατέστησε εφικτές μορφοποιήσεις οι οποίες ήταν αδύνατες ή πολύ δύσκολες μόνο με τη γλώσσα HTML. Η CSS βοηθάει στην ευκολότερη συντήρηση των ιστοσελίδων. Η εμφάνιση ενός ολόκληρου Website μπορεί να ελέγχεται από ένα μόνο εξωτερικό αρχείο .CSS. Έτσι, κάθε αλλαγή στην οπτική της ιστοσελίδας μπορεί να γίνεται με μια μοναδική αλλαγή σε αυτό το αρχείο, αντί για την επεξεργασία πολλών σημείων σε κάθε σελίδα που υπάρχει στο Website. Με τη CSS έχουμε μικρότερα σε μέγεθος αρχεία, δεδομένου ότι ο κάθε κανόνας μορφοποίησης γράφεται μόνο μια φορά και όχι σε κάθε σημείο που εφαρμόζεται. Εκτός αυτού, βελτιώνει την αναζήτηση στο διαδίκτυο SEO(Search engine optimization). Οι μηχανές αναζήτησης δεν «μπερδεύονται» ανάμεσα σε περιεχόμενο και τη μορφοποίηση του, αλλά έχουν πρόσβαση στο περιεχόμενο σκέτο, οπότε είναι πολύ ευκολότερο να το καταγράψουν και να το αρχειοθετήσουν (Indexing). Οι σελίδες γίνονται γρηγορότερες καθώς όταν χρησιμοποιούμε εξωτερικό αρχείο CSS, ο φυλλομετρητής Browser την πρώτη φορά που θα φορτώσει κάποια σελίδα ενός Website, το αποθηκεύει στην ζαχης, οπότε δεν χρειάζεται να το κατεβάσει ξανά κάθε φορά που κατεβάζει ο χρήστης του κάποια άλλη σελίδα του Website. Η ιστοσελίδα <http://flutrack.org> έχει δημιουργηθεί με την έκδοση CSS3.

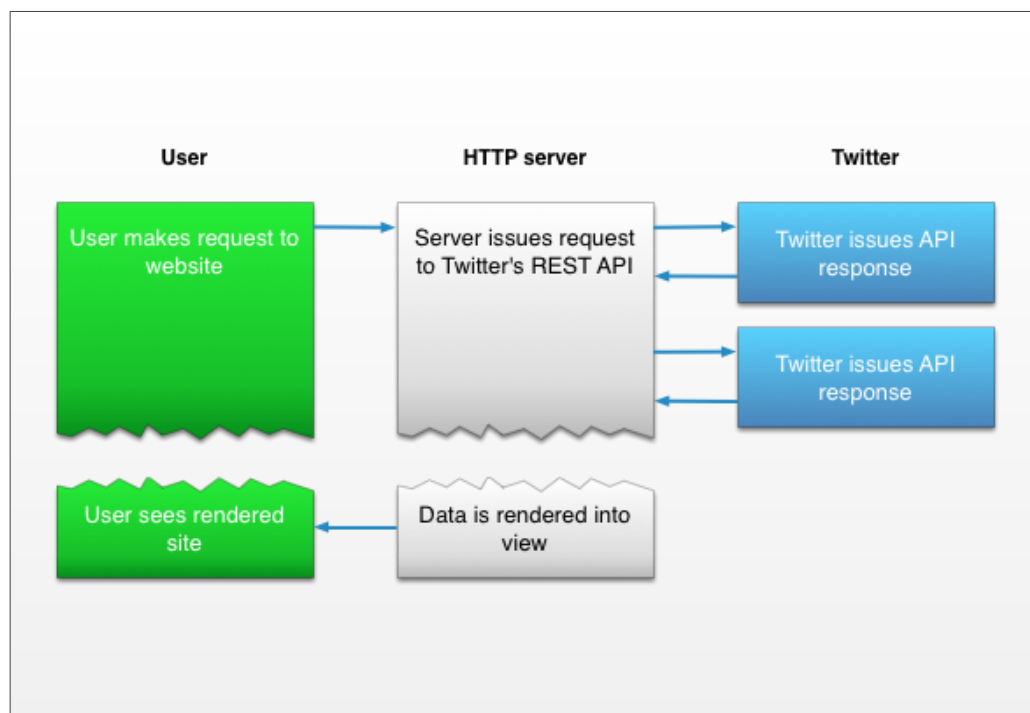
Γ'.2.4.7 Google Maps

Google Maps είναι μια WEB υπηρεσία χαρτογράφησης και παρέχεται από την Google, που τροφοδοτεί πολλές Map-based υπηρεσίες, όπως τους δικτυακούς τόπους Google Maps, Google Ride Finder, Google Transit, και ενσωματωμένους χάρτες σε ιστοσελίδες τρίτων μέσω του Google Maps API. Προσφέρει οδικούς χάρτες, σχεδιασμό διαδρομών με αυτοκίνητο, ποδήλατο (Beta), ή με τα μέσα μαζικής μεταφοράς σε πολλές χώρες σε όλο τον κόσμο. Το Google Maps API καλείται κάθε φορά που κάποιος χρήστης επισκέπτεται την ιστοσελίδα <http://flutrack.org> και οπτικοποιεί τον ψηφιακό χάρτη.

Γ'.2.5 Back-end τμήμα

Γ'.2.5.1 Αίτηση στη Διεπαφή Προγραμματισμού Εφαρμογών

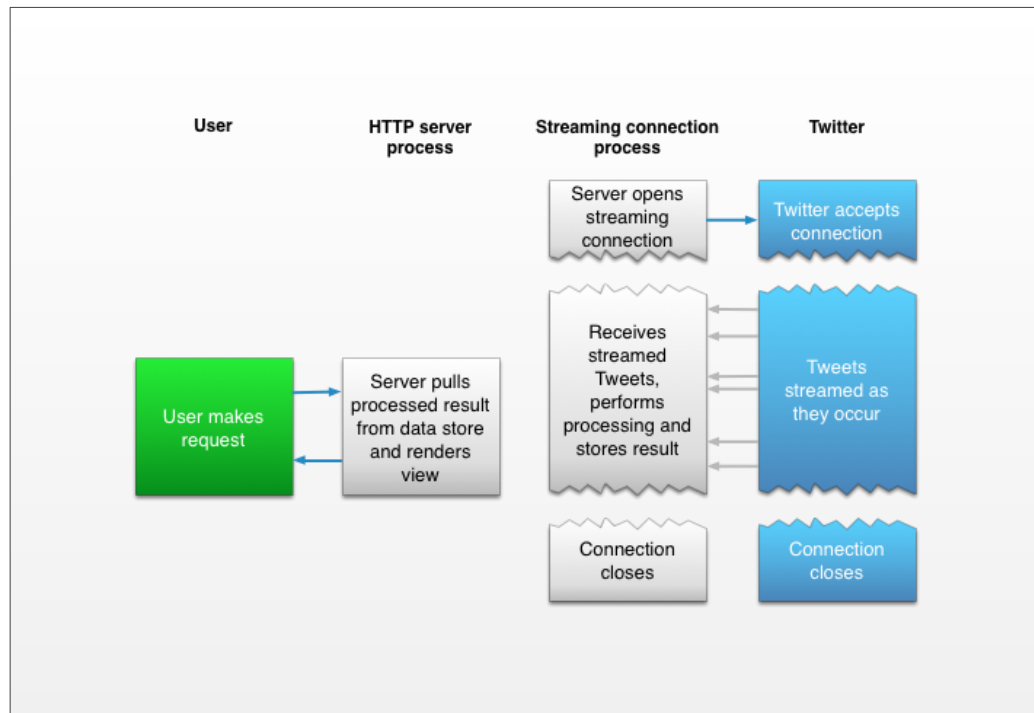
Σύμφωνα με το εννοιολογικό διάγραμμα Γ'.3, το σύστημα ξεκινά τη λειτουργία του καλώντας τη διεύθυνση του διακομιστή της Διεπαφής Προγραμματισμού Εφαρμογών (API) του Twitter, αποστέλλοντας μια αίτηση, η οποία περιέχει τις ετικέτες παρακολούθησης που προαναφέρθηκαν. Στη συνέχεια, η Διεπαφή Προγραμματισμού Εφαρμογών του Twitter αποστέλλει πίσω, τα δεδομένα που ζητήθηκαν, μέσω του διακομιστή. Τα δεδομένα αποστέλλονται μέσω ενός .JSON αρχείου. Η συγκεκριμένη τεχνολογία περιγράφεται στην παράγραφο των εργαλείων υλοποίησης. Το παρακάτω διάγραμμα Γ'.4 απεικονίζει τη διαδικασία της αίτησης και αποστολής δεδομένων.



Σχήμα Γ'.4: διάγραμμα διαδικασίας *Call Request*

Από τη στιγμή που γίνεται το αίτημα στο API του Twitter, εφόσον πληροί τις προϋποθέσεις αυθεντικοποίησης και αυθεντικότητας, γίνεται δεκτό και δημιουργείται ένα κανάλι ροής ανά-

μεσα στην εφαρμογή Φλυτραςκ και το Twitter. Η ροή διακόπτεται και το κανάλι τερματίζει όταν η ροή δεδομένων έχει τελειώσει. Για να δημιουργηθεί εκ νέου κανάλι ροής χρειάζεται ξεχωριστό αίτημα στο διακομιστή του Twitter.



Σχήμα Γ'.5: διάγραμμα διαδικασίας *Streaming*

Γ'.2.5.2 Ρυθμός παρακολούθησης

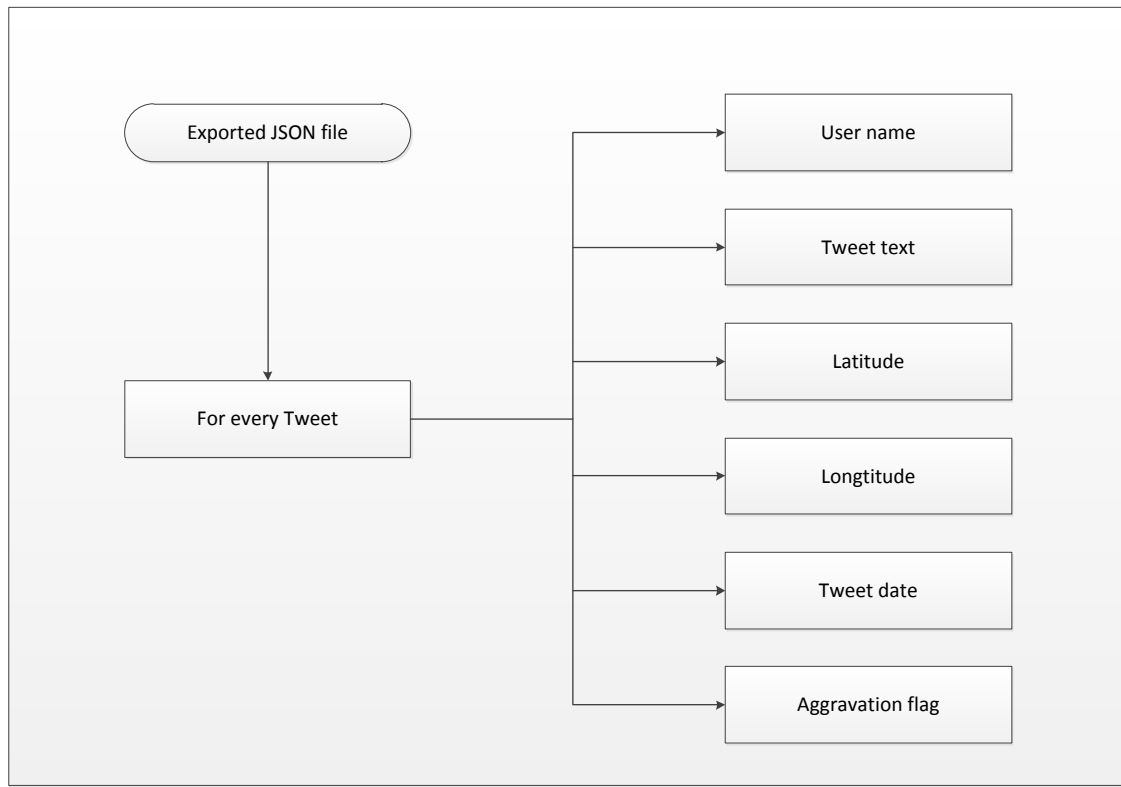
Λόγω του φόρτου εργασίας από χιλιάδες αιτήσεις ανά δευτερόλεπτο, το Twitter επιτρέπει σε κάθε εφαρμογή/χρήστη ένα συγκεκριμένο αριθμό αιτήσεων. Αν οι αιτήσεις στο API του Twitter είναι παραπάνω από αυτές που επιτρέπονται, τα δεδομένα που επιστρέφονται είναι τα ίδια με την τελευταία αίτηση ή δεν επιστρέφονται καθόλου δεδομένα. Συνεπώς θεωρήθηκε σκόπιμο, για το σύστημα Flutrack, να γίνονται αιτήσεις στο API κάθε 20 λεπτά, δηλαδή 5 φορές σε μια ώρα. Ο ρυθμός αυτός είναι ικανοποιητικός ώστε να αναλύονται και να οπτικοποιούνται τα Tweets σε πραγματικό χρόνο, με καθυστέρηση κάποιων λεπτών.

Γ'.2.5.3 Μεταδεδομένα

Για κάθε Tweet που επιστρέφεται και περιέχει κάποιες από τις ετικέτες παρακολούθησης, επιστρέφονται ταυτόχρονα μεταδεδομένα που χρησιμεύουν στην αποθήκευση, ταξινόμηση, αναζήτηση και την οπτικοποίηση των Tweets. Για κάθε Tweet τα μεταδεδομένα που ανασύρρονται φαίνονται στο παρακάτω διάγραμμα Γ'.6. Τα Longitude και Latitude είναι οι γεωγραφικές συντεταγμένες του σημείου από το οποίο ο χρήστης έκανε τη δημοσίευση. Οι χρήστες του Twitter, μπορούν να γνωστοποιήσουν τις συντεταγμένες στις οποίες βρίσκονται, μέσω συστημάτων GPS που περιέχουν τα περισσότερα "εξυπνα τηλέφωνα". Σε αντίθετη περίπτωση, εφόσον οι συντεταγμένες του χρήστη έχουν τιμή NULL, γίνεται εκ νέου αίτηση στο API του Twitter, ώστε να επιστρέψει την τοποθεσία που έχει δηλώσει ο χρήστης στο προφίλ του (Profile location). Η τοποθεσία προφίλ αποτελεί ένα όνομα και όχι έναν συνδυασμό συντεταγμένων ώστε να αναπαρασταθεί σε ψηφιακό χάρτη. Για το λόγο αυτό, γίνεται αίτημα στο API της υπηρεσίας Google Maps ώστε να γίνει η αντιστοίχιση της τοποθεσίας που έχει δηλώσει ο χρήστης, σε συντεταγμένες που μπορούν να προβληθούν σε χάρτη. Η ημερομηνία χρησιμοποιείται για να γνωρίζουμε ακριβώς πότε ο κάθε χρήστης δημοσίευσε την ασθένειά του.

Γ'.2.5.4 Επιδείνωση κλινικής κατάστασης χρήστη

Το μεταδεδομένο Aggravation flag, χρησιμοποιείται για να εξακριβώνεται αν το Tweet προέρχεται από χρήστη που υπάρχει ήδη αποθηκευμένος στη βάση δεδομένων του Flutrack ή όχι. Σε περίπτωση που ο ίδιος χρήστης έχει ξαναδημοσιεύσει Tweet σχετικό με τη γρίπη μέσα σε διάστημα μιας εβδομάδας και το Tweet αυτό περιέχει λέξεις σχετικές με την επιδείνωση της κλινικής κατάστασης του ασθενή, το Aggravation flag παίρνει την τιμή 1, διαφορετικά την τιμή 0. Για την αναζήτηση λέξεων επιδείνωσης, δημιουργήθηκε ένα λεξικό με όρους οι οποίοι θα μπορούσαν να σχετίζονται με συμπτώματα γρίπης ή φράσεις που παραπέμπουν σε ένδειξη επιδείνωσης. Ενδεικτικά, τέτοιες λέξεις είναι: worse, deteriorate, feeling bad again, higher fever κτλ. Το συνολικό λεξικό των όρων επιδείνωσης φαίνεται στον πηγαίο κώδικα της εφαρμογής.



Σχήμα Γ'.6: διάγραμμα μεταδεδομένων

Γ'.2.5.5 Φιλτράρισμα

Πριν αποθηκευτούν τα Tweets με τα μεταδεδομένα τους στη βάση δεδομένων της εφαρμογής, φιλτράρονται μέσω μιας αυτοματοποιημένης διαδικασίας με τη χρήση κανονικών εκφράσεων (Regular Expressions). Στην επιστήμη των υπολογιστών, μια κανονική έκφραση (Regex σε συντομογραφία), είναι μια ακολουθία χαρακτήρων που σχηματίζει ένα μοτίβο αναζήτησης, η οποία χρησιμοποιείται στην εύρεση και αντικατάσταση συμβολοσειρών. Με τον τρόπο αυτό, δίνεται η δυνατότητα αναζήτησης συγκεκριμένων ακολουθιών συμβολοσειρών στο κείμενο του κάθε Tweet, ώστε να αφαιρείται ο θόρυβος. Ως θόρυβος σε ένα Tweet, εννοούνται οι συμβολοσειρές που δεν προσφέρουν κάτι χρήσιμο στον εντοπισμό της γρίπης. Σύμβολα που αφαιρούνται είναι τα εξής:

- Έλεγχος αν το Tweet έχει λιγότερους από 5 χαρακτήρες.
- Έλεγχος στο κάθε Tweet για αν υπάρχουν μη ASCII χαρακτήρες. Οι μη ASCII χαρακτήρες διαγράφονται από το Tweet.
- Έλεγχος και διαγραφή των Hashtags. Ένα Hashtag είναι μια λέξη ή μια φράση, η οποία περιέχει το σύμβολο “.
- Έλεγχος και διαγραφή των Tags. Ένα Tag είναι μια λέξη ή μια φράση, η οποία περιέχει το σύμβολο .
- Έλεγχος και διαγραφή συμβολοσειρών που περιέχουν emoticons.
- Έλεγχος της τοποθεσίας προφίλ χρήστη για περιεχόμενο που δεν αντιστοιχεί σε πραγματική τοποθεσία. Σε περίπτωση που η τοποθεσία προφίλ περιέχει μη πραγματική τοποθεσία (όπως π.χ. home, paradise, heaven, hell κτλ.), διαγράφεται και δεν αναζητούνται οι συντεταγμένες.

Γ'.2.5.6 Βάση δεδομένων

Η βάση δεδομένων του Flutrack αποτελεί μια απλή σχεσιακή βάση δεδομένων. Αποθηκεύει δεδομένα μετά από κάθε φιλτράρισμα. Το πρωτεύον κλειδί που καθορίζει τη βάση ως μοναδική είναι το `tweet_id`, το οποίο για κάθε τweet που δημιουργείται είναι μοναδικό. Το κάθε μεταδεδομένο ενός τweet είναι και ένα πεδίο στη βάση δεδομένων σε μια στήλη. Τα Tweets αποθηκεύονται στη βάση δεδομένων έπειτα από το φιλτράρισμα. Αυτό έχει ως συνέπεια, να μην αποθηκευονται τα Tweets που δεν περιλαμβάνουν τις απαραίτητες πληροφορίες γεωγραφικών συντεταγμένων όπως και τα Tweets που περιέχουν λιγότερους από 5 χαρακτήρες. Εφόσον η διαδικασία της αποθήκευσης έχει έρθει εις πέρας, η βάση δεδομένων έχει μόνο Tweets που έχουν όλα τα πεδία των μεταδεδομένων τους συμπληρωμένα. Τότε ξεκινά η διαδικασία εξαγωγής ενός .JSON αρχείου, το οποίο περιέχει τα Tweets των τελευταίων 7 ημερών. Το διάστημα αυτό επιλέχτηκε με κριτήριο τη διάρκεια της εκδήλωσης της ασθένειας της γρίπης στον ασθενή ως την εξασθένηση των συμπτωμάτων. Το .JSON αρχείο που εξάγεται από τη βάση, είναι το αρχείο που τελικά οπτικοποιείται από Front-end τμήμα της εφαρμογής Flutrack. Για την οπτικοποίηση των δεδομένων δεν απαιτούνται όλα τα μεταδεδομένα. Για το λόγο αυτό, στο προς εξαγωγή αρχείο περιέχεται μόνο το κείμενο του Tweet, το γεωγραφικό πλάτος, το γεωγραφικό μήκος, και η τιμή του Aggravation flag.

Γ'.2.5.7 Διεπαφή Προγραμματισμού Εφαρμογών (Flutrack API)

Το σύστημα Flutrack, επαναδιαθέτει τα δεδομένα που εισάγει, επεξεργάζεται και αποθηκεύει μέσω μιας ανοιχτής Διεπαφής Προγραμματισμού Εφαρμογών (Flutrack API). Η εφαρμογή διαθέτει τα δεδομένα σε κάθε χρήστη ηλεκτρονικού υπολογιστή με διασύνδεση στο διαδίκτυο μέσω μιας αίτησης σε κάθε διαθέσιμο φυλλομετρητή διαδικτύου. Η Διεπαφή Προγραμματισμού Εφαρμογών του Flutrack, ταξινομεί και εξάγει τα δεδομένα σύμφωνα με:

- Τα συμπτώματα της γρίπης.
- Την επιδείνωση της κλινικής κατάστασης του ασθενή.
- Την ημερομηνία και την ώρα.
- Αριθμό των Tweets.

Οι παράμετροι που γίνονται αποδεκτές από το σύστημα είναι:

- Αναζήτηση:(s), Αναζήτηση συγκεκριμένων συμπτωμάτων με την υποστήριξη των AND, OR τελεστών. Παράδειγμα: <http://api.Flutrack.org/?s=feverANDcoughORfever>.
- Επιδείνωση:(a), Ορίζει αν ο χρήστης επιθυμεί μόνο τα Tweets που δείχνουν επιδείνωση της κλινικής κατάστασης του ασθενή, ή όχι. Παράδειγμα: <http://api.Flutrack.org/?a=TRUE>.
- Χρόνος: (time), Επιστρέφει αποτελέσματα που ανήκουν σε ένα διάστημα των τελευταίων 7 ημερών. Παράδειγμα: <http://api.Flutrack.org/?time=7>.
- Αριθμός Tweets: (limit), Αριθμός επιστρεφόμενων Tweets με όριο την τιμή 100. Παράδειγμα: <http://api.Flutrack.org/?limit=90>.

Εκτός από τα παραπάνω, κάθε χρήστης μπορεί ανά πάσα στιγμή να εξάγει από την εφαρμογή τα δεδομένα που οπτικοποιούνται σε πραγματικό χρόνο στέλνοντας αίτηση στη διαδικτυακή διεύθυνση: <http://Flutrack.org/results.json>. Με τον τρόπο αυτό αποστέλλεται κατευθείαν στο χρήστη ένα αρχείο .JSON με τα δεδομένα που ζήτησε.

Γ'.2.6 Front-end τμήμα

Γ'.2.6.1 Οπτικοποίηση

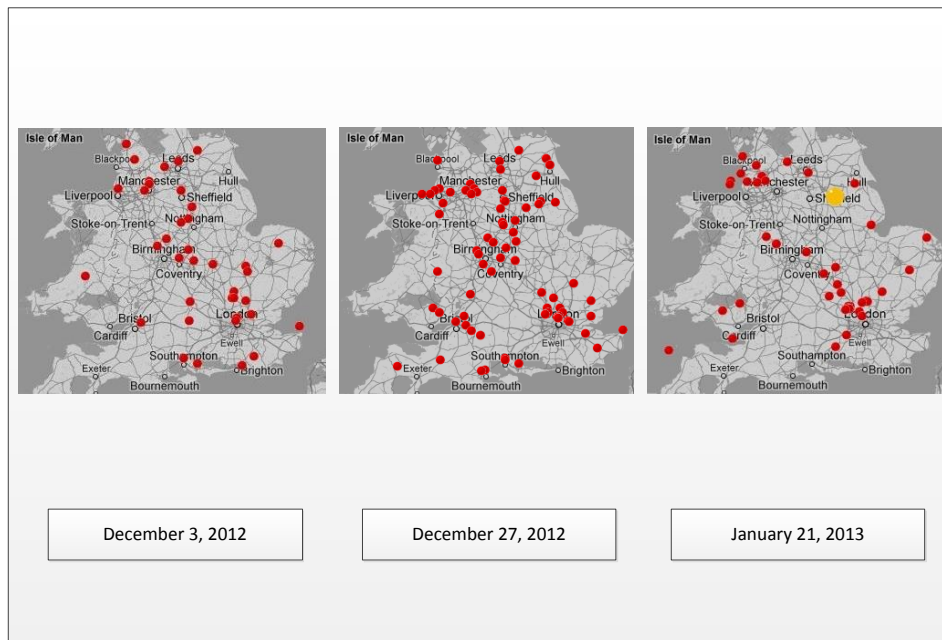
Από τη στιγμή που τα δεδομένα έχουν φιλτραριστεί, αποθηκευτεί και εξαχθεί με τον κατάλληλο τρόπο από τη βάση δεδομένων, ξεκινά η διαδικασία οπτικοποίησης τους. Το Front-end τμήμα εισάγει το αρχείο με τα προς οπτικοποίηση δεδομένα, και τα οπτικοποιεί σε έναν παγκόσμιας κλίμακας χάρτη χρησιμοποιώντας την υπηρεσία Google Maps.

Ο Google Map χάρτης έχει επεξεργαστεί έτσι ώστε να μην απεικονίζονται αχρείαστες πληροφορίες. Επικεντρώνεται στην απεικόνιση πληροφοριών σχετικών με τη γρίπη. Για παράδειγμα έχει φιλτραριστεί στοιχεία όπως χρωματισμός αυτοκινητοδρόμων, δασών, υδάτινων εκτάσεων κτλ. Ο συνολικός χάρτης έχει γκρι απόχρωση, η οποία αποσκοπεί στην έμφαση της απεικόνισης των σχετικών με τη γρίπη στοιχείων.

Κάθε Tweet, οπτικοποιείται στο ψηφιακό χάρτη με ένα κόκκινου χρώματος σημείο, ενώ αν δείχνει επιδείνωση οπτικοποιείται με σημείο κίτρινου χρώματος. Τα Tweets, τοποθετούνται στο χάρτη σύμφωνα με τις γεωγραφικές συντεταγμένες που περιέχονται στα μεταδεδομένα τους. Η απόκλιση της κλίμακας συντεταγμένων οπτικοποίησης με την πραγματική κλίμακα φτάνει το μέγιστο τα 20 μέτρα, οπότε χαρακτηρίζεται από μεγάλη ακρίβεια. Με την επιλογή ενός σημείου στο χάρτη, ο χρήστης μπορεί να παρακολουθήσει το περιεχόμενο του σχετικού με την εποχική γρίπη Tweet. Η επιλογή του κάθε σημείου ενεργοποιεί ένα συμβάν και μέσω της γλώσσας Javascript φέρνει δυναμικά τα δεδομένα που οπτικοποιούνται.

Η επιλογή του κέντρου του χάρτη, γίνεται με τρόπο δυναμικό. Από τη στιγμή που ο χρήστης εισέλθει στο σύστημα, του τίθεται ερώτημα να γνωστοποιήσει τη θέση του στο χάρτη μέσω της διεύθυνσης IP της σύνδεσής του στο διαδίκτυο. Αν ο χρήστης δεχτεί να γνωστοποιήσει τη θέση του, το κέντρο του χάρτη ορίζεται στο σημείο που βρίσκεται ο χρήστης. Σε διαφορετική περίπτωση, ως κέντρο του χάρτη ορίζεται ένα προεπιλεγμένο σημείο. Η αυξομείωση της απόστασης (zoom in/out), το κέντρο του χάρτη και η εμφάνιση (ή μη) των Tweets μπορούν να αλλάξουν έπειτα από παρέμβαση του χρήστη.

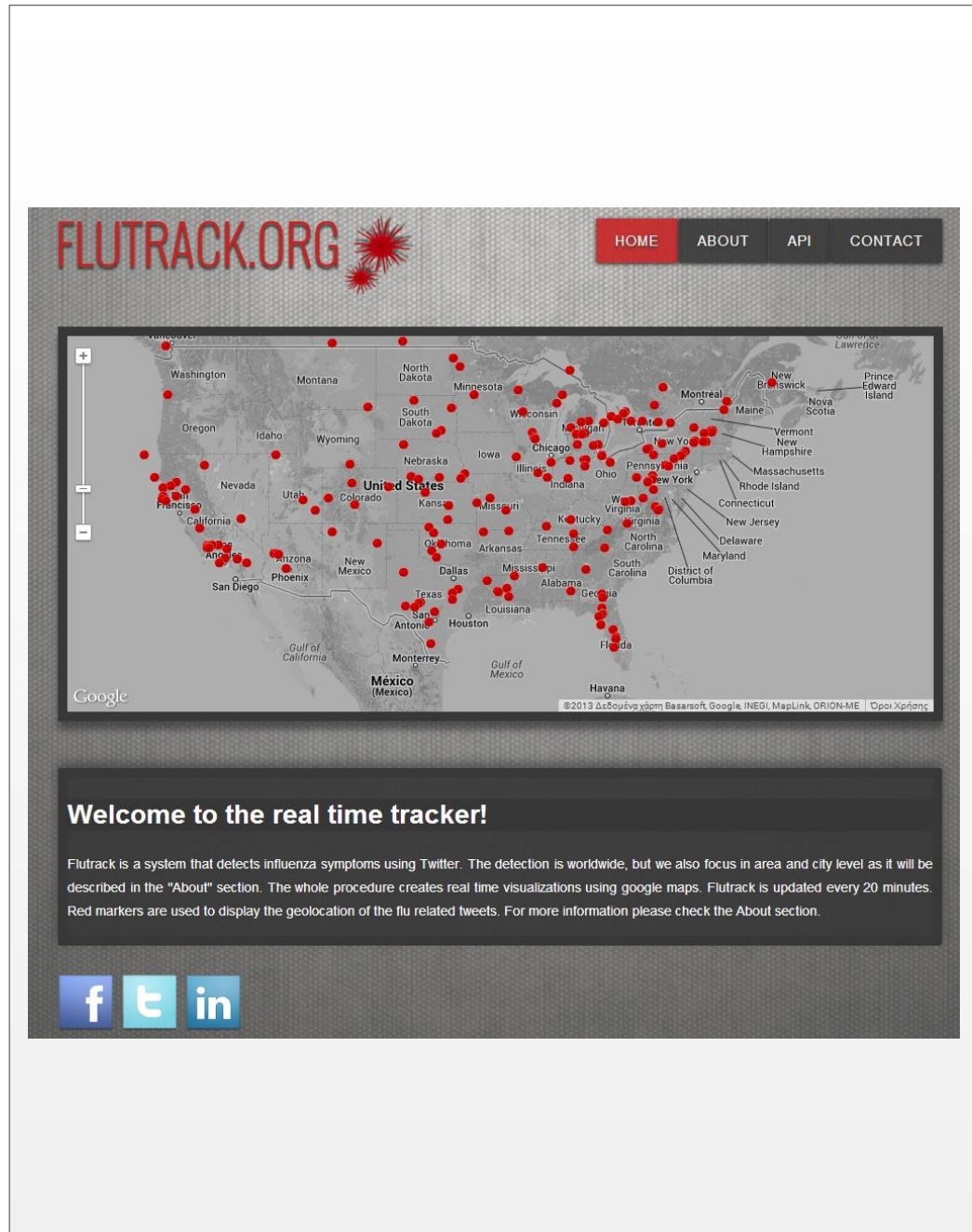
Κάθε 20 λεπτά, γίνεται εισαγωγή νέου αρχείου δεδομένων, όπως αυτό προκύπτει από το Back-end τμήμα της εφαρμογής. Το κάθε Tweet παραμένει στο χάρτη για 7 ημέρες ανεξάρτητα από τη νέα εισαγωγή δεδομένων στο χάρτη. Ενδεικτικά, στο παρακάτω σχήμα Γ'.7, φαίνεται η εξέλιξη της εξάπλωσης της εποχικής γρίπης σε συγκεκριμένα χρονικά διαστήματα σε μια περιοχή του Ηνωμένου Βασιλείου:



Σχήμα Γ΄.7: Παράδειγμα οπτικοποίησης σε συγκεκριμένες ημερομηνίες

Γ΄.2.6.2 Ιστοσελίδα εφαρμογής

Η εφαρμογή φιλοξενείται στην ιστοσελίδα <http://Flutrack.org>, η οποία δημιουργήθηκε για τον σκοπό αυτό. Η ιστοσελίδα περιέχει όλες τις απαραίτητες πληροφορίες ώστε να μπορεί ο χρήστης να παρακολουθήσει την εξάπλωση της γρίπης και να διαδράσσει με τον ψηφιακό χάρτη της εφαρμογής. Εκτός από τον ψηφιακό χάρτη, περιέχονται πληροφορίες σχετικές με τη λειτουργία του συστήματος και της διαδικασίας που ακολουθείται. Επίσης, παρέχονται πληροφορίες για τη Διεπαφή Προγραμματισμού Εφαρμογών Flutrack. Αναλύονται οι παράμετροι και δίνονται παραδείγματα για τους τρόπους που μπορεί κανείς να εξάγει δεδομένα από την πλατφόρμα. Η ιστοσελίδα αναπτύχθηκε ώστε να λειτουργεί σε σταθερούς υπολογιστές, σε laptops, tablets, smartphones με την ίδια ευκολία. Ακόμη, λειτουργεί σε όλους τους σύγχρονους περιηγητές-φυλλομετρητές διαδικτύου (όπως Google Chrome, Mozilla Firefox, IE, Safari, Opera κτλ.) με ελάχιστες διαφορές.



Σχήμα Γ'.8: Αρχική σελίδα του ιστοτόπου <http://flutrack.org>

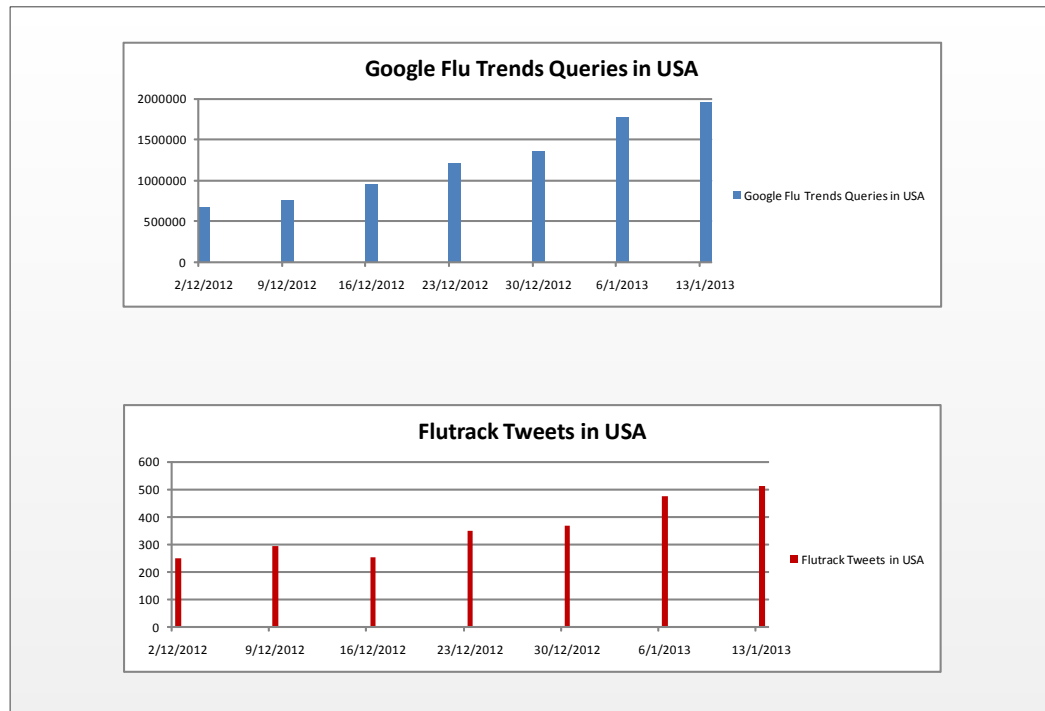
Κεφάλαιο Δ΄

Αποτελέσματα και συμπεράσματα

Δ΄.1 Αποτελέσματα

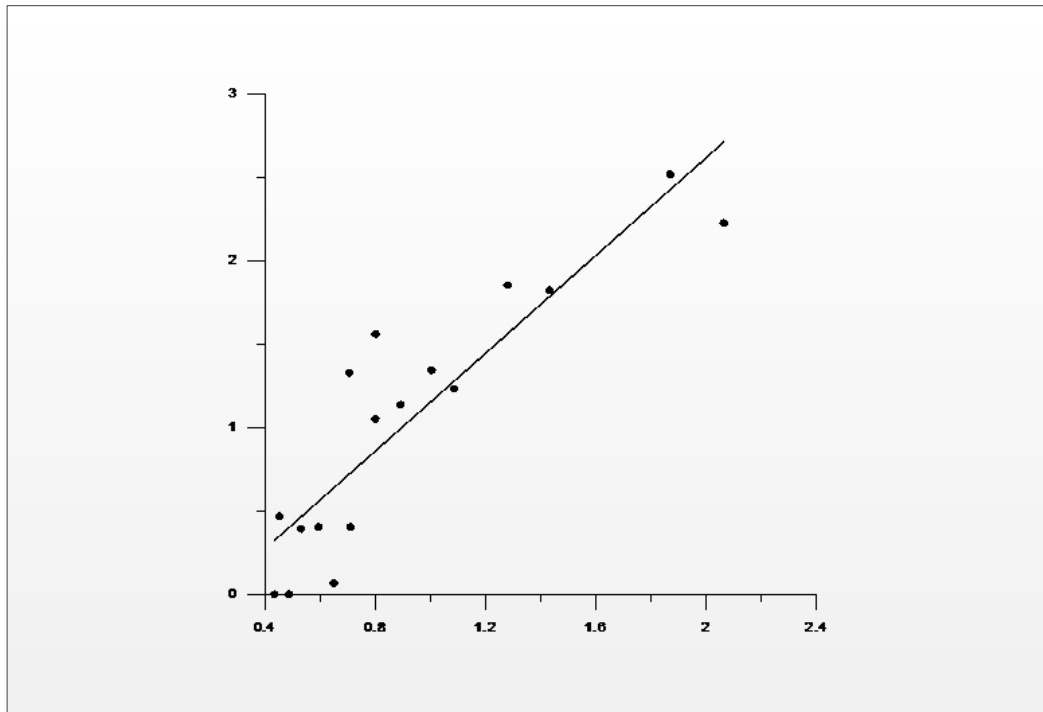
Για να αξιολογηθεί η ακρίβεια των στοιχείων που προέρχονται από το σύστημα Flutrack, έπρεπε να διερευνηθεί αν υπάρχει κάποιος συσχετισμός με επίσημα δεδομένα από άλλες υπηρεσίες. Η υπηρεσία που επιλεχτηκε για τη διερεύνηση αυτή είναι το Google Flu Trends. Ένας εύκολος τρόπος για να εξεταστεί ο συσχετισμός των δεδομένων των 2 συστημάτων, έπρεπε να επιλεχτούν δεδομένα από ίδιες ημερομηνίες. Το Google Flu Trends επαναδιαθέτει τα αποθηκευμένα του δεδομένα ταξινομημένα ανά εβδομάδα. Συνεπώς, τα δεδομένα της εφαρμογής Flutrack ταξινομήθηκαν με τον ίδιο τρόπο για τις ανάγκες αυτής της διαδικασίας.

Όλα τα Tweets και τα Queries των δύο εφαρμογών (παρόχων δεδομένων), προέρχονταν από περιοχές με συντεταγμένες στην επικράτεια των Ηνωμένων Πολιτειών Αμερικής (ΗΠΑ). Η επιλογή αυτή έγινε λόγω του ότι τα Tweets στις ΗΠΑ είναι στην αγγλική γλώσσα και η ποσότητα των αποθηκευμένων Tweets στη βάση δεδομένων του Flutrack αποτελεί ένα ικανοποιητικό δείγμα προς εξέταση. Επιπλέον, τα σχετικά με τη γρίπη δεδομένα του Google Flu Trends στις ΗΠΑ είναι πιο αναλυτικά από ότι σε άλλες περιοχές και έχουν ήδη συσχετιστεί με το Αμερικανικό Κέντρο Ελέγχου Ασθενειών. Τα δεδομένα που ταξινομήθηκαν και αναλύθηκαν ήταν από τις 2/12/2012 έως τις 07/04/2013. Το χρονικό αυτό διάστημα ανήκει στην κύρια περίοδο έξαρσης της εποχικής γρίπης στις περιοχές του βόρειου ημισφαιρίου.



Σχήμα Δ'.9: Συγκριτικοί πίνακες δεδομένων των υπηρεσιών Google Flu Trends και Flutrack ανά εβδομάδα, ενδεικτικά από τις 2/12/12 έως 13/1/2013.

Το παρακάτω διάγραμμα Δ'.10, απεικονίζει τη διασπορά των 2 κανονικοποιημένων πληθυσμών (των Tweets του Flutrack και των Search Queries του Google Flu Trends). Από την απεικόνιση φαίνεται ότι υπάρχει σαφής γραμμική συσχέτιση μεταξύ των 2 πληθυσμών. Προκειμένου να προσδιοριστεί ποσοτικά ο βαθμός της παρατηρούμενης γραμμικής συσχέτισης μεταξύ των 2 πληθυσμών, εκτελέστηκε μια ανάλυση γραμμικής παλινδρόμησης. Στο παραπάνω σχήμα ο συντελεστής προσδιορισμού υπολογίζεται σε 0.79, δείχνοντας έτσι έναν υψηλό βαθμό συσχέτισης μεταξύ του Google Flu Trends και της πλατφόρμας Flutrack.



Σχήμα Δ΄.10: διάγραμμα διασποράς δεδομένων των υπηρεσιών *Google Flu Trends* και *Flutrack*

Δ΄.2 Μελλοντική εργασία

Το σύστημα Flutrack αλλά και η ευρύτερη πλατφόρμα με παραμετροποιήσεις μπορούν να χρησιμοποιηθούν μελλοντικά για περαιτέρω ανάλυση. Σε ότι αφορά τη στατιστική ανάλυση, ένα βήμα εμπρός θα ήταν μια μελέτη στο χώρο και το χρόνο της εξέλιξης της γρίπης. Η εξέλιξη δηλαδή των μολυσμένων πληθυσμών σε διάφορες πόλεις ή περιοχές προοδευτικά με το χρόνο. Η μελέτη αυτή θα μπορούσε να γίνει με μοντέλα πολλαπλών κλιμάκων (multiscale models) που αναπτύχθηκαν για άλλες εργασίες. Τα μοντέλα αυτά έχουν τη δυνατότητα να χρησιμοποιούν "κοινωνικά δεδομένα", όπως αυτά του συστήματος Flutrack, προκειμένου να επιτευχθεί η έγκαιρη και ακριβής ανίχνευση των επιδημιών της εποχικής γρίπης.

Μια άλλη προοπτική έρευνας είναι η χρήση τεχνικών μηχανικής μάθησης που θα αποσκοπούν στην αυτόματη ανίχνευση της εποχικής γρίπης και της εξέλιξής της. Στον εντοπισμό δηλαδή

του αν η κλινική κατάσταση ενός ασθενή έχει επιδεινωθεί ή βελτιωθεί. Για το σκοπό αυτό, γλωσσικά χαρακτηριστικά (π.χ. unigrams λέξεις ή bigram φράσεις, συγκριτικά επίθετα, ποσοτικά επιρρήματα κ.λπ.), emoticons και άλλα κοινωνικά χαρακτηριστικά κειμένου , όπως επαναλαμβανόμενοι χαρακτήρες (π.χ. reaaaaally), οι οποίοι αντανακλούν τις αλλαγές της κλινικής κατάστασης, θα μπορούσαν να χρησιμοποιηθούν για να εντοπίσουν την έννοια του περιεχομένου του κάθε Tweet .Η αυτόματη αναγνώριση της επιδείνωσης των συμπτωμάτων και η επιδείνωση των συνθηκών θα μπορούσε να υποστηρίξει την έγκαιρη διάγνωση και την κατάλληλη θεραπεία . Η Εξόρυξη κειμένου (Text Mining) έχει προταθεί εκτενώς σε προηγούμενες εργασίες για την ανίχνευση της γρίπης.

Μια άλλη σχετική εργασία ταξινόμησης είναι η αναγνώριση της μιας περιγραφής της γρίπης σε ένα συγκεκριμένο Tweet, δηλαδή αν το Tweet αναφέρεται σε ένα πραγματικό ασθενή της εποχικής γρίπης σε μια συγκεκριμένη θέση, ή αν είναι πιο αφηρημένη και αναφέρεται στη γρίπη αόριστα. Κάνοντας χρήση ενός πιο εξελιγμένου συνόλου γλωσσικών χαρακτηριστικών στις λέξεις γύρω από γρίπη, όπως προτείνεται στο [παραπομπή], όπως και μέρος ετικετών και των σημείων στίξης , μπορούν να οδηγήσουν σε βελτιωμένη ακρίβεια ανίχνευσης της γρίπης.

Επιπλέον, κάνοντας χρήση του χρόνου και της τοποθεσίας των τweetς, η εξέλιξη των επιδημιών μπορεί να μοντελοποιηθεί και να προβλεφτεί η εξάπλωσή τους [παραπομπή] [παραπομπή]. Οι σχετικές προσεγγίσεις έχουν αποδειχθεί ότι συμβάλλουν σημαντικά στη δημόσια υγειονομική περίθαλψη. Επιπλέον, αντίκειμενο έρευνας μπορεί να αποτελέσει η συσχέτιση των δεδομένων συστήματος Flutrack, με δεδομένα μεταβολών θερμοκρασίας και καιρικών συνθηκών ώστε να μελετηθεί εαν οι μεταβολές αυτές μπορούν να επηρεάσουν την εξάπλωση επιδημιών γρίπης.

Δ'.3 Ηθικά ζητήματα

Πολλά είναι τα ηθικά ζητήματα που προκύπτουν από τα δεδομένα των χρηστών στα μέσα κοινωνικής δικτύωσης σε ότι αφορά τη παραβίαση της ιδιωτικότητας τους. Έτσι, γίνεται εύκολα αντιληπτό ότι παρόμοια θέματα προκύπτουν για το σύστημα Flutrack, ειδικά από τη στιγμή που αυτό εισάγει δεδομένα από κοινωνικά δίκτυα. Το σύστημα Flutrack αποθηκεύει τα 'account names' των χρηστών, κάτι που άρρει την ανωνυμότητα των ασθενών γρίπης. Το μέγεθος του ηθικού ζητήματος θα μπορούσε να μεγιστοποιηθεί αν τα δεδομένα δεν προέρχονταν από κοινωνικό δίκτυο, αλλά για παράδειγμα από SMS, όπου θα γνωστοποιούνταν αυτόματα και ο τηλεφωνικός αριθμός του χρήστη. Το σύστημα Flutrack δεν οπτικοποιεί καμία πλη-

ροφορία σχετική με την ιδιωτικότητα των χρηστών, ταυτόχρονα όμως απεικονίζει την ακριβή θέση τους στο χάρτη. Αναλογιζόμενοι την ευρύτερη χρήση της πλατφόρμας Flutrack, τα ηθικά ζητήματα που προκύπτουν είναι ακόμη περισσότερα.

Εκτός από την παραβίαση της ιδιωτικότητας, η χρησιμοποίηση ετικετών με τρόπο κακόβουλο, μπορεί να οδηγήσει σε σοβαρά ηθικά ζητήματα. Χρησιμοποιώντας ετικέτες με κακόβουλο τρόπο και λαμβάνοντας υπόψη ότι στα κοινωνικά δίκτυα υπάρχουν ανήλικοι χρήστες, ο εντοπισμός τους στο χάρτη σε πραγματικό χρόνο είναι γεγονός που δημιουργεί ανησυχία.

Κεφάλαιο Ε΄

Επίλογος

Ε΄.1 Σύνοψη

Στο παρόν κείμενο παρουσιάζεται μια πλατφόρμα ευρείας χρήσης που αναπτύχθηκε στα πλαίσια της εκπόνησης της διπλωματικής εργασίας του τμήματος Πληροφορικής του Ιονίου Πανεπιστημίου. Επιπλέον, αναλύεται η ανάπτυξη ενός πρωτοτύπου, το οποίο βασίζεται στην προαναφερθείσα πλατφόρμα και προτείνει τη μελέτη και εμφάνιση της εποχικής γρίπης σε παγκόσμια κλίμακα, όπως αυτή προκύπτει από τα σχετικά μηνύματα του κοινωνικού δικτύου Twitter. Το προτεινόμενο σύστημα επανιδιαθέτει τα δεδομένα που αποθηκεύει και οπτικοποιεί μέσω μιας ανοιχτής Διεπαφής Προγραμματισμού Εφαρμογών. Το συνολικό σύστημα μπορεί να λειτουργήσει, όχι μόνο μέσω της Διεπαφής Προγραμματισμού Εφαρμογών του Twitter, αλλά με οποιοδήποτε πάροχο δεδομένων. Το σύστημα θα μπορούσε να χρησιμοποιηθεί από ευαίσθητες κοινωνικά ομάδες (ηλικιωμένους, παιδιά μικρής ηλικίας, έγκυες γυναίκες κτλ.) για τον προγραμματισμό των δραστηριοτήτων τους σε συνδυασμό με την αποφυγή περιοχών που παρουσιάζουν έξαρση στην εποχική γρίπη. Ένα περαιτέρω βήμα, θα ήταν η χρησιμοποίηση των δεδομένων του Flutrack από οργανισμούς και υπηρεσίες με προϋπόθεση την παραμετροποίηση του συστήματος ανάλογα με τις εκάστοτε ανάγκες. Τέλος, εφαρμόζοντας τεχνικές μηχανικής μάθησης και συσχέτισης δεδομένων με άλλες υπηρεσίες, είναι δυνατό να επιτευχθεί μεγάλη ακρίβεια πρόγνωσης και πρόληψης επιδημιών εποχικής γρίπης.

Παράρτημα Α΄

Ενδεικτικό παράδειγμα παραρτήματος. Έπεται το κεφάλαιο της βιβλιογραφίας.

Βιβλιογραφία

- [1] Eccles R., *Understanding the symptoms of the common cold and influenza*, Vol 5, Issue 11, pp. 718-725, 2005.
- [2] Thursky K., *Working towards a simple case definition for influenza surveillance*, Journal of Clinical Virology, Vol. 27, Issue 2, pp.170-179, 2003.
- [3] Taubenburger JK., *Influenza: the once and future pandemic*, Public Health Reports (Washington, D.C.), pp.3:16-26, 2010.
- [4] Ghendon Y., *Introduction to pandemic influenza through history*, European Journal of Epidemiology, pp.451-453, 1994.
- [5] Beveridge W., *The Chronicle of influenza epidemics*, History and Philosophy of the Life Sciences, pp.13(2) : 223-234, 1991.
- [6] Belshe RB., *An introduction to influenza: lessons from the past in epidemiology, prevention and treatment*, Managed Care (Langhorne, Pa.), pp. 2-7, 2008.
- [7] Pak A., Paroubek P., *Twitter as a corpus for Sentiment Analysis and Opinion Mining*, Proceedings of LREC, 2010.
- [8] Agarwal, A., Xie, B., Vovsha, I., Rambow, O., Rassonneau, R., *Sentiment analysis of Twitter data*, LSM b●●11 Proceedings of the Workshop on Languages in Social Media, pp. 30-38, 2011.
- [9] Culotta A., *Towards detecting influenza epidemics by analyzing Twitter messages*, Proceedings of the First Workshop on Social Media Analytics, 2010.

- [10] Aramaki, E., Maskawa, S., Morita, M., *Twitter Catches The Flu, Detecting Influenza Epidemics using Twitter*, Conference on Empirical Methods in Natural Language Processing, EMNLP, 2011.
- [11] Norris B., Boicey C., Silverberg M., *MappyHealth application*, <http://mappyhealth.com> , 2012.
- [12] Sadilek A., Henry Kautz H., *Modelling the impact of lifestyle on health at scale*, Proceedings of the sixth ACM international conference on Web search and data mining, 2013.
- [13] Google, *Google Flu Trends*, <http://www.google.org/flutrends>, 2012.
- [14] Google Maps, *Google Maps API V.3*, <https://developers.google.com/maps>, 2012.
- [15] Courtney D. Corley, Diane J. Cook, Armin R. Mikler, and Karan P. Singh, *Text and Structural Data Mining of Influenza Mentions in Web and Social Media*, International Journal of Environmental Research and Public Health, Vol 7, pp. 596-615, 2010.
- [16] Collier N., *Uncovering text mining: A survey of current work on web-based epidemic intelligence*, Journal on Global Public Health, 2012.
- [17] Harshavardhan A., Avinash G., Ross L., Ssu-Hsin Y. and Benyuan L., *Twitter Improves Seasonal Influenza Prediction*, HEALTHINF, 2012.
- [18] Avlonitis M., Magkos E., Stefanidakis M. and Chrissikopoulos V., *A spatial stochastic model for worm propagation: scale effects*, J. Comput. Virol. 3:87b•“92, 2007.
- [19] Sadilek A., Henry A. Kautz, Silenzio V., *Predicting Disease Transmission from Geo-Tagged Micro-Blog Data*, AAAI, 2012.
- [20] Ginsberg J., Mohebbi M.H., Patel R.S., Brammer L., *Detecting influenza epidemics using search engine query data*, Nature 457, pp. 1012b•“1014, 2009.
- [21] Signorini A., Segre A.M., Polgreen P.M., *The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza A H1N1 Pandemic*, PLoS ONE, www.plosone.org , 2011.

Συντμήσεις

API	Application Programming Interface
CDC	Centers for Disease Control and Prevention
CSS	Cascading Style Sheets
IP	Internet Protocol
GISRS	Global Influenza Surveillance and Response System
GPS	Global Positioning System
HTML	HyperText Markup Language
MVC	Model View Controller
PHP	Hypertext Preprocessor
Regex	Regular Exprpressions
SMS	Short Message Service
MBS	Work Breakdown Structure

WEB

World Wide Web

WHO

World Health Organization

W3C

World Wide Web Consortium

Γλωσσάρι Ξενικών Όρων

Aggravation	Επιδείνωση
Application Programming Interface	Διεπαφή Προγραμματισμού Εφαρμογών
Browser	Φυλλομετρητής
Call Request	Αίτηση κλήσης
Conceptual Model	Εννοιολογικό μοντέλο
Data Structures	Δομές δεδομένων
Emoticon	Εικονογραφημένη αναπαράσταση της έκφρασης ενός προσώπου
Forum	Περιοχή συζητήσεων στο Διαδίκτυο
Global Influenza Surveillance and Response System	Παγκόσμιο σύστημα επιτήρησης και πρόληψης της γρίπης
Global Positioning System	Παγκόσμιο σύστημα εντοπισμού θέσης
Interactive Interface	διαδραστικός Διεπαφή
Interpreted language	Ερμηνευόμενη γλώσσα

Latitude	Γεωγραφικό πλάτος
Limit	Όριο
Linguistic filtering	Γλωσσική ανάλυση φιλτραρίσματος
Longitude	Γεωγραφικό μήκος
Mapping	Χαρτογράφηση
Object classes	Κλάσεις αντικειμένων
Optimization	Βελτιστοποίηση
Protocol	Πρωτόκολλο
Query	Ερώτημα
Regular Expressions	Κανονικές εκφράσεις
Search Engine	Μηχανή αναζήτησης
Script Language	Γλώσσα σεναρίων
User profile location	Γεωγραφική τοποθεσία προφίλ χρήστη
Text mining	Ανάσχυση κειμένου