

**ΑΕΙ ΠΕΙΡΑΙΑ Τ.Τ.  
ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΚΩΝ ΕΦΑΡΜΟΓΩΝ  
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΗΛΕΚΤΡΟΝΙΚΩΝ  
ΥΠΟΛΟΓΙΣΤΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ Τ.Ε.**

**ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

**Εργαλείο Ελέγχου Πληροφοριών σε Μέσα Κοινωνικής  
Δικτύωσης (Social Media) με εφαρμογή σε ειδικές ομάδες**

**Μαρία Π. Βουρνά**

**Εισηγητής: Δρ Παναγιώτης Γιαννακόπουλος, Καθηγητής**

**ΑΘΗΝΑ  
ΜΑΪΟΣ 2018**





Εργαλείο Ελέγχου Πληροφοριών σε Μέσα Κοινωνικής Δικτύωσης (Social Media) με εφαρμογή σε ειδικές ομάδες

## **ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

**Εργαλείο Ελέγχου Πληροφοριών σε Μέσα Κοινωνικής Δικτύωσης (Social Media) με εφαρμογή σε ειδικές ομάδες**

**Μαρία Π. Βουρνά  
Α.Μ. 39570**

**Εισηγητής:**

**Δρ Παναγιώτης Γιαννακόπουλος, Καθηγητής**

**Εξεταστική Επιτροπή:**

**Ημερομηνία εξέτασης /06/2018**

Εργαλείο Ελέγχου Πληροφοριών σε Μέσα Κοινωνικής Δικτύωσης (Social Media) με εφαρμογή σε ειδικές ομάδες

## ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

Η κάτωθι υπογεγραμμένη Βουρνά Μαρία, του Παναγιώτη, με αριθμό μητρώου 39570 φοιτήτρια του Τμήματος Μηχανικών Η/Υ Συστημάτων Τ.Ε. του Α.Ε.Ι. Πειραιά Τ.Τ. πριν αναλάβω την εκπόνηση της Πτυχιακής Εργασίας μου, δηλώνω ότι ενημερώθηκα για τα παρακάτω:

«Η Πτυχιακή Εργασία (Π.Ε.) αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο του συγγραφέα, όσο και του Ιδρύματος και θα πρέπει να έχει μοναδικό χαρακτήρα και πρωτότυπο περιεχόμενο.

Απαγορεύεται αυστηρά οποιοδήποτε κομμάτι κειμένου της να εμφανίζεται αυτούσιο ή μεταφρασμένο από κάποια άλλη δημοσιευμένη πηγή. Κάθε τέτοια πράξη αποτελεί προϊόν λογοκλοπής και εγείρει θέμα Ηθικής Τάξης για τα πνευματικά δικαιώματα του άλλου συγγραφέα. Αποκλειστικός υπεύθυνος είναι ο συγγραφέας της Π.Ε., ο οποίος φέρει και την ευθύνη των συνεπειών, ποινικών και άλλων, αυτής της πράξης.

Πέραν των όποιων ποινικών ευθυνών του συγγραφέα σε περίπτωση που το Ίδρυμα του έχει απονεμίσει Πτυχίο, αυτό ανακαλείται με απόφαση της Συνέλευσης του Τμήματος. Η Συνέλευση του Τμήματος με νέα απόφαση της, μετά από αίτηση του ενδιαφερόμενου, του αναθέτει εκ νέου την εκπόνηση της Π.Ε. με άλλο θέμα και διαφορετικό επιβλέποντα καθηγητή. Η εκπόνηση της εν λόγω Π.Ε. πρέπει να ολοκληρωθεί εντός τουλάχιστον ενός ημερολογιακού 6μήνου από την ημερομηνία ανάθεσης της. Κατά τα λοιπά εφαρμόζονται τα προβλεπόμενα στο άρθρο 18, παρ. 5 του ισχύοντος Εσωτερικού Κανονισμού».

Εργαλείο Ελέγχου Πληροφοριών σε Μέσα Κοινωνικής Δικτύωσης (Social Media) με εφαρμογή σε ειδικές ομάδες

## **ΕΥΧΑΡΙΣΤΙΕΣ**

Η παρούσα πτυχιακή εργασία ολοκληρώθηκε μετά από σημαντικές προσπάθειες παρά τις όποιες χρονικές και προσωπικές δυσκολίες, σε ένα ενδιαφέρον γνωστικό αντικείμενο, όπως αυτό του προγραμματισμού και τεχνητής νοημοσύνης . Την προσπάθειά μου αυτή υποστήριξε ο επιβλέπων καθηγητής μου, τον οποίο θα ήθελα να ευχαριστήσω.

Ακόμα θα ήθελα να ευχαριστήσω την οικογένεια μου για την υπομονή και την στήριξη που μου παρείχαν όλα αυτά τα χρόνια. Και φυσικά τους φίλους και τους συναδέλφους μου από το τμήμα αλλά και από την εταιρεία στην οποία εργάζομαι, για τη στήριξη και την ενθάρρυνσή τους, αλλά και για τις χρήσιμες συμβουλές τους.



Εργαλείο Ελέγχου Πληροφοριών σε Μέσα Κοινωνικής Δικτύωσης (Social Media) με εφαρμογή σε ειδικές ομάδες

## ΠΕΡΙΛΗΨΗ

Η παρούσα πτυχιακή ασχολείται με τα μέσα κοινωνικής δικτύωσης, και την ανάλυση συναισθημάτων. Αναλυτικότερα γίνεται αναφορά στην επεξεργασία της φυσικής γλώσσας και στην ανάλυση των συναισθημάτων που προκύπτει από την εξόρυξη των δεδομένων από τους ιστοτόπους των μέσων κοινωνικής δικτύωσης. Επίσης στα πλαίσια της εκπόνησης της πτυχιακής, δημιουργήθηκε ένα παράδειγμα συλλογής δεδομένων από τους χρήστες του Twitter, τα οποία αφορούν μια συγκεκριμένη ειδική ομάδα όπως είναι αυτή της ναυτιλίας.

## ABSTRACT

The subject of this thesis focuses on the social media network and data sentiment analysis. Specifically, presented definitions for natural language processing and the analysis of emotions which resulting from data mining network. Also, created a data mining collector from Twitter, which applies them some rules for sentiment analysis, in a group like shipping.

ΕΠΙΣΤΗΜΟΝΙΚΗ ΠΕΡΙΟΧΗ: Προγραμματισμός  
ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: μέσα κοινωνικής δικτύωσης, εξόρυξη δεδομένων, ανάλυση δεδομένων, ανάλυση συναισθημάτων, social media, twitter, data mining, data analysis, sentiment analysis.

Εργαλείο Ελέγχου Πληροφοριών σε Μέσα Κοινωνικής Δικτύωσης (Social Media) με εφαρμογή σε ειδικές ομάδες

## ΠΕΡΙΕΧΟΜΕΝΑ

<b>1. ΕΙΣΑΓΩΓΗ</b> .....	<b>11</b>
<b>2.ΜΕΣΑ ΚΟΙΝΩΝΙΚΗΣ ΔΙΚΤΥΩΣΗΣ ΚΑΙ ΠΟΛΙΤΙΚΗ ΠΡΟΣΤΑΣΙΑΣ ΤΗΣ ΙΔΙΩΤΙΚΟΤΗΤΑ ΣΕ ΑΥΤΑ</b> .....	<b>13</b>
2.1 Βασικές γνώσεις .....	13
2.2 Εισαγωγή στα Social Media.....	14
2.2.1 Δημοφιλέστερα Social Media .....	15
2.3 Πολιτική Προστασίας της Ιδιωτικότητας στα Μέσα Κοινωνικής Δικτύωσης.	18
<b>3.ΑΝΑΛΥΣΗ ΣΥΝΑΙΣΘΗΜΑΤΩΝ</b> .....	<b>23</b>
3.1 Γενικά για την ανάλυση συναισθημάτων .....	23
3.2 Επεξεργασία Φυσικής γλώσσας. (NLP).....	23
3.3 Ανάλυση συναισθημάτων (Sentiment Analysis).....	24
3.4 Κατηγοριοποίηση κειμένου (Text classification).....	25
<b>4.Η ΓΛΩΣΣΑ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ ΡΥΘΩΝ</b> .....	<b>27</b>
4.1 Η γλώσσα προγραμματισμού Python .....	27
4.2 Δημιουργία λογαριασμού και API στο Twitter .....	28
4.2.1 Βιβλιοθήκες της Python που χρησιμοποιήθηκαν.....	31
4.3 Πρόσβαση στα δεδομένα με REST APIs και Streaming APIs.....	31
<b>5.ΔΗΜΙΟΥΡΓΙΑ ΕΡΓΑΛΕΙΟΥ ΣΥΛΛΟΓΗΣ TWEETS</b> .....	<b>32</b>
5.1 Περιγραφή εργαλείου συλλογής tweets .....	32
5.1.1 Ανάπτυξη εργαλείου .....	33
5.2 Αποτελέσματα εφαρμογής .....	36
<b>6.ΣΥΜΠΕΡΑΣΜΑΤΑ</b> .....	<b>40</b>
6.1 Ανασκόπηση.....	40
6..2 Επίτευξη στόχων .....	40
6.3 Πιθανές επεκτάσεις.....	41
<b>ΠΑΡΑΡΤΗΜΑ Α</b> .....	<b>444</b>
<b>ΒΙΒΛΙΟΓΡΑΦΙΑ</b> .....	<b>448</b>

Εργαλείο Ελέγχου Πληροφοριών σε Μέσα Κοινωνικής Δικτύωσης (Social Media) με εφαρμογή σε ειδικές ομάδες

## ΚΕΦΑΛΑΙΟ 1

### ΕΙΣΑΓΩΓΗ

Την τελευταία δεκαετία έχει παρατηρηθεί ραγδαία αύξηση της δραστηριότητας των χρηστών στα μέσα κοινωνικής δικτύωσης, που ως αποτέλεσμα έχει την ολοένα και αυξανόμενη χρήση του όγκου των δεδομένων που ανταλλάσσονται στα social media. Τα μέσα κοινωνικής δικτύωσης και οι εφαρμογές τους είναι πλέον ευρέως διαδεδομένα και οι ενεργοί χρήστες καθώς και ο χρόνος τον οποίο παραμένουν συνδεδεμένοι σε αυτές αυξάνεται συνεχώς σε όλες τις χώρες. Ο βαθμός με τον οποίο τα κοινωνικά δίκτυα επιδρούν στο ευρύ κοινό είναι μεγάλος καθώς εκτός από μέσο επικοινωνίας αποτελούν και άμεσο μέσο έκφρασης ιδεών, συναισθημάτων αλλά και αναπαραγωγής ειδήσεων και σχολιασμού.

Ένα από τα πιο δημοφιλή μέσα κοινωνικής δικτύωσης είναι το Twitter, το οποίο μετρά μηνιαίως 330 εκατομμύρια ενεργούς χρήστες κατά μέσο όρο μέχρι τα τέλη του 2017, ενώ οι εγγεγραμμένοι χρήστες ξεπερνούν το 1 δισεκατομμύριο παγκοσμίως. Το twitter είναι από τα πιο άμεσα και πιο εύκολα προσβάσιμο δίκτυο επικοινωνίας και ενημέρωσης. Για το λόγο αυτό, είναι διαχειρίσιμο στο κομμάτι της έρευνας που αφορά την γρήγορη και αποδοτική αξιοποίηση του μεγάλου όγκου πληροφοριών που δημοσιεύονται σε αυτό, που ως σκοπό έχουν την δημιουργία υπηρεσιών και γνώσεων προς το χρήστη.

Ο σκοπός της παρούσας πτυχιακής εργασίας είναι η διερεύνηση της τεχνολογίας ανάλυσης συναισθημάτων μέσα από την επεξεργασία των κειμένων που δημοσιεύονται στα social media. Εν συνεχεία θα αναπτυχθεί ένα παράδειγμα εργαλείου με το πλαίσιο εφαρμογής που μπορεί να έχει αυτό για μια συγκεκριμένη ειδική ομάδα όπως αυτή των ναυτικών.

Στο 2<sup>ο</sup> Κεφάλαιο γίνεται παρουσίαση των δημοφιλέστερων μέσων κοινωνικής δικτύωσης, αλλά και των νομικών κανόνων και νόμων που διέπουν στη προστασία των προσωπικών δεδομένων των χρηστών.

Στο 3<sup>ο</sup> Κεφάλαιο παρουσιάζονται γίνεται αναφορά στην ανάλυση συναισθημάτων ως κομμάτι της επεξεργασίας της φυσικής γλώσσας, ενώ γίνεται αναφορά στην κατηγοριοποίηση κειμένων.

Στο 4<sup>ο</sup> Κεφάλαιο παρουσιάζεται η γλώσσα προγραμματισμού python και οι μέθοδοι που χρησιμοποιεί για την άντληση πληροφοριών από το Twitter. Καθώς και οι βιβλιοθήκες που χρησιμοποιήθηκαν για την υλοποίηση του εργαλείου συλλογής και ανάλυσης συναισθημάτων αυτών.

Στο 5<sup>ο</sup> Κεφάλαιο γίνεται ανάλυση του προγράμματος που υλοποιήθηκε κατά την εκπόνηση της πτυχιακής εργασίας. Παρουσιάζονται με γραφήματα το ποσοστό των συναισθημάτων των tweet που συλλέχθηκαν και αφορούν μια ναυτιλιακή εταιρεία.

Τέλος, στα συμπεράσματα γίνεται μια γενική αποτίμηση της πτυχιακής εργασίας, βγαίνουν τα ανάλογα συμπεράσματα και γίνονται προτάσεις για περαιτέρω ανάπτυξη του συστήματος που εξετάζεται στο μέλλον.

## ΚΕΦΑΛΑΙΟ 2

### ΜΕΣΑ ΚΟΙΝΩΝΙΚΗΣ ΔΙΚΤΥΩΣΗΣ ΚΑΙ ΠΟΛΙΤΙΚΗ ΠΡΟΣΤΑΣΙΑΣ ΤΗΣ ΙΔΙΩΤΙΚΟΤΗΤΑ ΣΕ ΑΥΤΑ

#### 2.1 Βασικές γνώσεις

Η ιστορία του internet ξεκινάει από τα τέλη της δεκαετίας του 60 στις ΗΠΑ. Όπου ο στρατός των ΗΠΑ αναπτύσσει πειραματικά το δίκτυο ARPANET (αρχικά) που ως αποτέλεσμα έχει την εξέλιξη της επιστήμης και της τεχνολογίας. Το **1969** κατασκευάζεται το πρώτο δίκτυο που συνδέει καλωδιακά 4 κόμβους (Πανεπιστήμιο της Καλιφόρνια-Λος Άντζελες, Πανεπιστήμιο SRI (Στανφορντ), Πανεπιστήμιο Σαντα Μπαρμάρκ Καλιφόρνια, Πανεπιστήμιο της Γιούτα), με ταχύτητα 50 Kbps. Στις 19 Οκτώβρη 1969 πραγματοποιείται η πρώτη απόπειρα να σταλθεί το πρώτο μήνυμα από ηλεκτρονικό υπολογιστή, η προσπάθεια αυτή γίνεται από το Πανεπιστήμιο της UCLA που προσπαθεί να στείλει τη λέξη "Log in " στο Πανεπιστήμιο του Stamford, μια προσπάθεια η οποία πέφτει στο κενό αφού το σύστημα καταρρέει. Το **1972** το Arpanet έχει γίνει διεθνές και απαριθμεί πάνω από 2000 χρήστες. Το **1973** σχεδιάζεται το Πρωτόκολλο Ελέγχου Μετάδοσης TCP που ως σκοπό έχει να ενωθούν όλο και περισσότεροι υπολογιστές με στόχο να επικοινωνούν μεταξύ τους. Στη συνέχεια και μέσα από έρευνα δημιουργείται μια νέα τεχνική το Internet Protocol (IP:Πρωτόκολλο Διαδικτύωσης), από την οποία θα πάρει αργότερα το όνομα του το σημερινό σε εμάς internet. Το **1974** οι Vinton Cerf και Bob Kahn χρησιμοποιούν σε ένα έγγραφό τους για το Πρωτόκολλο Ελέγχου Μετάδοσης για πρώτη φορά τη λέξη Internet. Το **1978** το TCP πρωτόκολλο χωρίζεται στο TCP και IP. Το **1982** τα TCP και IP καθορίζονται ως ο συνδυασμός πρωτοκόλλων του Arpanet. Αυτό έχει ως απόρροια να γίνεται για πρώτη φορά λόγος στον όρο internet ως ένα συνδεδεμένο σύνολο από δίκτυα που χρησιμοποιούν το TCP/IP. Το **1984** υλοποιείται το Domain Name System (DNS) στο οποίο καταγράφονται 1000 κεντρικοί κόμβοι. Έτσι οι υπολογιστές αναγνωρίζονται από διευθύνεις ip κωδικοποιημένων αριθμών. Το **1988** με τη δημιουργία εθνικών και άλλων δικτύων που συνδέονται στο διαδίκτυο έχουμε την εξάπλωση του. Ταυτόχρονα έχουμε και την εμφάνιση του πρώτου worm το οποίο



επηρεάζει τη λειτουργία 6000 από τους 60.000 εξυπηρετητές του δικτύου. Το **1990** σταματάει η λειτουργία του Arpanet καθώς οι περισσότεροι κόμβοι του διαδικτύου έχουν ήδη συνδεθεί με άλλα δίκτυα. Το **1991** η Ελλάδα συνδέεται για πρώτη φορά στο διαδίκτυο μέσω του NSFNET. Το **1992** το CERN θέτει σε λειτουργία το World Wide Web (WWW). Ο παγκόσμιος ιστός με τη σειρά του έφερε το *HTML* και το *e-mail* καθώς όμως η τεχνολογία εξελισσόταν είδαμε να κάνουν την εμφάνιση τους νέες υπηρεσίες και μπήκαμε σε μια **νέα εποχή (web 2.0, Google, skype, twitter, facebook, wikipedia κ.α.)**, όπου οι χρήστες δεν περιορίζονται στη παθητική παρακολούθηση των πληροφοριών που τους παρέχονται μέσω των ιστοσελίδων, έχουν τη δυνατότητα να επικοινωνούν μεταξύ τους μέσω των ιστότοπων κοινωνικής δικτύωσης, να χρησιμοποιούν διαδικτυακές εφαρμογές κ.α., με αποτέλεσμα να συνεισφέρουν στο περιεχόμενο του ιστοχώρου.

Οι υπηρεσίες που προσφέρει το internet σήμερα έχουν ως αποτέλεσμα όλο και περισσότεροι άνθρωποι ανεξαρτήτως ηλικίας να ξοδεύουν ένα αρκετά μεγάλο μέρος του χρόνου τους καθημερινά μέσα στο διαδίκτυο, είτε λόγο δουλειάς ή για εκπαιδευτικούς λόγους ή ακόμα και για λόγους ψυχαγωγίας. Τη τελευταία δεκαετία ραγδαία είναι η ανάπτυξη των μέσων κοινωνικής δικτύωσης και εκατομμύρια άνθρωποι είναι ενεργοί χρήστες τους. Παρακάτω παρουσιάζονται κάποια από τα δημοφιλέστερα μέσα κοινωνικής δικτύωσης και η επιρροή των μέσων αυτών στον άνθρωπο.

## 2.2 Εισαγωγή στα Social Media

Κοινωνική δικτύωση είναι η συμμετοχή των ανθρώπων σε συγκεκριμένες ομάδες. Ο όρος κοινωνικό δίκτυο έχει παραπάνω από έναν ορισμούς. Ως κοινωνικό δίκτυο μπορούμε να ορίσουμε:

- πολυδιάστατα συστήματα επικοινωνίας και διαμόρφωσης της ανθρώπινης πρακτικής και της κοινωνικής ταυτότητας.
- Το σύνολο των προσωπικών επαφών μέσω των οποίων το άτομο δημιουργεί και συντηρεί την κοινωνική του ταυτότητα, λαμβάνει συναισθηματική υποστήριξη, υλική ενίσχυση και συμμετοχή στις υπηρεσίες, έχει πρόσβαση στις πληροφορίες και δημιουργεί νέες κοινωνικές επαφές.

Το πρώτο κοινωνικό δίκτυο στο οποίο εντάσσεται το άτομο είναι αυτό της οικογένειας και στη συνέχεια δημιουργεί νέα όπως αυτό των φίλων, των γνωστών, των συναδέλφων. Άρα ένα κοινωνικό δίκτυο περιλαμβάνει τρεις σημαντικές έννοιες:

1. Το εύρος ή το μέγεθος, το οποίο αναφέρεται στον αριθμό των ατόμων που συμμετέχουν στο δίκτυο
2. Το ποσοστό συμμετοχής μελών του ευρύτερου κύκλου (οικογένεια, φίλοι κ.α) στο δίκτυο και
3. Το πόσο συχνά τα μέλη ενός κοινωνικού δικτύου αλληλοεπιδρούν μεταξύ τους.

Στις μέρες μας τα κοινωνικά δίκτυα δημιουργούνται και μέσα σε ιστότοπους (online) και λειτουργούν ως online κοινότητες των χρηστών του Διαδικτύου. Ανάλογα με τα χαρακτηριστικά του ιστότοπου, τα μέλη της online κοινότητας έχουν κοινά ενδιαφέροντα και χαρακτηριστικά (χόμπι, επάγγελμα, θρησκεία πολιτική κ.α). Η εγγραφή του ατόμου σε ένα μέσο κοινωνικής δικτύωσης

### **2.2.1 Δημοφιλέστερα Social Media**

Το **Facebook** ξεκίνησε τη λειτουργία του το 2004, όταν ιδρύθηκε από τον φοιτητή του Harvard Mark Zuckerberg. Αρχικός του στόχος, ήταν να αποτελέσει έναν χώρο κοινωνικής δικτύωσης, στον οποίο θα είχαν δικαίωμα συμμετοχής μόνο οι φοιτητές του Harvard. Οι χρήστες της υπηρεσίας μπορούσαν να εγγραφούν δωρεάν και να δημιουργήσουν προσωπικό λογαριασμό χρήστη. Κατόπιν, μπορούσαν να επικοινωνούν μέσω μηνυμάτων με τις επαφές τους, να προσθέτουν και να ανανεώνουν τις προσωπικές πληροφορίες τους και να δημιουργούν ομάδες για οποιοδήποτε θέμα. Το εντυπωσιακό στοιχείο σε σχέση με το Facebook έχει να κάνει με το ρυθμό αύξησης των εγγραφών. Καθημερινά προστίθενται 600.000 νέοι χρήστες στους ήδη υπάρχοντες. Σήμερα, το Facebook έχει 500 εκατομμύρια μοναδικούς ενεργούς χρήστες. Συνολικά κάθε μήνα ανεβαίνουν περίπου 700 εκατομμύρια φωτογραφίες και 4 εκατομμύρια βίντεο. Όσον αφορά στο γεωγραφικό προσδιορισμό, το 30% της κίνησης προέρχεται από τις ΗΠΑ και το υπόλοιπο 70% από τις άλλες χώρες του πλανήτη. Κατά μέσο όρο, κάθε χρήστης της υπηρεσίας έχει πάνω από 100 φίλους, ενώ ο συνολικός χρόνος παραμονής στο site ανέρχεται σε 2,6 δισεκατομμύρια λεπτά ανά ημέρα.

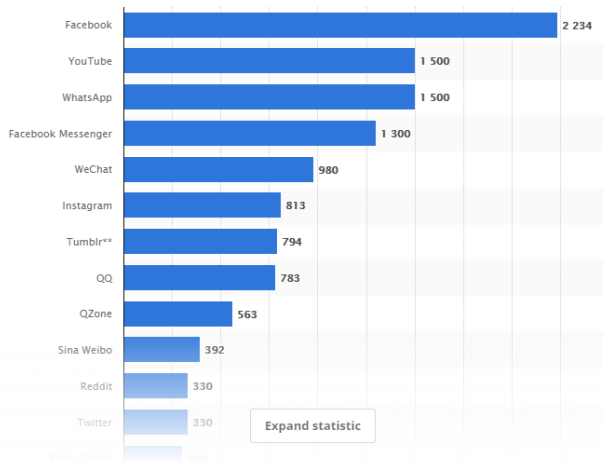
Το **Twitter** δημιουργήθηκε το 2006 από τους Jack Dorsey, Noah Glass, Biz Stone και Evan Williams έχει ήδη περίπου 330 εκατομμύρια ενεργούς χρήστες, οι οποίοι δημοσιεύουν καθημερινά 65 εκατομμύρια μηνύματα. Τα μηνύματα είναι μικρού μεγέθους, αρχικά ήταν μέχρι 140 χαρακτήρες αλλά από το 2017 οι χρήστες μπορούν να εισάγουν μέχρι 280 χαρακτήρες . Τα μέλη μπορούν να ακολουθούν τα μηνύματα άλλων χρηστών καθώς και να απαντούν σε αυτά. Το δίκτυο έχει απαγορευθεί στην Κίνα (λογοκρισία). Το δίκτυο προσφέρει στους χρήστες του τη δυνατότητα να αναφέρουν ενόχληση από άλλα μέλη του δικτύου, των οποίων ο λογαριασμός μπορεί να ανασταλεί γι' αυτό το λόγο. Όταν ένας χρήστης δημοσιεύσει ένα μήνυμα, δεν είναι δυνατόν να το αλλάξει, αλλά μόνο να το διαγράψει.

Το **Instagram** το όνομα του προέρχεται από τις λέξεις **Instast** που σημαίνει στιγμιαίο και **telegram** που σημαίνει τηλεγράφημα. Δημιουργήθηκε τον Οκτώβριο του 2010 από τους Kevin Systrom και Mike Krieger, αποφοίτους του πανεπιστημίου του Stanford και το 2012 αγοράστηκε από το Facebook. Μέσα σε μόλις δυο μήνες από τη δημιουργία του (Δεκέμβριος 2010), έφτασε να μετράει 1.000.000 εγγεγραμμένους χρήστες. Σήμερα η εφαρμογή μετράει πάνω από 200 εκατομμύρια χρήστες και 20 δισεκατομμύρια φωτογραφίες από όλο το κόσμο. Το κύριο χαρακτηριστικό της εφαρμογής είναι η δημοσίευση φωτογραφιών και βίντεο. Ο χρήστης μπορεί είτε να τραβήξει με το κινητό του μια φωτογραφία ή βίντεο, είτε να χρησιμοποιήσει μια ήδη υπάρχουσα φωτογραφία ή ένα υπάρχων βίντεο και στις δυο περιπτώσεις πριν τη κοινοποίησή τους μπορεί να τα επεξεργαστεί με ειδικά εργαλεία που του προσφέρει η εφαρμογή και να εφαρμόσει ειδικά σχεδιασμένα φίλτρα. Στη συνέχεια μπορεί να προχωρήσει στη κοινοποίησή τους και να επισημάνει άλλους χρήστες (tag), να προσθέσει περιγραφή και τοποθεσία και να τα κοινοποιήσει απευθείας στο λογαριασμό του σε Facebook, Twitter, Fourquare και άλλα social media που συνεργάζονται με την συγκεκριμένη εφαρμογή.

Το **LinkedIn** αποτελεί ένα δικτυακό τόπο κοινωνικής δικτύωσης που εστιάζει κυρίως σε επιχειρηματικές δραστηριότητες. Δίνει τη δυνατότητα στους χρήστες να δημιουργήσουν ένα δίκτυο από υπάρχουσες και νέες επαγγελματικές επαφές. Όμως, προχωρά και σε άλλες δράσεις. Συγκεκριμένα όταν μια επιχείρηση προσθέτει μια νέα σύνδεση, οι συνδέσεις αυτής και οι συνδέσεις αυτών των συνδέσεων επίσης προστίθενται στο δίκτυο. Αυτό προσφέρει σ' έναν επαγγελματία μεγάλο πεδίο επαγγελματικών επαφών . Σύμφωνα με τη Wikipedia, το LinkedIn έχει περισσότερους από 30 εκατομμύρια χρήστες και το επισκέπτονται περισσότεροι από 3,2 εκατομμύρια επισκέπτες κάθε μήνα σε ένα συνολικό αριθμό από περισσότερες από 150 διαφορετικές επιχειρήσεις.

Το **Google plus (Google+)** είναι πολύ απλά ένα νέο κοινωνικό δίκτυο από την μεγαλύτερη εταιρεία διαδικτυακών υπηρεσιών το οποίο φέρει και το όνομα της. Η Google μετά από κάποιες αποτυχημένες προσπάθειες κοινωνικών δικτύων (π.χ. Buzz, Wave κτλ.) δημιούργησε το Google plus με στόχο να εδραιωθεί και να ανταγωνιστεί άλλα μεγάλα κοινωνικά δίκτυα όπως το Facebook, Twitter κτλ. Η βασική δραστηριότητα του είναι η επικοινωνία των χρηστών και η ανταλλαγή πληροφοριών. Μπορείτε να προσθέσετε τους φίλους σας, να επικοινωνήσετε μαζί τους, να κάνετε βίντεο κλήση, να μοιραστείτε φωτογραφίες και πληροφορίες και πολλά άλλα. Υπάρχει η δυνατότητα διαφοροποίησης του περιεχομένου που μοιράζεις ανάλογα στο "circle"(ομάδα-group χρηστών σαν και τις λίστες φίλων που μπορείς να φτιάξεις στο facebook).

**Most popular social networks worldwide as of April 2018, ranked by number of active users (in millions)**



**Εικόνα 2.1:**Most popular social networks

### 2.3 Πολιτική Προστασίας της Ιδιωτικότητας στα Μέσα Κοινωνικής Δικτύωσης.

Η δημιουργία νέου λογαριασμού σε κάποιο από τους προαναφερθέντες ιστοτόπους μέσων κοινωνικής δικτύωσης προϋποθέτει τη συναίνεση του χρήστη στους όρους χρήσης και στη πολιτική προστασίας της ιδιωτικότητας και των προσωπικών δεδομένων. Πρέπει ο χρήστης πριν συμφωνήσει με τα προαναφερθέντα να έχει ενημερωθεί για το σκοπό και για το τρόπο επεξεργασίας των προσωπικών πληροφοριών που θα διαθέσει στον ιστότοπο.

Κάθε εγγεγραμμένο μέλος είναι υπεύθυνο για την ορθότητα και τον έλεγχο των προσωπικών πληροφοριών που έχει δηλώσει στο διαδίκτυο, όπως για παράδειγμα ποιες πληροφορίες καταχωρούνται στο προφίλ και ποιος έχει πρόσβαση σε αυτές.

Τι ορίζονται προσωπικά δεδομένα και τι νόμοι και κανόνες ισχύουν για την προστασία αυτών αναλύεται ακολούθως.

Τα προσωπικά δεδομένα διαχωρίζονται σε:

- Δεδομένα προσωπικού χαρακτήρα:
  - ο Στοιχεία αναγνώρισης (ονοματεπώνυμο, ηλικία, οικογενειακή κατάσταση κ.ο.κ)

- Φυσικά χαρακτηριστικά (χρώμα δέρματος, ύψος, χρώμα μαλλιών, ματιών κ.α)
- Εκπαίδευση
- Εργασία
- Οικονομική κατάσταση
- Ενδιαφέροντα
- Δραστηριότητες
- Συνήθειες
- Ευαίσθητα προσωπικά δεδομένα:
  - Φυλετική ή εθνική προέλευση
  - Πολιτικά φρονήματα (Συμμετοχή σε συνδικαλιστική οργάνωση)
  - Θρησκευτικές οι φιλοσοφικές πεποιθήσεις
  - Υγεία και κοινωνική πρόνοια
  - Ποινικές διώξεις ή καταδίκες
  - Ερωτική ζωή

Στην Ελλάδα, όπως και στις υπόλοιπες Ευρωπαϊκές χώρες υπάρχει ειδική νομοθεσία που προστατεύει τους χρήστες από την ανεξέλεγκτη χρήση των δεδομένων τους. Στην χώρα μας υπεύθυνη είναι η Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα για την εφαρμογή του της νομοθεσίας (νόμοι 2472/1997 «Ειδική νομοθεσία για την προστασία των προσωπικών δεδομένων» και 3471/2006 «Προστασία δεδομένων προσωπικού χαρακτήρα και της ιδιωτικής ζωής στον τομέα των ηλεκτρονικών επικοινωνιών»).

Για τη δημιουργία λογαριασμού σε οποιοδήποτε Μέσο Κοινωνικής Δικτύωσης ο χρήστης πρέπει να δηλώσει κάποιες προσωπικές πληροφορίες σχετικά με αυτόν. Τις περισσότερες φορές για την δημιουργία του προφίλ ζητείται να δηλώσει το όνομα και το επίθετο του, προαιρετικά τηλέφωνο σταθερό ή κινητό καθώς και περιοχή διαμονής, πληροφορίες σχετικές με την ερωτική του ζωή (αν είναι σε σχέση ή όχι), ενδιαφέροντα και χόμπι. Αυτές είναι κάποιες από τις πληροφορίες που καλείται να αναφέρει ώστε να αποκτήσει ταυτότητα στο εκάστοτε ιστότοπο κοινωνικής δικτύωσης, για να διαμορφώσει ένα πλήρες προφίλ μπορεί να ανεβάσει κάποια προσωπική του φωτογραφία ή βίντεο να κοινοποιήσει την παρουσία του σε οποιαδήποτε τοποθεσία έχει επισκεφτεί μόνος ή με παρέα να

αναφέρει τα άτομα που είναι μαζί (tag). Μπορεί να παρουσιάσει τις απόψεις και τις συνήθειες που έχει, να γίνει μέλος σε κάποια ομάδα ή ακόμα και ο ίδιος να δημιουργήσει κάποια ομάδα. Εξάλλου στις μέρες μας μεγαλώνει το ποσοστό των ανθρώπων που χρησιμοποιούν τα Μέσα Κοινωνικής Δικτύωσης για την προβολή της δουλειάς τους και για την προώθηση των προϊόντων τους (Social Media Marketing).

Έχοντας αντιληφθεί την απήχηση που έχουν στον άνθρωπο το Μάρτιο του 2008 οι Αρχές Προστασίας Δεδομένων της Ευρωπαϊκής Ένωσης και η Διεθνής Ομάδα για την Προστασία των Προσωπικών Δεδομένων στις Τηλεπικοινωνίες (IWGDPT) εκδίδουν ένα έγγραφο συστάσεων για τις υπηρεσίες κοινωνικής δικτύωσης, το οποίο απευθύνεται στους παρόχους αλλά και στους χρήστες των υπηρεσιών αυτών. Το έγγραφο αυτό (για την διαμόρφωση του οποίου πήραν μέρος και οι Ελληνικές Αρχές) προτείνει ενδεικτικά κάποια από τα παρακάτω μέτρα για τους παρόχους:

- Οι χρήστες πρέπει να πληροφορούνται μέσα από κείμενο που θα περιέχει τους όρους και τις προϋποθέσεις καθώς και την πολιτική ιδιωτικότητας του παρόχου καθώς και επιμέρους σημεία κατά την χρήση των υπηρεσιών που προσφέρει (ιδιωτικότητα κατά τη δημιουργία ενός φωτογραφικού άλμπουμ). Οι πάροχοι πρέπει να παρέχουν στο χρήστη την επιλογή να προβεί σε ενημέρωση ή διαγραφή προσωπικών του δεδομένων.
- Οι πάροχοι πρέπει να ενημερώνουν τους χρήστες για κινδύνους που μπορεί να κρύβει η δημοσίευση προσωπικών πληροφοριών και φωτογραφιών στο διαδίκτυο. Καθώς και να προειδοποιούν πως είναι παράνομο να αναδημοσιεύουν πληροφορίες τρίτων χωρίς τη συγκατάθεση του. Επίσης κάθε σελίδα κοινωνικής δικτύωσης πρέπει να αναφέρει πως τα δεδομένα του εκάστοτε χρήστη μπορούν να χρησιμοποιηθούν για σκοπούς στοχευμένης διαφήμισης.
- Πρέπει να παρέχουν στους χρήστες τη δυνατότητα να μπορούν να ελέγξουν ποιος μπορεί να έχει πρόσβαση στις δημοσιευμένες πληροφορίες ή ποιος μπορεί να δημοσιεύσει δικές τους κοινοποιήσεις. Καθώς και τη δυνατότητα οι ομάδες να μπορούν να αρνηθούν ή να διαγράψουν μέλη όταν το κρίνουν σκόπιμο.

- Πρέπει να υπάρχει μηχανισμός στον οποίο οι χρήστες θα μπορούν να παραπονεθούν για τυχόν προβλήματα που αντιμετωπίζουν ή για άλλους χρήστες σε περίπτωση που υπάρξουν προβλήματα μεταξύ τους. Ο πάροχος θα πρέπει να είναι σε θέση να λάβει τα απαραίτητα μέτρα, όπως τη διαγραφή κάποιου λογαριασμού που δεν συμμορφώνεται στους κανονισμούς του ιστότοπου.

Επίσης, για τους χρήστες των μέσων κοινωνικής δικτύωσης ενδεικτικά συνίσταται:

- Να είναι προσεκτικοί στη δημοσίευση των προσωπικών και ευαίσθητων δεδομένων, καθώς από την στιγμή που δημοσιεύονται είναι αυτομάτως διαθέσιμα σε μεγάλο αριθμό ατόμων που χρησιμοποιούν το διαδίκτυο.
- Να παίρνουν πρώτα τη συγκατάθεση από άλλους χρήστες πριν προβούν στην δημοσίευση προσωπικών τους δεδομένων.
- Να χρησιμοποιούν τις ρυθμίσεις που ορίζουν την ιδιωτικότητα του προφίλ τους, ώστε να μπορούν να ελέγχουν ποιος μπορεί να έχει πρόσβαση στο προφίλ τους ή να περιορίσουν την διαθεσιμότητα των δεδομένων τους σε μηχανές αναζήτησης.
- Να χρησιμοποιούν διαφορετικούς κωδικούς για κάθε ιστότοπο που χρησιμοποιούν και να τους ανανεώνουν ανά τακτά χρονικά διαστήματα.

Στη πτυχιακή θα αναπτυχθεί ένα εργαλείο ελέγχου πληροφοριών μέσων κοινωνικής δικτύωσης για ειδικές ομάδες όπως είναι η αστυνομία, ο στρατός, η ναυτιλία, οι ηλικιωμένοι κ.α.

Στις μέρες μας τα μέσα κοινωνικής δικτύωσης χρησιμοποιούνται από όλες τις κοινωνικές ομάδες, από μεγάλο αριθμό ανθρώπων ανεξαρτήτως επαγγέλματος και όλες τις ηλικίες από έφηβους μέχρι και ανθρώπους της τρίτης ηλικίας. Έχουν μεγάλη απήχηση και πολλές φορές οι άνθρωποι εκφράζουν την άποψη τους και τα συναισθήματα τους ή παίρνουν θέση και λένε τη γνώμη τους για πολιτικά ζητήματα και θέματα που απασχολούν τη καθημερινότητα. Δεν είναι λίγες οι φορές που η δίωξη ηλεκτρονικού εγκλήματος της χώρας μας έχει αποτρέψει εγκλήματα ή αυτοκτονίες χρηστών που κάνουν χρήση των μέσων κοινωνικής



δικτύωσης ή έχει βρει ίχνη (ηλεκτρονικά) καταζητούμενων και έχει προβεί η αστυνομία στη σύλληψη τους.

Αν και στη χώρα μας η νομοθεσία εμφανίζει κενά σχετικά με την προστασία των προσωπικών δεδομένων μέσα στο διαδίκτυο και τις ποινικές κυρώσεις που επιβάλλονται όταν κάποιος δεν συμμορφώνεται με τους κανονισμούς. Όπως προαναφέρθηκε οι νόμοι που έχουν θεσπιστεί είναι νόμοι 2472/1997 «Ειδική νομοθεσία για την προστασία των προσωπικών δεδομένων» και 3471/2006 «Προστασία δεδομένων προσωπικού χαρακτήρα και της ιδιωτικής ζωής στον τομέα των ηλεκτρονικών επικοινωνιών». Πάνω σε αυτούς προστίθενται ή ακυρώνονται διατάξεις και οδηγίες που προέρχονται από κανόνες της Ευρωπαϊκής Ένωσης.

Για τις ειδικές ομάδες που θα ασχοληθεί η πτυχιακή δεν υπάρχει κάποια επίκαιρη νομοθεσία, που να θεσπίζει τη χρήση των social media από τις ομάδες αυτές.

Πρέπει να σημειωθεί πως τα δεδομένα που θα συλλεχθούν έχουν ως μοναδικό σκοπό την επεξεργασία τους για τη πτυχιακή και όχι την αλλοίωση τους, ούτε και την δημοσίευσή για προσβλητικούς λόγους.

## ΚΕΦΑΛΑΙΟ 3

### ΑΝΑΛΥΣΗ ΣΥΝΑΙΣΘΗΜΑΤΩΝ

#### 3.1 Γενικά για την ανάλυση συναισθημάτων

Με την ραγδαία εξάπλωση των μέσων κοινωνικής δικτύωσης και της καθημερινής τους χρήσης από το κοινωνικό σύνολο, όλο και περισσότερο δίνεται έμφαση στη συλλογή και στη κατανόηση των συναισθημάτων που κρύβονται πίσω από μια δημοσίευση σε κάποιο από τα social media. Αυτό είχε ως αποτέλεσμα να έρθουν στο προσκήνιο πρωτοποριακοί αλγόριθμοι για την ανάλυση συναισθημάτων. Οι αλγόριθμοι αυτοί χρησιμοποιούν την επεξεργασία φυσικής γλώσσας, τις στατιστικές και την ανάλυση κειμένου για να εξαγάγουν και να προσδιορίσουν το συναίσθημα του κειμένου σε θετικές, αρνητικές ή ουδέτερες κατηγορίες.

#### 3.2 Επεξεργασία Φυσικής γλώσσας (NLP)

Η επεξεργασία φυσικής γλώσσας είναι ένα πεδίο της επιστήμης Η/Υ, της τεχνητής νοημοσύνης και της υπολογιστικής γλωσσολογίας που αφορά την αλληλεπίδραση που υπάρχει ανάμεσα στους υπολογιστές και στην ανθρώπινη γλώσσα που χρησιμοποιούμε. Σκοπός, της επιστήμης αυτής είναι να κατανοήσει την ανθρώπινη γλώσσα δίνοντας στη συνέχεια τη δυνατότητα στα υπολογιστικά συστήματα να εξαγάγουν συμπεράσματα όταν δέχονται ως είσοδο την φυσική γλώσσα.

Στις μέρες οι αλγόριθμοι NLP βασίζονται στην μηχανική μάθηση (machine learning) και εστιάζουν περισσότερο στα στατιστικά μοντέλα τα οποία λαμβάνουν αποφάσεις βασισμένες σε πιθανότητες και στον υπολογισμό και ορισμό πραγματικών τιμών (βάρη) στην είσοδο που δέχονται.

Αναφορικά, θα παρουσιαστούν κάποιοι τομείς έρευνας της επιστήμης NLP.

- **Αυτόματη περίληψη- automatic summarization**  
Από ένα κομμάτι κειμένου που δέχεται ως είσοδο, συνήθως χρησιμοποιούνται άρθρα ή οικονομικές αναλύσει παράγει μία περίληψη τους.
- **Παραγωγή φυσικής γλώσσας- natural language generation**  
Παίρνει τη πληροφορία από μια βάση δεδομένων και τη μετατρέπει σε γλώσσα κατανοητή για τον άνθρωπο.
- **Οπτική αναγνώριση χαρακτήρων- optical character recognition**  
Από μια εικόνα που δέχεται ως είσοδο γίνεται αναγνώριση του κειμένου το οποίο απεικονίζεται με χαρακτήρες.
- **Ανάλυση συναισθημάτων- sentiment analysis**  
Από ένα σύνολο κειμένων που δέχεται ως είσοδο γίνεται εξαγωγή μιας υποκειμενικής πληροφορίας που βασίζεται στον καθορισμό της «πολικότητας» συγκεκριμένων ζητημάτων. Αποτελεί, σημαντικό εργαλείο, διότι είναι ένας δείκτης που ορίζει τις τάσεις τις κοινής γνώμης στα social media.
- **Αναγνώριση ομιλίας -speech recognition**  
Ως είσοδο δέχεται ένα φωνητικό ή ηχητικό ντοκουμέντο και το μετατρέπει σε λεκτική αναπαράσταση. Αυτό είναι κάτι αρκετά δύσκολο, αν σκεφτεί κανείς ότι στην ομιλία οι παύσεις μπορεί να είναι πολύ μικρές και πολλές φορές καθώς μιλάμε δεν υπάρχουν καθόλου κενά ανάμεσα στις λέξεις.

### **3.3 Ανάλυση συναισθημάτων (Sentiment Analysis)**

Η ανάλυση συναισθημάτων στοχεύει στο να εξακριβώσει το ύφος του ανθρώπου που παράγει ένα κείμενο σχετικά με ένα συγκεκριμένο θέμα ή στη συνολική πολικότητα του περιεχομένου του κειμένου. Το ύφος μπορεί να είναι η συναισθηματική κατάσταση στην οποία βρίσκεται το πρόσωπο που το γράφει εκείνη τη στιγμή ή το συνολικό συναίσθημα που αποπνέει στον αναγνώστη όταν το διαβάζει. Για το λόγο αυτό εφαρμόζεται σε δημοσιεύσεις στα μέσα κοινωνικής δικτύωσης που ως σκοπό έχουν να παρακολουθούν την τάση της κοινής γνώμης.

Η πολικότητα ενός κειμένου σε επίπεδο εγγράφου, πρότασης ή χαρακτηριστικού, δηλαδή αν η γνώμη που εκφράζεται στο συγκεκριμένο τμήμα

που αναλύεται είναι θετική, αρνητική ή ουδέτερη είναι από τις κυρίες εφαρμογές της ανάλυσης συναισθημάτων. Ωστόσο, πολιτισμικοί παράγοντες, γλωσσολογικές ιδιαιτερότητες καθιστούν εξαιρετικά δύσκολη την μετατροπή ενός τμήματος κειμένου σε κάποιο απλό θετικό ή αρνητικό συναίσθημα. Το γεγονός ότι ακόμα και οι άνθρωποι συχνά δεν συμφωνούν για το συναίσθημα των κειμένων, αποδεικνύει το πόσο δύσκολο είναι για ένα υπολογιστή να το εξάγει σωστά. Όσο πιο σύντομο είναι το κείμενο, τόσο πιο δύσκολο γίνεται.

Αν και τα μικρά σε μήκος κείμενα είναι δύσκολο να βγάλουν ένα σαφές αποτέλεσμα, η ανάλυση συναισθημάτων σε ότι έχει να κάνει με τα κείμενα που δημοσιεύονται στο twitter έχει δώσει αρκετά αξιόπιστα αποτελέσματα σε ότι έχει να κάνει με το marketing ή τις πολιτικές απόψεις.

### 3.4 Κατηγοριοποίηση κειμένου (Text classification)

Η ανάλυση συναισθημάτων αξιοποιεί βασικές μεθόδους κατηγοριοποίησης κειμένου για την αναγνώριση και εξαγωγή πληροφορίας από διάφορες πηγές. Η κατηγοριοποίηση κειμένου είναι μια διαδικασία στη διάρκεια της οποίας δίνονται προκαθορισμένες κατηγορίες σε έγγραφα κειμένου. Παράδειγμα, σε αυτό είναι στα emails όταν διαχωρίζονται σε spam και non-spam.

Σε αντίθεση με την δημιουργία αυτόματων κανόνων η στατιστική κατηγοριοποίηση κειμένου χρησιμοποιεί μεθόδους μηχανικής μάθησης. Ένα κείμενο αναπαρίσταται ως ένα διάνυσμα χαρακτηριστικών:

$$\hat{x} = (x^{(1)}, \dots, x^{(p)})$$

τα χαρακτηριστικά  $x^{(i)}$  είναι συνήθως η συχνότητα εμφάνισης λέξεων ή διαφόρων ν-γραμμάτων που υπάρχουν μέσα στις λέξεις, στις φράσεις ή στις οντότητες κ.λ.π που αποτελούν ένα κείμενο. Για τον υπολογισμό των τιμών των χαρακτηριστικών αυτών για ένα συγκεκριμένο κείμενο d είναι η μέθοδος “bag of words”.

Μια επιμέρους προσέγγιση είναι η TF-IDF (term frequency – inverse document frequency), “συχνότητα όρου – αντίστροφη συχνότητα εγγράφου”, που βασίζεται στην ανάθεση ενός βάρους για κάθε όρο. Στην πιο απλή μορφή της παριστάνεται ως:

$$x^{(i)} = \mathbf{TF}(i, d) \cdot \mathbf{IDF}(i)$$

Όπου:

- **TF(i,d)**- term frequency, είναι ο αριθμός των εμφανίσεων του όρου  $i$  στο έγγραφο  $d$ .
- **IDF(i)= $\log(N/n_i)$** , είναι η inverse document frequency, με το  $N$  να δηλώνει τον συνολικό αριθμό των εγγράφων σε μια συλλογή, και το  $n_i$  τον αριθμό των εγγράφων που περιέχουν τον όρο  $i$ .

## ΚΕΦΑΛΑΙΟ 4

### Η ΓΛΩΣΣΑ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ PYTHON

Στη παρούσα πτυχιακή θα αναπτυχθεί ένα εργαλείο εξόρυξης δεδομένων από το Twitter. Για τις ανάγκες ανάπτυξης του χρησιμοποιήθηκε η γλώσσα προγραμματισμού Python 3.6.1 καθώς και βιβλιοθήκες της.

#### 4.1 Η γλώσσα προγραμματισμού Python

Η Python δημιουργήθηκε από τον Ολλανδό Guido van Rossum το 1990 και αποτελεί μια υψηλού επιπέδου γλώσσα προγραμματισμού. Το όνομα της γλώσσας προέρχεται από την ομάδα άγγλων κωμικών Μόντυ Πάιθον.

Ο κύριος στόχος της είναι η αναγνωσιμότητα του κώδικά της και η ευκολία χρήσης της και το συντακτικό της επιτρέπει στους προγραμματιστές να εκφράσουν έννοιες σε λιγότερες γραμμές κώδικα απ' ό,τι θα ήταν δυνατόν σε γλώσσες όπως η C++ ή η Java. Διακρίνεται λόγω του ότι έχει πολλές βιβλιοθήκες που διευκολύνουν ιδιαίτερα αρκετές συνηθισμένες εργασίες και για την ταχύτητα εκμάθησής της. Ένα ιδιαίτερο χαρακτηριστικό της γλώσσας είναι η χρήση κενών διαστημάτων (whitespace) για τον διαχωρισμό των συντακτικών δομών που προγράμματος, σε αντίθεση με την πρακτική σε άλλες γλώσσες όπου για τον ίδιο σκοπό χρησιμοποιούνται ειδικά σύμβολα (πχ αγκύλες). Αυτό, σε συνδυασμό με το ότι χρησιμοποιεί πλήρεις αγγλικές λέξεις στη θέση συμβόλων, καθιστούν τον κώδικα της Python ευανάγνωστο από όσους έχουν βασική γνώση των αγγλικών.

Για τη συγγραφή προγραμμάτων είναι απαραίτητος ένας κειμενογράφος ή ακόμα καλύτερα ένα ολοκληρωμένο περιβάλλον ανάπτυξης (Integrated Development Environment - IDE), το οποίο είναι ένα ειδικό λογισμικό για την ανάπτυξη εφαρμογών. Η Python έρχεται μαζί με ένα εύχρηστο περιβάλλον ανάπτυξης με την ονομασία IDLE. Τα αρχικά του έρχονται από τις λέξεις Interactive DeveLopment Environment και είναι γραμμένο σε Python από τον Guido van Rossum. Χρησιμοποιεί τη βιβλιοθήκη γραφικών Tkinter, οπότε μπορεί να εκτελεσθεί σε περιβάλλον Linux, Windows και Mac OS X. Το IDLE μας δίνει τη

δυνατότητα να χρησιμοποιήσουμε διαδραστικά τον διερμηνευτή της γλώσσας, να γράψουμε και να επεξεργαστούμε προγράμματα, να τα αποθηκεύσουμε σε αρχεία, να τα εκτελέσουμε, να κάνουμε

#### **4.2 Δημιουργία λογαριασμού και API στο Twitter**

Για τη συλλογή δεδομένων από το Twitter δημιουργήθηκε ένας λογαριασμός σε αυτό, ο οποίος χρησιμοποιήθηκε για την πρόσβαση στο Twitter API (Application Programming Interface/Διασύνδεση Προγραμματισμού Εφαρμογών) ώστε να είναι εφικτή η απόκτηση των secrets και tokens keys (κωδικοί αυθεντικότητας) τα οποία κάνουν μοναδική τη πρόσβαση ανά χρήση, λειτουργούν δηλαδή ως «ταυτότητα» για την εφαρμογή.

Για την απόκτηση των προαναφερθέντων κωδικών έγινε σύνδεση στο Twitter Developer ώστε να δημιουργηθεί μια νέα εφαρμογή. Τα βήματα που ακολουθήθηκαν είναι τα εξής:

1. Στο Twitter Apps δημιούργησα μια νέα εφαρμογή πατώντας την αντίστοιχη επιλογή «Create App».
2. Στη συνέχεια ζητήθηκε να δοθούν ένα σύντομο όνομα ,μια περιγραφή για την εφαρμογή καθώς και μια διεύθυνση ιστότοπου.
3. Τέλος, πρέπει να γίνει συναίνεση με τους όρους που πρέπει να τηρούνται αμφότερα και να επιλεγεί το Create your Twitter Application.

## Εργαλείο Ελέγχου Πληροφοριών σε Μέσα Κοινωνικής Δικτύωσης (Social Media) με εφαρμογή σε ειδικές ομάδες

### Application Details

**Name \***

Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens. 32 characters max.

**Description \***

Your application description, which will be shown in user-facing authorization screens. Between 10 and 200 characters max.

**Website \***

Your application's publicly accessible home page, where users can go to download, make use of, or find out more information about your application. source attribution for tweets created by your application and will be shown in user-facing authorization screens. (If you don't have a URL yet, just put a placeholder here but remember to change it later.)

**Callback URLs**

Where should we return after successfully authenticating? OAuth 1.0a applications must explicitly specify their oauth\_callback URL(s) here, as well as request token step. To restrict your application from using callbacks, leave this field blank.

### Developer Agreement

Yes, I have read and agree to the [Twitter Developer Agreement](#).

**Εικόνα 4.1:** Create App

Έχοντας γίνει δεκτή η εγγραφή στο, δημιουργήθηκε το αναγνωριστικό πρόσβασης (secrets and tokens keys) από την αντίστοιχη καρτέλα Keys and Access Tokens.



[Details](#) [Settings](#) [Keys and Access Tokens](#) [Permissions](#)

### Application Settings

*Keep the "Consumer Secret" a secret. This key should never be human-readable in your application.*

Consumer Key (API Key)	<input type="text"/>
Consumer Secret (API Secret)	<input type="text"/>
Access Level	Read and write ( <a href="#">modify app permissions</a> )
Owner	<input type="text"/>
Owner ID	<input type="text"/>

### Your Access Token

*This access token can be used to make API requests on your own account's behalf. Do not share your access token secret with anyone.*

Access Token	<input type="text"/>
Access Token Secret	<input type="text"/>
Access Level	Read and write
Owner	<input type="text"/>
Owner ID	<input type="text"/>

**Εικόνα 4.2:**Application Settings

#### 4.2.1 Βιβλιοθήκες της Python που χρησιμοποιήθηκαν

Για τις ανάγκες ανάπτυξης της εφαρμογής χρησιμοποιήθηκαν οι ακόλουθες βιβλιοθήκες:

- Tweepy: βιβλιοθήκη για συλλογή δεδομένων από το twitter και πρόσβαση στο twitter Api.
- Textblob: είναι μια βιβλιοθήκη της Python για την επεξεργασία δεδομένων κειμένου. Παρέχει ένα απλό API για κοινές εργασίες επεξεργασίας φυσικής γλώσσας (NLP), ανάλυση συναισθημάτων κ.α
- Matplotlib.pyplot: βιβλιοθήκη για κατασκευή διαγραμμάτων.
- Datetime και time: βιβλιοθήκες για να ορίζεται η ημερομηνία και η ώρα σε σωστό format.
- Tkinter: βιβλιοθήκη γραφικών της Python.
- sqlite3: βιβλιοθήκη C, που παρέχει μια ελαφριά βάση δεδομένων, οποία επιτρέπει πρόσβαση στην βάση δεδομένων χρησιμοποιώντας μια μη τυπική παραλλαγή της γλώσσας ερωτήματος SQL.

#### 4.3 Πρόσβαση στα δεδομένα με REST APIs και Streaming APIs

Το Twitter παρέχει δύο τύπους API για να έχει κάποιος χρήστης πρόσβαση στα δεδομένα τους:

1. REST full APIs (Representational State Transfer): χρησιμοποιείται για τη παροχή δεδομένων σε προγενέστερο χρόνο από την εκτέλεση του ερωτήματος. Επίσης υπάρχει η δυνατότητα ανάκτησης δεδομένων από διάφορους χρήστες και η παρακολούθηση θεμάτων που είναι δημοφιλή εκείνη τη συγκεκριμένη περίοδο.
2. Streaming APIs: χρησιμοποιείται για live καταστάσεις tweets καθώς αποστέλλονται. Υπάρχουν τρία διαφορετικά είδη ροών. Η πρώτη είναι με δημόσια δεδομένα όπου μας δίνεται πρόσβαση σε όλα τα tweet που διέρχονται στο σύστημα. Μπορούμε να επιλέξουμε θέματα τα οποία πραγματεύεται η ροή. Το δεύτερο και το τρίτο είδος ροής αναφέρονται σε αναρτήσεις χρηστών. Κάθε λογαριασμός μπορεί να έχει μόνο μια ανοικτή σύνδεση με το API.

## ΚΕΦΑΛΑΙΟ 5

### ΔΗΜΙΟΥΡΓΙΑ ΕΡΓΑΛΕΙΟΥ ΣΥΛΛΟΓΗΣ TWEETS

#### 5.1 Περιγραφή εργαλείου συλλογής tweets

Δημιουργήθηκε με τη γλώσσα προγραμματισμού python ένα πρόγραμμα συλλογής δεδομένων από τον ιστότοπο του Twitter βάση μιας λέξης κλειδί που ορίζει ο χρήστης και ενός συγκεκριμένου αριθμού tweets που επιθυμεί να συλλέξει.

Αναλυτικότερα, η εφαρμογή στοχεύει στη συλλογή δεδομένων που αφορούν κάποια από τις ειδικές ομάδες όπως είναι η αστυνομία, ο στρατός ή η ομάδα των ναυτικών. Οι ομάδες αυτές τα τελευταία χρόνια έχουν αρχίσει και δημοσιεύουν στους ιστοτόπους των μέσων κοινωνικής πληροφορίας ως αναφορά τη δράση τους. Η ομάδα για την οποία θα γίνει συλλογή δεδομένων από το Twitter είναι αυτή των ναυτικών. Πιο συγκεκριμένα, θα γίνει εξόρυξη δεδομένων για μια ναυτιλιακή εταιρεία και στη συνέχεια ανάλυση συναισθημάτων των tweets που συλλέχθηκαν. Όπως σε όλες τις εταιρείες έτσι και σε μια ναυτιλιακή εταιρεία παίζει σημαντικό ρόλο να γνωρίζει την απήχηση που έχει στον κόσμο, καθώς και όσα δημοσιεύονται για αυτή αν είναι θετικά ή αρνητικά και σε τι ποσοστό

Σκοπός της εφαρμογής είναι να συλλέξει και να αποθηκεύσει σε μια βάση δεδομένων ένα εύρος tweets τα οποία αφορούν τη συγκεκριμένη ναυτιλιακή εταιρεία και στη συνέχεια να παρουσιάσει μέσω διαγράμματος για το αν τα δεδομένα που συλλέχθηκαν έχουν θετικό, αρνητικό ή ουδέτερο περιεχόμενο, καθώς επίσης και το ποσοστό τους.



Εργαλείο Ελέγχου Πληροφοριών σε Μέσα Κοινωνικής Δικτύωσης (Social Media) με εφαρμογή σε ειδικές ομάδες

Δημιουργούνται λίστες (για πολικότητα και αριθμό δεδομένων) ώστε να αποθηκεύσουμε τα δεδομένα μας και να είναι ποιο εύκολη η απεικόνιση του συναισθήματος στο διάγραμμα. Ορίζουμε μετρητές counter, για τα συναισθήματα αρνητικά, θετικά και ουδέτερα.

Δημιουργούμε, μια σύνδεση στη βάση δεδομένων και ένα πίνακα με την ονομασία tweets που έχει ως στήλες τα keyword, username, tweets, polarity.

```
root = Tk()

label1 = Label(root, text="Search")
E1 = Entry(root, bd =5)

label2 = Label(root, text="Sample Size")
E2 = Entry(root, bd =5)

def getE1():
    return E1.get()

def getE2():
    return E2.get()

def getData():
    getE1()
    keyword = getE1()

    getE2()
    numberOfTweets = getE2()
    numberOfTweets = int(numberOfTweets)

    #Where the tweets are stored to be plotted
    polarity_list = []
    numbers list = []
    number = 1

    poscounter=0
    neutcounter=0
    negcounter=0

    conn = sqlite3.connect('Twitter.db')
    c = conn.cursor()

    c.execute("CREATE TABLE IF NOT EXISTS tweets (keyword varchar,username varchar,tweets text, polarity int)")
    conn.commit()
```

**Εικόνα 5.3:** Δημιουργία GUI, σύνδεση στη βάση και δημιουργία functions

Μέσα στο βρόγχο for χρησιμοποιώντας τη μέθοδο REST API το πρόγραμμα βρίσκει και αποθηκεύει τα tweets που σχετίζονται με τη λέξη κλειδί και τα συναισθήματα του κάθε tweet. Η αποθήκευση γίνεται στην αντίστοιχη βάση που αναφέραμε παραπάνω.

```
for tweet in tweepy.Cursor(api.search, keyword, lang="en").items(numberOfTweets):
    try:
        analysis = TextBlob(tweet.text)
        analysis = analysis.sentiment
        polarity = analysis.polarity
        polarity_list.append(polarity)
        numbers_list.append(number)
        number = number + 1

        tweets=tweet.text.encode('utf-8')

        if polarity > 0:
            poscounter= poscounter+1
        elif polarity == 0:
            neutcounter=neutcounter+1
        else:
            negcounter=negcounter+1

        tweets=tweet.text.encode('utf-8')
        username = tweet.user.screen_name
        tweets=tweet.text.encode('utf-8')
        print(username,tweets,polarity)

    # print([(tweet.created_at, tweet.text.encode('utf-8')),polarity])

    c.execute("INSERT INTO tweets (keyword,username,tweets,polarity)VALUES(?,?,?,?)",
              (keyword,username,tweets,polarity))
    conn.commit()

except tweepy.TweepError as e:
    print(e.reason)

except StopIteration:
    break
```

**Εικόνα 5.4:** REST API, εύρεση tweet βάση keyword και υπολογισμός συναισθημάτων

Για την απεικόνιση των διαγραμμάτων βρίσκουμε το συνολικό συναίσθημα των tweets που συγκεντρώθηκαν και υπολογίζουμε το μέσο όρο σε όλα τα tweet που έχουν συλλεχθεί μέχρι εκείνη τη στιγμή.

Για την ορθή αποτύπωση των συναισθημάτων βρίσκουμε το ποσοστό των θετικών, αρνητικών και ουδέτερων tweets που έχουν δημοσιευθεί για την ναυτιλιακή.

```
pos_perc=format(100*(poscounter)/(numberOfTweets))
neg_perc=format(100*(negcounter)/(numberOfTweets))
neut_perc=format(100*(neutcounter)/(numberOfTweets))

print("Positive tweets percentage: {} %".format(100*(poscounter)/(numberOfTweets)))
print("Negative tweets percentage: {} %".format(100*(negcounter)/(numberOfTweets)))
print("Neutral tweets percentage: {} %".format(100*(neutcounter)/(numberOfTweets)))
```

**Εικόνα 5.5:** Εύρεση ποσοστού συναισθημάτων

Εργαλείο Ελέγχου Πληροφοριών σε Μέσα Κοινωνικής Δικτύωσης (Social Media) με εφαρμογή σε ειδικές ομάδες

Τέλος φτιάχνουμε τους άξονες για την απεικόνιση των συναισθημάτων κάθε tweet και το πλαίσιο με τις πληροφορίες για την ημερομηνία και ώρα άντλησης των δεδομένων και για το μέσο όρο όλων των δεδομένων που συλλέχθηκαν.

```
plt.figure(1)
axes = plt.gca()
axes.set_ylim([-1, 2])

plt.scatter(numbers_list, polarity_list)

averagePolarity = (sum(polarity_list))/len(polarity_list)
averagePolarity = "%.0f%%".format(averagePolarity * 100)
time = datetime.now().strftime("%A: %B:%d:%m: %m-%d-%y")

plt.text(0, 1.25, "Average Sentiment: " + str(averagePolarity) + "%\n" + time, fontsize=12, bbox = dict(facecolor='none', edgecolor='black', boxstyle='square, pad = 1'))

plt.title("Sentiment of " + keyword + " on Twitter")
plt.xlabel("Number of Tweets")
plt.ylabel("Sentiment")
```

Εικόνα 5.6: Σχεδίαση πίνακα απεικόνισης συναισθήματος tweet

Τέλος, δημιουργήθηκε μία πίτα που απεικονίζει το ποσοστό των συναισθημάτων βάση του δείγματος των tweets που συλλέγησαν.

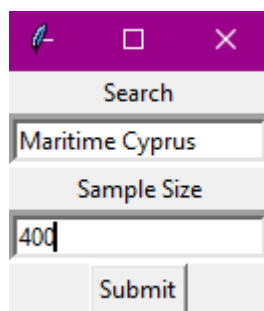
```
plt.figure(2)
values = [poscounter, negcounter, neutcounter]
labels = ['positive', 'negative', 'neutral']
plt.pie(values, labels=labels, autopct='%.1f%%')

plt.show()
```

Εικόνα 5.7: Σχεδίαση πίτας με συνολικά ποσοστά συναισθημάτων

## 5.2 Αποτελέσματα εφαρμογής

Όπως αναφέρθηκε και στην αρχή η εφαρμογή χρησιμοποιεί για keyword την ονομασία μιας ναυτιλιακής εταιρείας και ένα δείγμα συλλογής δεδομένων. Τρέχοντας την εφαρμογή εμφανίζεται το παράθυρο στο οποίο ο χρήστης καλείται να ορίσει τη λέξη κλειδί και τον αριθμό του δείγματος που επιθυμεί να συλλέξει.



Εικόνα 5.2.1: GUI

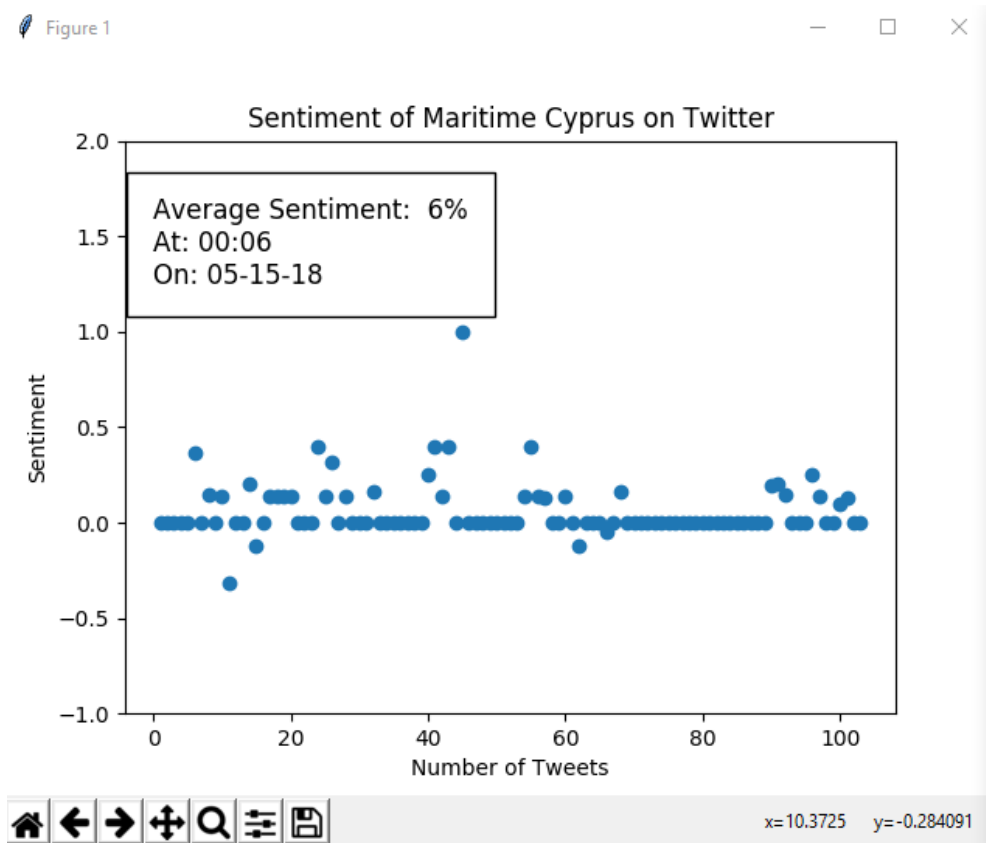
Πατώντας submit το πρόγραμμα βρίσκει και αποθηκεύει βάση της λέξης που όρισε ο χρήστης τα 400 tweets που έχουν δημοσιευτεί μέχρι σήμερα και αφορούν το keyword.

	keyword	ernan	tweets	polarity
	Filter		Filter	Filter
1	Maritime Cyprus	un...	Infographic - The Age of Explorati...	0
2	Maritime Cyprus	tra...	RT @ZSchneeweiss: Cyprus's loc...	0
3	Maritime Cyprus	nch...	RT @ZSchneeweiss: Cyprus's loc...	0
4	Maritime Cyprus	ma...	RT @ZSchneeweiss: Cyprus's loc...	0
5	Maritime Cyprus	ma...	RT @SeafarersWeek: #Cyprus's f...	0.16
6	Maritime Cyprus	ma...	International Chamber of Shipping...	0
7	Maritime Cyprus	ma...	Cruises via cargo ships !!!	0
8	Maritime Cyprus	ma...	Amazing new cruise ship https://t...	0.3681818181...
9	Maritime Cyprus	ma...	International Chamber of Shipping...	0
10	Maritime Cyprus	ma...	RT @cyprusmail: Cypriot sailor Ko...	0.15
11	Maritime Cyprus	ma...	RT @CruiseFever: World's newes...	0
12	Maritime Cyprus	ma...	RT @CruiseFever: Photos of new ...	0.1363636363...
13	Maritime Cyprus	ma...	RT @swansea_dst: A small remin...	-0.3125
14	Maritime Cyprus	ma...	Infographic – The Age of Explorati...	0
15	Maritime Cyprus	ma...	RT @americanson57: One would ...	0.2

**Εικόνα 5.2.2:** Απεικόνιση δεδομένων (δείγμα) όπως αποθηκεύτηκαν στη βάση

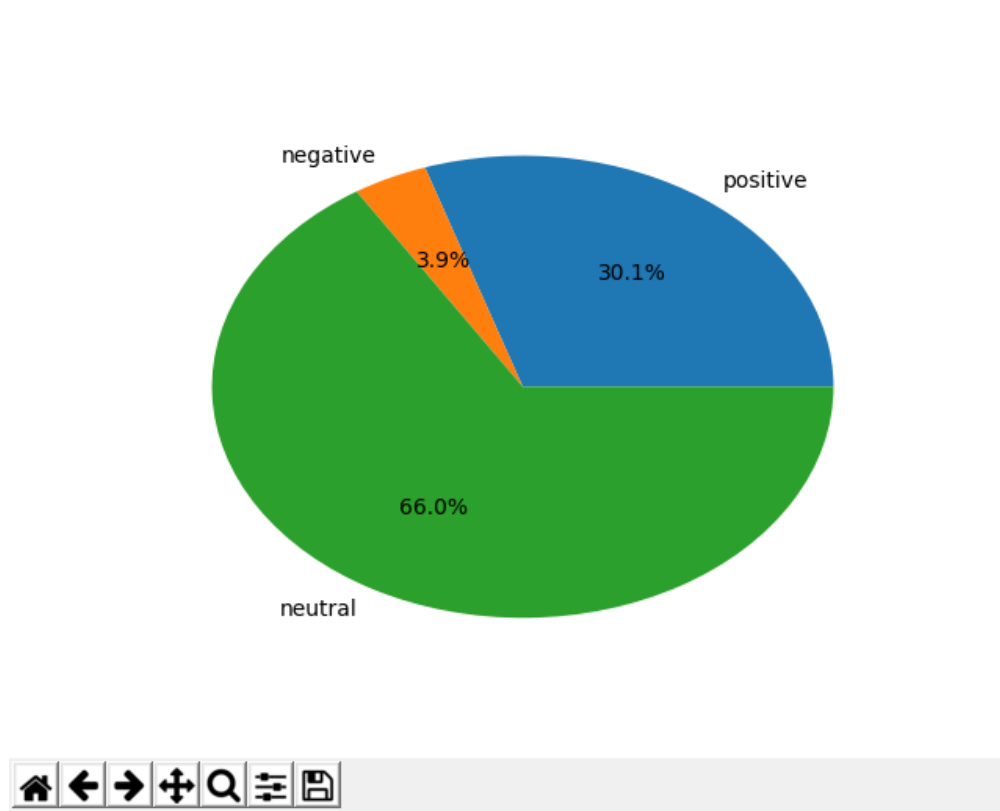


Γράφημα με τα συναισθήματα κάθε tweet που έχει καταγραφεί.



Εικόνα 5.2.3: Σχεδίαση συναισθημάτων

Σχεδίαση πίτας με το συνολικό ποσοστό των συναισθημάτων.



Εικόνα 5.2.3: Απεικόνιση συνολικού ποσοστού συναισθημάτων

## ΚΕΦΑΛΑΙΟ 7

### ΣΥΜΠΕΡΑΣΜΑΤΑ

#### 7.1 Ανασκόπηση

Η παρούσα πτυχιακή εργασία αποτέλεσε μια προσπάθεια κατανόησης της λειτουργίας των μέσων κοινωνικής δικτύωσης καθώς και της επεξεργασίας των δημοσιεύσεων των χρηστών σε αυτά. Για τη θεμελίωση της προσπάθειας κατανόησης της επεξεργασίας κειμένων στα social media έγινε αναφορά στην επεξεργασία φυσικής γλώσσας και ειδικότερα στο κομμάτι της ανάλυσης συναισθημάτων.

Στη συνέχεια παρουσιάστηκαν οι γλώσσα προγραμματισμού python, οι μέθοδοι που χρησιμοποιεί για την συλλογή πληροφοριών από το twitter καθώς και ο τρόπος με τον οποίο γίνεται σύνδεση μια εφαρμογής που δημιουργείται με τη συγκεκριμένη γλώσσα και του twitter

Βασικό μέλος είναι η δημιουργία εφαρμογής με τη συγκεκριμένη γλώσσα και με την χρήση των κατάλληλων βιβλιοθηκών της για την εξαγωγή και επεξεργασία δεδομένων από το twitter.

Τέλος, παρουσιάζεται η λειτουργία της εφαρμογής, η αναζήτηση ενός συγκεκριμένου δείγματος δεδομένων βάση μιας λέξης κλειδί για μια συγκεκριμένη ομάδα και η απεικόνιση διαγραμμάτων σχετικά με το αν τα tweets που αντλήθηκαν είχαν θετικό, αρνητικό ή ουδέτερο περιεχόμενο. Καθώς και το συνολικό ποσοστό τους

#### 7.2 Επίτευξη στόχων

Οι στόχοι της εργασία εκτιμάται ότι πραγματοποιήθηκαν σε ικανοποιητικό βαθμό. Η έρευνα πραγματοποιήθηκε μέσω του διαδικτύου, των εκπαιδευτικών συγγραμμάτων του προγράμματος σπουδών του τμήματος Μηχανικών Ηλεκτρονικών Υπολογιστικών Συστημάτων, αλλά και σχετικών βιβλίων από τη Βιβλιοθήκη του Ανώτατου Τεχνολογικού Εκπαιδευτικού Ιδρύματος Πειραιά Τεχνολογικού Τομέα.

### **7.3 Πιθανές επεκτάσεις**

Μια επέκταση της εργασίας θα μπορούσε να αποτελέσει η εξόρυξη δεδομένων σε πραγματικό χρόνο (real-time), και η γραφική απεικόνιση τους. Το οποίο θα μπορέσει να υλοποιηθεί, μέσα από την εμβάθυνση στο γνωστικό πεδίο της μηχανικής μάθησης και της ανάλυσης των αλγορίθμων που χρησιμοποιούνται για την εξόρυξη δεδομένων.



## Εργαλείο Ελέγχου Πληροφοριών σε Μέσα Κοινωνικής Δικτύωσης (Social Media) με εφαρμογή σε ειδικές ομάδες

```
for tweet in tweepy.Cursor(api.search, keyword, lang="en").items(numberOfTweets):
    try:
        analysis = TextBlob(tweet.text)
        analysis = analysis.sentiment
        polarity = analysis.polarity
        polarity_list.append(polarity)
        numbers_list.append(number)
        number = number + 1

        tweets=tweet.text.encode('utf-8')

        if polarity > 0:
            poscounter=poscounter+1
        elif polarity == 0:
            neutcounter=neutcounter+1
        else:
            negcounter=negcounter+1

        tweets=tweet.text.encode('utf-8')
        username = tweet.user.screen_name
        tweets=tweet.text.encode('utf-8')

        print(username,tweets,polarity)

        c.execute("INSERT INTO tweets (keyword,username,tweets,polarity)VALUES(?,?,?,?)",
            (keyword,username,tweets,polarity))
        conn.commit()

    except tweepy.TweepError as e:
        print(e.reason)

    except StopIteration:
        break

pos_perc=format(100*(poscounter)/(numberOfTweets))
neg_perc=format(100*(negcounter)/(numberOfTweets))
neut_perc=format(100*(neutcounter)/(numberOfTweets))
```

```
#Plotting
plt.figure(1)
axes = plt.gca()
axes.set_ylim([-1, 2])

plt.scatter(numbers_list, polarity_list)

averagePolarity = (sum(polarity_list))/(len(polarity_list))
averagePolarity = "{0:.0f}%".format(averagePolarity * 100)
time = datetime.now().strftime("%A: %H:%M:%S: %m-%d-%y")

plt.text(0, 1.25, "Average Sentiment: " + str(averagePolarity) + "%\n" + time, fontsize=12, bbox = dict(facecolor='none', edgecolor='black', boxstyle='square, pad = 1'))

plt.title("Sentiment of " + keyword + " on Twitter")
plt.xlabel("Number of Tweets")
plt.ylabel("Sentiment")
```

```
plt.figure(2)
values = [pos_perc, neg_perc, neut_perc]
labels = ['positive', 'negative', 'neutral']
plt.pie(values, labels=labels, autopct='%1.2f')

plt.show()

submit = Button(root, text = "Submit", command = getData)

labell.pack()
E1.pack()
label2.pack()
E2.pack()
submit.pack(side =BOTTOM)

root.mainloop()
```

## ΒΙΒΛΙΟΓΡΑΦΙΑ

Βλαχαβάς Ι,Κεφαλάς Π.Βασιλειάδης Ν.,Κοκκορας Φ.,Σακελλαρίου Η, (2011).

*Τεχνητη Νοημοσυνη Γ έκδοση*. Εκδόσεις Πανεπιστημίου Μακεδονίας.

Ricardo Baeza-Yates,Berthier Riberio-Neto. (2015). *Ανάκτηση πληροφορίας οι*

*έννοιες και οι τεχνολογίες πίσω από την αναζήτηση*. Θεσσαλονίκη: Τζιόλας.

Παναγιώτης Ηρ.Γιαννακόπουλος (2009). Πληροφορική & Κοινωνία.

Mathew A.Russell (2013). *Mining the Social Web*.

T. Mitchell (1997). *Machine Learning*. McGraw Hill

<http://www.itbiz.gr/κοινωνικα-δικτυα-social-networks-σε-μια-επιχειρηση/>.

<https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>

<http://www.tweepy.org/>

[https://matplotlib.org/users/pyplot\\_tutorial.html](https://matplotlib.org/users/pyplot_tutorial.html)

<http://textblob.readthedocs.io/en/dev/>

<https://www.analyticsvidhya.com/blog/2018/02/natural-language-processing-for-beginners-using-textblob/>

<https://docs.python.org/3/library/sqlite3.html>

<https://docs.python.org/2/library/tkinter.html>

<https://developer.twitter.com/en/docs>

[https://en.wikipedia.org/wiki/Python\\_\(programming\\_language\)](https://en.wikipedia.org/wiki/Python_(programming_language))