



Πανεπιστήμιο Πειραιώς – Τμήμα Πληροφορικής

Πρόγραμμα Μεταπτυχιακών Σπουδών

«Πληροφορική»

Μεταπτυχιακή Διατριβή

Τίτλος Εργασίας	Συλλογή Δεδομένων από WhatsApp και Telegram Data crawling on WhatsApp and Telegram
Όνοματεπώνυμο Φοιτητή	Λήτος Γιώργης
Πατρώνυμο	Θωμάς
Αριθμός Μητρώου	ΜΠΠΛ/ 16028
Επιβλέπων Καθηγητής	Πατσάκης Κωνσταντίνος, Επίκουρος Καθηγητής

Ημερομηνία Παράδοσης **Ιούλιος 2020**

Τριμελής Εξεταστική Επιτροπή

(υπογραφή)

Κωνσταντίνος Πατσάκης
Επίκουρος Καθηγητής

(υπογραφή)

Ευθύμιος Αλέπης
Αναπληρωτής Καθηγητής

(υπογραφή)

Γεώργιος Τσιχριντζής
Καθηγητής

Περίληψη

Οι εφαρμογές αποστολής μηνυμάτων (Text Message Apps) έχουν αλλάξει δραστικά τον τρόπο επικοινωνίας και αλληλεπίδρασης των ανθρώπων. Η σωρεία της πληροφορίας και η ευκολία διακίνησης της ωστόσο προκαλούν σοβαρά ζητήματα σε διάφορους τομείς της καθημερινότητας, καθώς μπορεί το περιεχόμενο των μηνυμάτων να αναφέρεται σε ψευδείς ειδήσεις, παράνομο υλικό κι αλλού είδος παράνομες ή ανήθικες δραστηριότητες. Έτσι, δημιουργείται η ανάγκη έγκαιρης και ακριβής εύρεσης μηνυμάτων, κυρίως σε μαζικά κανάλια επικοινωνίας, που ενδεχομένως να περιέχουν αθέμιτο υλικό. Ωστόσο, η ανάλυση των δεδομένων που διακινούνται μέσω των υπηρεσιών αυτών με σκοπό την αναγνώριση του θέματος συζήτησης των ομάδων χρηστών (groups, channels) είναι μια πρόκληση, καθώς τα μηνύματα αυτά συνήθως έχουν μικρό μέγεθος ή περιλαμβάνουν πολυμέσα. Μια εξίσου σημαντική πρόκληση είναι η πρόληψη και η αποτροπή παραβίασης των όρων χρήσης των υπηρεσιών αυτών από κακόβουλους χρήστες, όπου επίσης απαιτεί την ανάλυση περιεχομένου έναντι των όρων.

Σκοπός της παρούσας εργασίας είναι να βοηθήσει, μέσω μιας αρχικής μαζικής ανάλυσης των μηνυμάτων, προς την κατεύθυνση της έγκαιρης εύρεσης μηνυμάτων σε κανάλια επικοινωνίας κοινωνικών δικτύων που ενδεχομένως να αποτελούν ενδείξεις αθέμιτων ενεργειών ή παραβίασης όρων. Συγκεκριμένα, εστιάσαμε στην συλλογή και ανάλυση δεδομένων σε δύο από τις πιο δημοφιλείς υπηρεσίες στον χώρο των Text Message Apps, το WhatsApp και το Telegram. Για να προσεγγίσουμε το πρόβλημα, αρχικά δημιουργήσαμε δύο σύνολα δεδομένων από συνομιλίες μελών διαφόρων ομάδων για την κάθε εφαρμογή ξεχωριστά και στη συνέχεια αναλύσαμε τα δεδομένα αυτά με την χρήση του αλγορίθμου K-means καθώς και με την πιθανοτική κατανομή LDA για να προσδιορίσουμε λέξεις κλειδιά που θα μας βοηθήσουν να εντοπίσουμε τα θέματα συζήτησης κάθε ομάδας. Στη συνέχεια, παρουσιάζουμε για κάθε cluster ή topic αντίστοιχα, τους όρους που φαίνεται να είναι πιο δημοφιλείς, με την βοήθεια των Word clouds. Τέλος, για κάθε μήνυμα στο κανάλι, υπολογίζουμε την πιθανότητα αυτό να ανήκει σε ένα από τα cluster/topic, και για κάθε κανάλι επιστρέφουμε την κατανομή των μηνυμάτων στα αντίστοιχα clusters ή topics.

Abstract

Text Message Apps have dramatically changed the way people communicate and interact. However, the plethora of information and the ease of sharing it hinder serious problems in various areas of our daily life, given that the content of the messages may refer to fake news, illegal material and other types of illegal or immoral activities. This generates the need to timely and accurately identify messages that may contain illegal material, especially in mass media communication channels. Yet, analysis of the data communicated through these services with the objective to identify the topic of discussion in the different groups or channels is challenging, as messages are typically small in size and may contain multimedia. An equally important challenge is to deter and prevent the breach of the terms of use of these services by malicious users, which also requires content analysis against the terms.

The goal of this thesis is to assist in the direction of timely identifying messages on social media channels that may be indicative of unfair actions or breach of terms, through a primary collective analysis of messages. Specifically, we focus on data collection and analysis in two of the most popular services in the field of Text Message Apps, namely WhatsApp and Telegram. We approach the problem in distinct phases. First, we create two unique datasets for each application, derived from conversations among members of different groups within each application. Second, we analyse this data using the K-means algorithm as well as the LDA probability distribution, aiming to identify keywords that will help us decipher key discussion topics within the groups. Next, for each cluster or topic respectively we depict the terms that seem to be most popular using Word Clouds. Finally, for each message on the channel, we calculate the probability that this message

belongs to one of a specific cluster / topic, and for each channel we return the distribution of messages to the respective clusters or topics.

Λέξεις κλειδιά:

WhatsApp, Telegram, data crawling, WhatsApp groups, Telegram groups, Telegram channels, messaging, k-means algorithm, python, clustering, TF-IDF, LDA Topic Modeling.

Πίνακας Περιεχομένων

Κατάλογος Εικόνων και Διαγραμμάτων.....	6
Εισαγωγή.....	8
Κεφάλαιο 1. Messaging.....	9
Κεφάλαιο 2. Συλλογή δεδομένων.....	11
2.1 Μεθοδολογία συλλογής δεδομένων.....	11
2.1.1 Μεθοδολογία συλλογής δεδομένων για την εφαρμογή WhatsApp.....	11
2.1.2 Μεθοδολογία συλλογής δεδομένων για την εφαρμογή Telegram.....	11
2.2 Ηθικά ζητήματα.....	12
Κεφάλαιο 3. Telegram.....	13
3.1 Στατιστικά χρηστών και χρήσης του Telegram.....	14
3.2 Ρυθμός αύξησης χρηστών του Telegram.....	14
3.3 Σε ποιες χώρες είναι πιο δημοφιλές το Telegram;.....	16
3.4 Στατιστικά στοιχεία για τα κανάλια (channel) του Telegram.....	18
Κεφάλαιο 4. WhatsApp.....	20
4.1 Στατιστικά στοιχεία χρηστών του WhatsApp.....	21
4.2 Χρήστες του WhatsApp ανά χώρα.....	23
4.3 WhatsApp Groups.....	24
Κεφάλαιο 5. Συλλογή και ανάλυση δεδομένων Telegram.....	25
5.1 Ανάλυση δεδομένων με τον αλγόριθμο K-means (clustering).....	39
5.2 Ανάλυση δεδομένων με LDA (Topic Extraction).....	47
Κεφάλαιο 6. Συλλογή και ανάλυση δεδομένων από το WhatsApp.....	59
6.1 Εγκατάσταση απαιτούμενων εργαλείων.....	59
6.2 Συλλογή δεδομένων.....	66
6.3 Ανάλυση δεδομένων.....	69
6.3.1 Ανάλυση δεδομένων με τον αλγόριθμο K-means (clustering).....	79
6.3.2 Ανάλυση δεδομένων με LDA (Topic Extraction).....	99
Κεφάλαιο 7. K-means clustering (ομαδοποίηση).....	144
Κεφάλαιο 8. TF-IDF.....	146
Κεφάλαιο 9. Topic Modeling με Latent Dirichlet Allocation (LDA).....	149
ΣΥΜΠΕΡΑΣΜΑΤΑ.....	151
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	153

Κατάλογος Εικόνων και Διαγραμμάτων

Κατάλογος Διαγραμμάτων

- Διάγραμμα 1. Telegram MAU έναντι άλλων εφαρμογών μηνυμάτων
- Διάγραμμα 2. Αύξηση μηνιαίως ενεργών χρηστών του Telegram 2014 – 2018
- Διάγραμμα 3. Εγκατάσταση για πρώτη φορά της εφαρμογής Telegram στο Χονγκ Κονγκ
- Διάγραμμα 4. Κορυφαία κανάλια του Telegram το 2019 (σε εκατομμύρια χρήστες)
- Διάγραμμα 5. Πόσα άτομα χρησιμοποιούν το WhatsApp; Απρίλιος 2013-Δεκέμβριος 2017
- Διάγραμμα 6. Πόσα άτομα χρησιμοποιούν το WhatsApp έναντι άλλων εφαρμογών μηνυμάτων
- Διάγραμμα 7. Εφαρμογές του Facebook βάση αριθμού χρηστών
- Διάγραμμα 8. Ποσοστό διείσδυσης του WhatsApp ανά χώρα

Κατάλογος Εικόνων

- Εικόνα 1. Δημοφιλία του Telegram ανά τον κόσμο.
- Εικόνα 2. Δημοφιλία του WhatsApp ανά τον κόσμο.
- Εικόνα 3. - Εικόνα 4. Screenshots από τα βήματα που ακολουθήθηκαν για την απόκτηση `api_id` και `api_hash` για το Telegram.
- Εικόνα 5. Screenshot από την επικοινωνία με τον BotFather για την απόκτηση token για το Telegram.
- Εικόνα 6. - Εικόνα 9. Screenshots από την εκτέλεση του προγράμματος για το Telegram (με K-means).
- Εικόνα 10. – Εικόνα 21. Wordclouds για το κάθε cluster που δημιουργήθηκε για το Telegram (με K-means).
- Εικόνα 22. – Εικόνα 24. Screenshots από την εκτέλεση του προγράμματος για το Telegram (με LDA).
- Εικόνα 25. – Εικόνα 37. Wordclouds για το κάθε topic που δημιουργήθηκε για το Telegram (με LDA)
- Εικόνα 38. – Εικόνα 53. Screenshots από την εγκατάσταση των που απαιτούμενων εργαλείων για το WhatsApp.
- Εικόνα 54. – Εικόνα 57. Screenshots από την συλλογή δεδομένων για το WhatsApp.
- Εικόνα 58. – Εικόνα 60. Screenshots από την εκτέλεση του προγράμματος για το WhatsApp (με K-means).
- Εικόνα 61. – Εικόνα 75. Wordclouds για το κάθε cluster που δημιουργήθηκε για το WhatsApp (με K-means).
- Εικόνα 76. - Εικόνα 78. Screenshots από την εκτέλεση του προγράμματος για το WhatsApp (με LDA)
- Εικόνα 79. – Εικόνα 91. Wordclouds για το κάθε topic που δημιουργήθηκε για το WhatsApp (με LDA).
- Εικόνα 92. Clustering για $K=3$.
- Εικόνα 93. Τύπος υπολογισμού Συχνότητας Όρου (TF).
- Εικόνα 94. Τύπος υπολογισμού Αντίστροφης Συχνότητας Εγγράφων (IDF).
- Εικόνα 95. Απεικόνιση παραδείγματος για τον υπολογισμό του TF – IDF.
- Εικόνα 96. Το διάγραμμα του μοντέλου LDA.

Κατάλογος Πινάκων

- Πίνακας 1. Τελικά αποτελέσματα Telegram με K-means.
- Πίνακας 2. Τελικά αποτελέσματα Telegram με LDA.
- Πίνακας 3. Τελικά αποτελέσματα WhatsApp με K-means.

Πίνακας 4. Τελικά αποτελέσματα WhatsApp με LDA.

Κατάλογος Script

- Script 1. settings.ini (Telegram)
- Script 2. collector.py (Telegram)
- Script 3. document.py (Telegram)
- Script 4. word_cloud_helper.py (Telegram)
- Script 5. text_proccesing.py (Telegram)
- Script 6. files_management.py (Telegram)
- Script 7. corpusCreator.py (Telegram)
- Script 8. Kmeans.py (Telegram)
- Script 9. LDA.py (Telegram)
- Script 10. tests.py (Telegram)
- Script 11. test_lda.py (Telegram)
- Script 12. document.py (WhatsApp)
- Script 13. word_cloud_helper.py (WhatsApp)
- Script 14. text_proccesing.py (WhatsApp)
- Script 15. files_management.py (WhatsApp)
- Script 16. corpusCreator.py (WhatsApp)
- Script 17. Kmeans.py (WhatsApp)
- Script 18. LDA.py (WhatsApp)
- Script 19. tests.py (WhatsApp)
- Script 20. test_lda.py (WhatsApp)

Εισαγωγή

Τα τελευταία χρόνια παρατηρούμε μια συνεχή αύξηση της χρήσης διαφόρων εφαρμογών αποστολής και λήψης μηνυμάτων από εκατομμύρια χρήστες παγκοσμίως. Οι εφαρμογές αυτές προσφέρουν τις υπηρεσίες τους δωρεάν στο ευρύ κοινό και επιπλέον οι χρήστες τους έχουν την δυνατότητα να έχουν εύκολη πρόσβαση σε αυτές είτε από κάποιον υπολογιστή όταν βρίσκονται στο σπίτι ή το γραφείο τους, είτε από κάποια συσκευή κινητής τηλεφωνίας όταν βρίσκονται καθόδων. Εκτός της απλής αποστολής και λήψης μηνυμάτων προσφέρονται στον χρήστη πλείστες άλλες δυνατότητες με αποτέλεσμα η χρήση των συγκεκριμένων εφαρμογών να ξεπερνά τα όρια της απλής επικοινωνίας. Η παραδοσιακή αποστολή και λήψη μηνυμάτων από χρήστη σε χρήστη περνά πλέον στο παρελθόν και μέσω των εφαρμογών αυτών ο χρήστης πλέον δύναται να στέλνει και να λαμβάνει μηνύματα από πολλούς χρήστες ταυτόχρονα μεσώ των ομάδων χρηστών που δημιουργούνται. Στις ομάδες αυτές οι χρήστες που πολλές φορές είναι και άγνωστοι μεταξύ τους συζητάνε για διάφορα θέματα του κοινού τους ενδιαφέροντος. Για παράδειγμα υπάρχουν ομάδες που έχουν ως θέμα συζήτησης τους τα διαδικτυακά παιχνίδια, ομάδες που έχουν ως θέμα συζήτησης το εμπόριο, την διαφήμιση, το ερωτικό περιεχόμενο, τις γνωριμίες κ.α. Συνήθως οι διαχειριστές των ομάδων αυτών δίνουν στις ομάδες τους κάποιο όνομα το οποίο σχετίζεται με το θέμα συζήτησης των χρηστών τους, με αποτέλεσμα μέσω του ονόματος μιας ομάδας να μπορούμε να βγάλουμε ένα συμπέρασμα για το τι περίπου συζητάνε οι χρήστες - μέλη της. Υπάρχουν όμως και πολλές άλλες ομάδες που δύσκολα μπορούμε να αντιληφθούμε ποιο είναι το θέμα για το οποίο συζητάνε οι χρήστες τους.

Και με αυτό ακριβώς θα ασχοληθούμε στην παρούσα εργασία. Θα συλλέξουμε και θα αναλύσουμε δεδομένα από ομάδες (groups) που χρησιμοποιούν ως μέσο για την αποστολή και λήψη μηνυμάτων, τις γνωστές εφαρμογές WhatsApp και Telegram. Κύριος στόχος μας είναι να επεξεργαστούμε τα δεδομένα αυτά έτσι ώστε να μπορέσουμε να καταλάβουμε για ποια θέματα πιθανών να συζητούν τα μέλη σε κάθε ομάδα του WhatsApp και του Telegram.

Δεδομένου του ότι τα τελευταία χρόνια οι δυο συγκεκριμένες εφαρμογές και γενικότερα οι εφαρμογές αυτού του είδους γίνονται όλο και πιο δημοφιλή, αντικαθιστώντας παλαιότερους μεθόδους αποστολής και λήψης μηνυμάτων όπως είναι τα SMS, εύκολα αντιλαμβανόμαστε την σπουδαιότητα εξαγωγής αποτελεσμάτων μέσω των δεδομένων που διακινούνται στα μέσα αυτά. Η εργασία περιλαμβάνει μια μικρή ιστορική αναδρομή γύρω από την αποστολή και λήψη μηνυμάτων, λίγα λόγια καθώς και δημογραφικά στοιχεία για την χρήση των δύο μέσων (WhatsApp και Telegram) και τα βήματα που έγιναν για την συλλογή και την ανάλυση των δεδομένων.

Λέξεις κλειδιά:

WhatsApp, Telegram, data crawling, WhatsApp groups, Telegram groups, Telegram channels, messaging, k-means algorithm, python, clustering, TF-IDF, LDA Topic Modeling.

Κεφάλαιο 1. Messaging

Το 1984 ο Γερμανός μηχανικός Friedhelm Hillebrand και ο γάλλος συνάδελφός του Bernard Ghillebaert εργάστηκαν σκληρά αναπτύσσοντας ένα σύστημα το οποίο θα επέτρεπε την ανταλλαγή γραπτών μηνυμάτων μέσω του υφιστάμενου τηλεφωνικού δικτύου κάνοντας χρήση της χωρητικότητας δικτύου που βρίσκονταν εκτός χρήσης. Το σύστημα βασίστηκε στα πρότυπα GSM (Global System for Mobile Communications).

Στα αρχικά βήματα της έρευνάς τους, οι Hillebrand και Ghillebaert διαπίστωσαν ότι ήταν απαραίτητο να περιοριστεί το μήκος των μηνυμάτων σε 128 byte, προκειμένου να χωρέσει στις υπάρχουσες μορφές σήματος. Αυτό βελτιώθηκε αργότερα σε 160 χαρακτήρες. Ο Hillebrand έφτασε στο συμπέρασμα αυτό μετά από την ανάλυση μηνυμάτων που αποστέλλονταν μέσω κάρτ ποστάλ και του Telex και έδειχνε ότι οι 160 χαρακτήρες ήταν αρκετοί για να περιγράψουν συνοπτικά τα περισσότερα μηνύματα. Ήταν η ίδια αρχή, βάση της οποίας πιο πρόσφατα στις μέρες μας οι 140 χαρακτήρες θεσπίστηκαν ως το όριο των χαρακτήρων στο Twitter.

Παρόλο που οι Hillebrand και Ghillebaert ήταν αυτοί που έθεσαν τα θεμέλια για τα SMS μηνύματα, η εφευρέση του SMS όπως το γνωρίζουμε σήμερα πιστώνεται στον φινλανδό μηχανικό κινητής τηλεφωνίας Matti Makkonen. Ο Makkonen εργαζόταν τότε για τη Nokia και βοήθησε την εταιρεία να αναπτύξει τα πρώτα «SMS friendly» τηλέφωνα.

Το πρώτο μήνυμα SMS στάλθηκε στις 3 Δεκεμβρίου 1992 μέσω του δικτύου GSM της Vodafone στο Ηνωμένο Βασίλειο. Το μήνυμα εστάλη από τον Neil Papworth της Sema Group (τώρα Mavenir Systems) με την χρήση ενός ηλεκτρονικού υπολογιστή στον Richard Jarvis της Vodafone σε μια συσκευή Orbitel 901 Handset. Το κείμενο του μηνύματος ήταν "Merry Christmas".

Τα πρώτα χρόνια η υπηρεσία αποστολής μηνυμάτων SMS ξεκίνησε με πολύ μικρά βήματα. Το 1995, ο μέσος αριθμός μηνυμάτων που στάλθηκαν ήταν μόλις 0,4 μηνύματα ανά πελάτη το μήνα. Ωστόσο, αυτό οφειλόταν σε μεγάλο βαθμό στους παρόχους της υπηρεσίας και όχι στους πελάτες. Για παράδειγμα η συμφωνία που θα επέτρεπε στους χρήστες να στέλνουν μηνύματα SMS σε χρήστες άλλων δικτύων, επτεύχθη το 1999!

Μόλις ρυθμίστηκαν τα αρχικά ζητήματα χρέωσης και μειώθηκαν οι περιορισμοί, ξεκλειδώθηκε η πλήρης χρησιμότητα της υπηρεσίας των μηνυμάτων SMS και οι αριθμοί αποστολής μηνυμάτων αυξήθηκαν με γοργούς ρυθμούς. Έως το τέλος του 2000, οι χρήστες του Ηνωμένου Βασιλείου έστελναν κατά μέσο όρο 35 SMS ανά μήνα. Το 2010, αναφέρθηκε ότι είχαν σταλεί συνολικά 6.1 τρισεκατομμύρια μηνύματα το οποίο ισοδυναμεί με 193.000 SMS ανά δευτερόλεπτο.

Στις 24 Φεβρουαρίου 2009 στην Καλιφόρνια, ο Jan Koum δημιούργησε την εταιρεία WhatsApp Inc. και άρχισε να αναπτύσσει αυτό που θα έμελλε να γίνει η μεγαλύτερη εφαρμογή ανταλλαγής μηνυμάτων στον κόσμο μέσα σε λίγα μόλις χρόνια. Το WhatsApp ξεκίνησε ως ο βασικός τρόπος αποστολής ενημερώσεων κατάστασης για τις επαφές που βρίσκονται αποθηκευμένες στα κινητά τηλέφωνα, αλλά καθώς ο κόσμος άρχισε να το χρησιμοποιεί ως τρόπο αποστολής μηνυμάτων στους φίλους τους, ο Koum συνειδητοποίησε ότι δημιούργησε ακούσια μια υπηρεσία ανταλλαγής μηνυμάτων. Όπως είπε ο Koum, "η δυνατότητα να επικοινωνήσετε με κάποιον που βρίσκεται στην άλλη άκρη του κόσμου άμεσα, με μια συσκευή που είναι πάντα μαζί σας, είναι κάτι πολύ ισχυρό".

Την εποχή εκείνη, το κόστος αποστολής μηνυμάτων SMS μέσω του κινητού τηλεφώνου ήταν πολύ μεγάλο, και ειδικά η αποστολή SMS σε άλλες χώρες ήταν πανάκριβη. Μια εναλλακτική ήταν το BBM, αλλά η υπηρεσία αυτή ήταν συμβατή μόνο για τους κατόχους συσκευών BlackBerrie.

Μια άλλη εναλλακτική του οποίου η χρήση ήταν ελεύθερη ήταν το Skype, άλλα η χρήση του απαιτούσε τη δημιουργία ενός εντελώς νέου κοινωνικού γραφήματος. Το WhatsApp από την άλλη ήταν κάτι απίστευτα απλό και μοναδικό. Η μόνη απαίτηση για να συνδεθεί κάποιος στην υπηρεσία και να συνδεθεί άμεσα με όλους τους χρήστες τον οποίων τα τηλέφωνα ήταν αποθηκευμένα στις επαφές του τηλεφώνου του, ήταν η εγγραφή στην υπηρεσία με τον τηλεφωνικό του αριθμό. Μετά

την κυκλοφορία της έκδοσης 2.0 του WhatsApp στο iPhone με ενσωματωμένη την υπηρεσία αποστολής μηνυμάτων, ο αριθμός των χρηστών του WhatsApp αυξήθηκε στους 250.000 περίπου. Αμέσως μετά άρχισαν να πληθύνουν οι αιτήσεις από τους χρήστες της υπηρεσίας για να κυκλοφορήσει η υπηρεσία και σε άλλες πλατφόρμες, έτσι ώστε οι χρήστες iPhone να μπορούν να μιλούν με τους φίλους που έχουν συσκευές Nokia και Blackberrie σε όλο τον κόσμο δωρεάν. Άμεσα καταβλήθηκαν προσπάθειες για την επέκταση του WhatsApp και σε άλλες πλατφορμες και έως το 2011 η εφαρμογή WhatsApp ήταν σταθερά στις 20 κορυφαίες εφαρμογές του αμερικανικού App Store. Το 2013, ο αριθμός των χρηστών της εφαρμογής WhatsApp είχε αυξηθεί στους 200 εκατομμύρια χρήστες σε όλο τον κόσμο. Το 2014, ο αριθμός των χρηστών αυξήθηκε στους 450 εκατομμύρια χρήστες, ενώ καθημερινά ένα εκατομμύριο χρήστες προστίθενται στον αριθμό αυτό.

Σήμερα, οι εφαρμογές μηνυμάτων (συμπεριλαμβανομένων των Facebook Messenger, Telegram, WhatsApp, WeChat στην Ασία, Snapchat κ.α.) δίνουν την δυνατότητα στους χρήστες τους εκτός των απλών μηνυμάτων κειμένου να μπορούν να στείλουν επίσης εικονίδια, gifs, βίντεο κ.α. και όλα αυτά χωρίς κάποια χρέωση, πράγμα το οποίο έχει περιορίσει πλέον κατά πολύ τα μηνύματα SMS. Στις περισσότερες χώρες τα μηνύματα SMS έχουν σχεδόν εκλείψει και εκτός αυτού παρατηρείται επίσης και μια μείωση στην χρήση των κοινωνικών δίκτυων, τόσο στον συνολικό αριθμό των χρηστών τους όσο και στον ρυθμό ανάπτυξης τους. Το Facebook Messenger χρειάστηκε περίπου 40 μήνες για να φτάσει τα 500 εκατομμύρια χρήστες, και μόλις 20 μήνες αργότερα ο αριθμός αυτός διπλασιάστηκε και έφτασε τους 1 δισεκατομμύριο χρήστες, ένα επίτευγμα που θεωρείται απίστευτο. Η εφαρμογή WeChat αύξησε τους χρήστες της από 100 εκατομμύρια μηνιαία ενεργούς χρήστες σε 500 εκατομμύρια σε μόλις ενάμιση χρόνο. Σήμερα, έξι από τις δέκα πιο δημοφιλείς εφαρμογές στο iTunes και το Google Play είναι εφαρμογές ανταλλαγής μηνυμάτων.

Λαμβάνοντας υπόψιν όλα τα παραπάνω καταλαβαίνουμε την σπουδαιότητα του να συλλέξουμε και να αναλύσουμε τα δεδομένα που διακινούνται καθημερινά μέσα από τις συγκεκριμένες εφαρμογές και συγκεκριμένα στις δύο από τις πιο δημοφιλείς από αυτές όπως είναι το WhatsApp και το Telegram. Και αυτό ακριβώς θα κάνουμε στα πλαίσια αυτής της πτυχιακής εργασίας.

Κεφάλαιο 2. Συλλογή δεδομένων

Σε αυτή την ενότητα περιγράφεται η μεθοδολογία που ακολουθήθηκε για την συλλογή των δεδομένων, καθώς και οι περιορισμοί και τα ηθικά ζητήματα που προκύπτουν από την συλλογή των δεδομένων αυτών.

2.1 Μεθοδολογία συλλογής δεδομένων

2.1.1 Μεθοδολογία συλλογής δεδομένων για την εφαρμογή WhatsApp

Για την συλλογή των δεδομένων από τα WhatsApp groups χρησιμοποιήθηκε ένας ηλεκτρονικός υπολογιστής σχετικά χαμηλών δυνατοτήτων καθώς και μια κινητή συσκευή με εγκατεστημένη την εφαρμογή WhatsApp. Για την κινητή συσκευή απαιτείται επιπλέον μια κάρτα SIM με έναν ενεργό τηλεφωνικό αριθμό έτσι ώστε να λάβουμε το SMS επιβεβαίωσης για την εγγραφή μας στην υπηρεσία.

Σαν πρώτο βήμα για την συλλογή των δεδομένων, θα χρειαστεί να εγγραφούμε σε μερικές ομάδες (public groups). Θα μπορούσε ενδεχομένως κάποιος ερευνητής να επιλέξει στοχευμένα ομάδες με συγκεκριμένα χαρακτηριστικά ανάλογα με το πεδίο της έρευνας του. Στην παρούσα εργασία έγινε προσπάθεια να επιλεγθούν τυχαίες ομάδες με τρόπο τέτοιο ώστε να καλύπτεται όλο το φάσμα των θεμάτων συζήτησης. Για την εύρεση των ομάδων έγινε χρήση της μηχανής αναζήτησης της Google καθώς και ιστοσελίδων που περιέχουν συνδέσμους για public groups. Αυτό το επιτεύχθηκε με την αναζήτηση συνδέσμων που περιέχουν την κατάληξη *chat.whatsapp.com*, και έγινε εγγραφή σε περίπου 280 public groups.

Μετά την εγγραφή μας στα group, κάναμε χρήση του WebWhatsApp, ένα API του οποίου η χρήση είναι ελεύθερη στο Github στον σύνδεσμο <https://github.com/mukulhase/WebWhatsApp-Wrapper>, και το οποίο μας βοηθάει να συλλέξουμε στον υπολογιστή μας τα μηνύματα από τα groups του WhatsApp. Για την χρήση του WebWhatsApp χρειάστηκε να εγκαταστήσουμε στο σύστημα μας ορισμένα εργαλεία όπως είναι το Gecko Driver και ένα εργαλείο αυτοματοποίησης προγράμματος περιήγησης το Selenium (αναλυτικότερα βλ. Κεφάλαιο 6.1). Πρέπει να έχουμε υπόψιν ότι για την σύνδεση του web interface (*web.whatsapp.com*) με τον λογαριασμό της συσκευής Android που θα χρησιμοποιήσουμε για να συλλέξουμε τα δεδομένα χρειάζεται κάθε φορά σύνδεση μέσω σάρωσης ενός κωδικού QR.

2.1.2 Μεθοδολογία συλλογής δεδομένων για την εφαρμογή Telegram

Όπως και στην περίπτωση του WhatsApp, για την συλλογή των δεδομένων από το Telegram χρησιμοποιήθηκε ένας ηλεκτρονικός υπολογιστής σχετικά χαμηλών δυνατοτήτων καθώς και μια τηλεφωνική συσκευή με μια ενεργή κάρτα SIM την οποία θα χρειαστούμε για να λάβουμε το SMS επιβεβαίωσης για την εγγραφή μας στην υπηρεσία Telegram.

Μετά την εγγραφή μας στην υπηρεσία θα χρειαστεί να εγγραφούμε σε ομάδες (public groups). Όπως για το WhatsApp έτσι και στην περίπτωση του Telegram έγινε προσπάθεια να επιλεγθούν ομάδες που καλύπτουν όλο το φάσμα των θεμάτων συζήτησης. Συνολικά έγινε εγγραφή σε περίπου 40 ομάδες. Για την εύρεση τους έγινε χρήση της μηχανής αναζήτησης της Google καθώς και ιστοσελίδων που περιέχουν συνδέσμους για public groups και channels.

Για να συλλέξουμε τα δεδομένα κάναμε χρήση της Telethon. Η Telethon είναι μια βιβλιοθήκη της Python για το API του Telegram. Για να αλληλεπιδράσουμε με το Telegram API θα πρέπει αρχικά να αποκτήσουμε ένα *api_id* και ένα *api_hash*. Για την απόκτηση τους ακολουθήσαμε τις οδηγίες από τον επίσημο ιστότοπο: https://core.telegram.org/api/obtaining_api_id. Εκτός του *api_id* και του *api_hash* θα χρειαστούμε και ένα *token*. Για την απόκτηση του ξεκινάμε έναν διάλογο με τον

BotFather και αφού μας ζητηθεί να δώσουμε ένα *name* και ένα *username* για το bot μας, μας αποστέλλεται το *token*. Όλα τα παραπάνω τα αποθηκεύσαμε στο *setting.ini* (Script 1.), μαζί με την μεταβλητή *phone*, στην οποία καταχωρούμε τον αριθμό της SIM που χρησιμοποιήσαμε για να κάνουμε εγγραφή στην υπηρεσία Telegram.

Στο script *collector.py* που χρησιμοποιούμε για την συλλογή των μηνυμάτων θα παρατηρήσουμε ότι υπάρχουν κάποιοι όροι οι οποίοι μπορεί να είναι λίγο συγκεχυμένοι. Για να γίνει πιο εύκολη η κατανόηση τους, διευκρινίζουμε ότι ο όρος *dialogs* αντιπροσωπεύει τις συνομιλίες από το ιστορικό των συνομιλιών για κάθε group, ο όρος *chats* αντιπροσωπεύει τις ομάδες και τα channels, ο όρος *messages* είναι το τελευταίο μήνυμα που αποστέλλεται σε κάθε συνομιλία και τέλος *users* είναι οι χρήστες με τους οποίους έχουμε άμεση (εκτός group) επικοινωνία ή ο αποστολέας του τελευταίου μηνύματος σε κάποια από τις ομάδες που είμαστε εγγεγραμμένοι (αναλυτικότερα τα βήματα που ακολουθήθηκαν βλ Κεφάλαιο 5.).

2.2 Ηθικά ζητήματα

Είναι σαφές ότι η παραπάνω μεθοδολογία έχει την δυνατότητα συλλογής μεγάλου όγκου δεδομένων-μηνυμάτων που αποστέλλονται από πλήθος χρηστών παγκοσμίως. Εγείρονται επομένως ορισμένα ζητήματα απορρήτου που πρέπει να ληφθούν υπόψη. Αυτό που θα πρέπει να τονιστεί ιδιαίτερω είναι ότι δεν θα πρέπει να δημοσιεύονται οι τηλεφωνικοί αριθμοί που συλλέγονται. Επομένως για την προστασία των χρηστών θα πρέπει να αποκρύπτεται ο κωδικός χώρας από τον τηλεφωνικό τους αριθμό και να συλλέγεται ως αναγνωριστικό μόνο το ψευδώνυμο τους ή το κομμάτι του τηλεφωνικού τους αριθμού αφού τους έχει αφαιρεθεί πρώτα ο κωδικός χώρας. Επίσης θα ήταν θεμιτό, μετά την ολοκλήρωση της διαδικασίας της εξαγωγής δεδομένων να διαγράφονται τα δεδομένα από την συσκευή WhatsApp ή Telegram που χρησιμοποιήθηκε για την συλλογή τους.

Ιδιαίτερη προσοχή πρέπει να δοθεί και στο είδος των groups από τα οποία θα συλλέγονται τα δεδομένα. Αν και όλες οι ομάδες είναι δημόσιες και οι χρήστες τους ξέρουν ότι τα μηνύματα τους κοινοποιούνται σε άτομα άγνωστα προς αυτούς, θα έπρεπε να αποφεύγεται η συλλογή μηνυμάτων κυρίως από group με περιεχόμενο ενηλίκων καθώς υπάρχει κίνδυνος λήψης απαγορευμένου υλικού κατά την λειτουργία αυτόματης συλλογής δεδομένων.

Τέλος να τονίσουμε ότι στην πολιτική απορρήτου για τους χρήστες του WhatsApp δηλώνεται ρητά ότι οι χρήστες μοιράζονται τα μηνύματα και τις πληροφορίες τους (συμπεριλαμβανομένου του τηλεφωνικού τους αριθμού) με τα υπόλοιπα μέλη των ομάδων (<https://www.whatsapp.com/legal/>). Στην περίπτωση του Telegram κανένα μέρος δεν μπορεί να δει τον αριθμό τηλεφώνου κάποιου άλλου εκτός εάν αυτό έχει επιτραπεί από τον χρήστη από τις ρυθμίσεις απορρήτου (<https://telegram.org/faq>).

Κεφάλαιο 3. Telegram

Το Telegram είναι μια υπηρεσία άμεσης ανταλλαγής μηνυμάτων η οποία βασιζόμενη στην τεχνολογία του νέφους (Cloud) επιτρέπει στους χρήστες να του να στέλνουν μηνύματα πολυμέσων και να πραγματοποιούν κλήσεις φωνής και βίντεο.

Ιδρύθηκε από τους Νικολάι και Πάβελ Ντούροφ το 2013, οι οποίοι ήταν και ιδρυτές της ρωσικής πλατφόρμας κοινωνικών μέσων VK (η μεγαλύτερη στην εγχώρια αγορά). Με αρχική έδρα την Αγία Πετρούπολη, η ομάδα του Telegram έχει επανειλημμένα αλλάξει έδρα για να αποφύγει προβληματικούς κανονισμούς. Η επιχείρηση βρίσκεται επί του παρόντος στο Ντουμπάι, μετά από σειρά περιορισμών που αντιμετώπισαν έχοντας έδρα στο Λονδίνο, το Βερολίνο και τη Σιγκαπούρη.

Το Telegram δημιουργήθηκε με σκοπό να αμφισβητήσει την υπεροχή του WhatsApp στον χώρο και διαφέρει από το τελευταίο, επιτρέποντας στους χρήστες του να έχουν πρόσβαση στους λογαριασμούς τους από πολλές συσκευές. Οι ιδρυτές της ισχυρίζονται επίσης ότι η υποδομή των κέντρων δεδομένων και η κρυπτογράφηση τους τους καθιστούν ασφαλέστερο μέσω από τον ανταγωνιστή τους που πλέον ανήκει στο Facebook.

Οι χρήστες εκτός των συνηθισμένων συνομιλιών μπορούν να έχουν και «μυστικές συνομιλίες». Στην περίπτωση αυτή τα μηνύματα αποθηκεύονται στις συσκευές των χρηστών σε αντίθεση με τις κανονικές συνομιλίες που αποθηκεύονται στο νέφος (Cloud). Οι χρήστες έχουν επίσης την δυνατότητα μέσω των ρυθμίσεων να επιλέξουν την αυτοκαταστροφή των μηνυμάτων όπως ακριβώς συμβαίνει και με το Snapchat.

Το Telegram, ωστόσο, έχει επικριθεί από τους ειδικούς στον τομέα της ασφάλειας, οι οποίοι επισημαίνουν την έλλειψη προεπιλεγμένης από άκρο σε άκρο κρυπτογράφησης και του ιδιόκτητου πρωτόκολλου κρυπτογράφησης. Οι κωδικοί μιας χρήσης (one-time login passwords) που αποστέλλονται μέσω SMS έχουν παρακαμφθεί από τους χάκερ σε διάφορες χώρες όπως το Ιράν, τη Ρωσία, τη Βραζιλία και τη Γερμανία (στην τελευταία περίπτωση από την αστυνομία).

Το Telegram χρησιμοποιεί ένα ανοιχτό API και χαιρετίζει τους προγραμματιστές να δημιουργήσουν τις δικές τους εφαρμογές (Telegram apps), δυνατότητα την οποία αξιοποίησε η Samsung με την δημιουργία του Socializer Messenger το 2015.

Οι προγραμματιστές έχουν επιπλέον την δυνατότητα να δημιουργήσουν και τα δικά τους bots και δίνεται επίσης η δυνατότητα στους χρήστες να μπορούν να δημιουργήσουν τα δικά τους stickers. Εκτός από την αποστολή μηνυμάτων, οι χρήστες μπορούν να εγγραφούν σε 'κανάλια' (channels), τα οποία επιτρέπουν στους δημιουργούς τους να στέλνουν μηνύματα σε συνδρομητές.

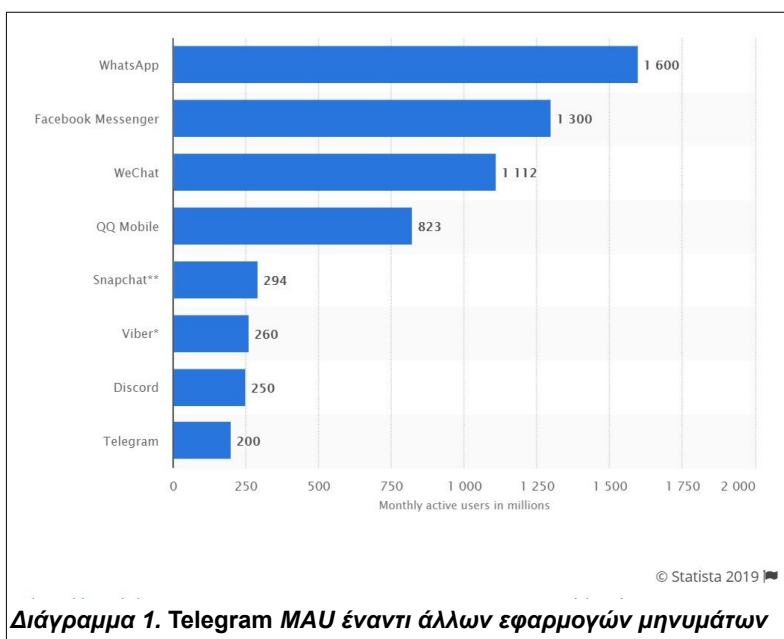
Το Telegram δεν έχει έσοδα από διαφημίσεις, δηλώνοντας ότι η πρόσβαση σε προσωπικά δεδομένα που αποκτούνται από άτομα ή εταιρίες που θέλουν να διαφημιστούν θα ήταν αντίθετη προς το ήθος του. Η μέχρι τώρα χρηματοδότηση του έρχεται ιδιωτικά από τον Pavel Durov ο οποίος έχει δηλώσει ότι θα στραφεί σε δωρεές χρηστών ή σε ένα μοντέλο τύπου freemium για να αυξήσει τα έσοδα εάν αυτό είναι απαραίτητο. Ο ίδιος έχει δηλώσει, "Η δημιουργία κέρδους δεν θα είναι ποτέ ο τελικός στόχος για το Telegram".

Τον Μάρτιο του 2018, το Telegram λάνσαρε ιδιωτικά το κρυπτονόμισμα Gram για λίγους και επιλεγμένους επενδυτές. Μέχρι στιγμής το κρυπτονόμισμα αυτό δεν έχει διατεθεί ακόμα στο κοινό.

Γενικά το Telegram λειτουργεί με κάποια σχετική μυστικότητα και υπογραμμίζει συνεχώς τη δέσμευσή του να απαλλαγεί από τις πιέσεις της αγοράς καθώς και από οποιονδήποτε περιορισμό που επιβάλλεται σε εθνικό επίπεδο. Συνεπώς, η ροή πληροφοριών είναι σχετικά περιορισμένη.

3.1 Στατιστικά χρηστών και χρήσης του Telegram

Τα τελευταία επίσημα στατιστικά στοιχεία χρηστών του Telegram χρονολογούνται από το Μάρτιο του 2018, όταν η εταιρεία ανακοίνωσε ότι είχε πετύχει τον αριθμό των 200 εκατομμυρίων μηνιαίως ενεργών χρηστών (MAU). Το Telegram αναμένει μερικά πολύ φιλόδοξα ορόσημα τα επόμενα χρόνια και έχει θέσει ως στόχο τους 1 δισεκατομμύριο χρήστες μέχρι το 2022. Πάντως φαίνεται να έχει ακόμα πολύ δρόμο μπροστά της καθώς παρατηρώντας το γράφημα στον Διάγραμμα 1. συμπεραίνουμε σαφώς ότι υστερεί ακόμη κατά πολύ σε σύγκριση με άλλες εφαρμογές του είδους του, όπως είναι το WhatsApp, το Facebook Messenger, το WeChat και η QQ. Από την άλλη, βλέπουμε ότι η χρήση του κινείται στα ίδια επίπεδα με την χρήση του Snapchat, του Viber και του Discord (έχοντας υπόψιν ότι ορισμένοι από αυτούς τους αριθμούς είναι εκτιμήσεις, δεδομένου ότι για παράδειγμα, το Snapchat δημοσιεύει μόνο στοιχεία ημερησίως ενεργών χρηστών (DAU)).



Σύμφωνα με την Sensor Tower, για το Telegram έχουν καταγράψει συνολικά 365 εκατομμύρια downloads σε παγκόσμια κλίμακα.

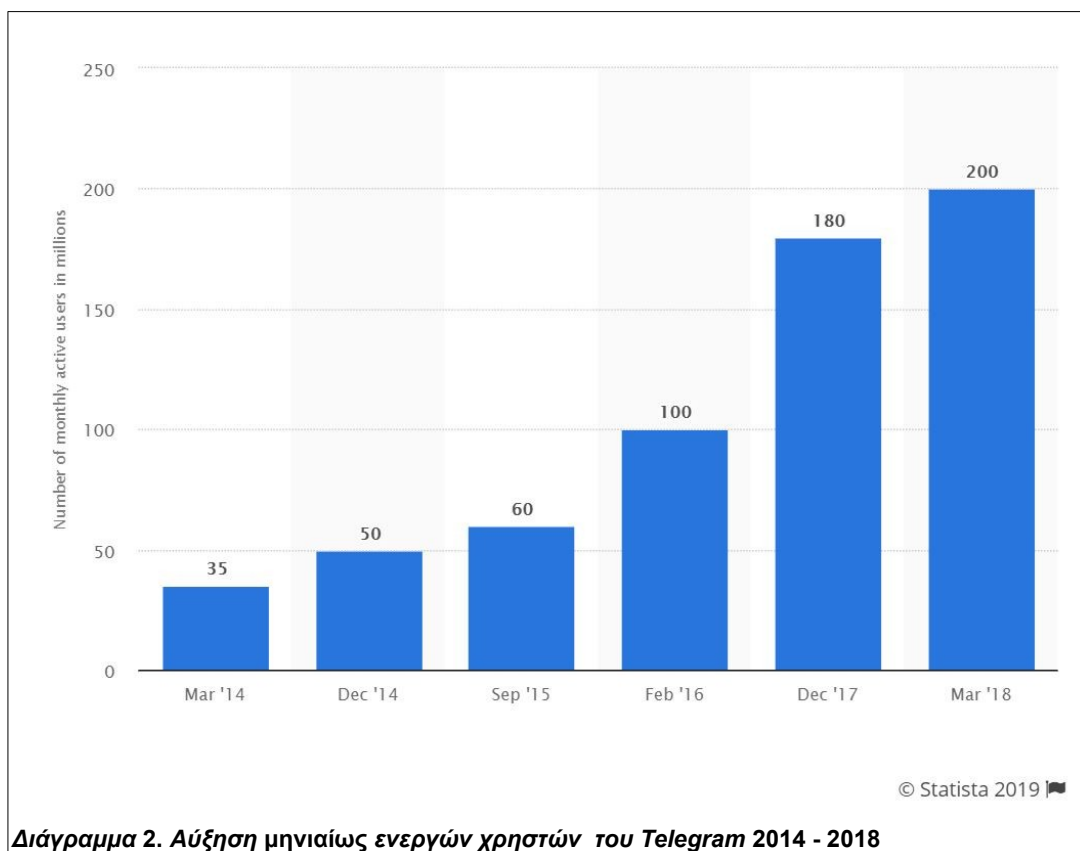
3.2 Ρυθμός αύξησης χρηστών του Telegram.

Το Telegram αποδείχθηκε αρκετά δημοφιλές στα πρώτα του βήματα καθώς έφτασε τους 100.000 χρήστες μέχρι τα τέλη Οκτωβρίου του 2013 αν και λανσαρίστηκε τον Αύγουστο του ίδιου έτους για το iOS και τον Οκτώβριο για το Android.

Έως το Μάρτιο του 2014, η εφαρμογή έφτασε τον αριθμό των 35 εκατομμυρίων μηνιαίως ενεργών χρηστών (MAU), από τους οποίους τα 15 εκατομμύρια ήταν ημερησίως ενεργοί χρήστες (DAU), και ο αριθμός αυτός έφτασε τους 50 εκατομμύρια χρήστες μέχρι το τέλος του έτους. Ο

αριθμός αυτός θα διπλασιαστεί μέχρι τον Φεβρουάριο του 2016 με μία εντυπωσιακή αύξηση από τον Σεπτέμβριο του 2015 και μετά.

Μέχρι το τέλος του 2017 ο αριθμός των χρηστών του Telegram έφτασε τους 180 εκατομμύρια χρήστες, και μαζί με τους 20 εκατομμύρια χρήστες που προστέθηκαν τους επόμενους μήνες έφτασε τον Μάρτιο του 2018 τον αριθμό των 200 εκατομμυρίων χρηστών (βλ. Διάγραμμα 2.).

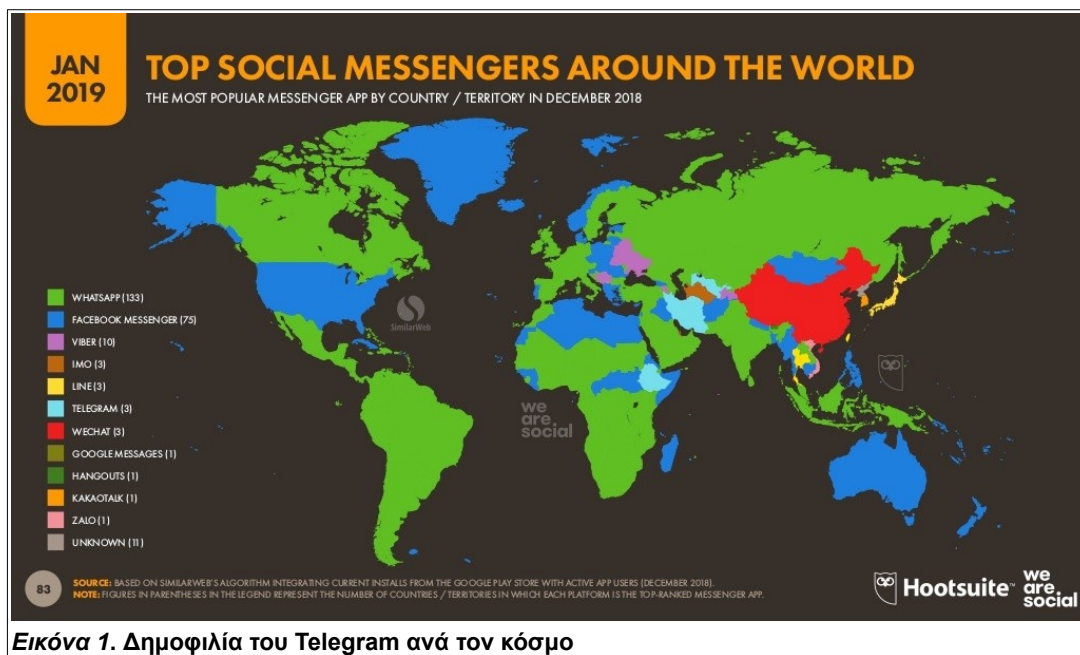


Τον Μάρτιο του 2019, αναφέρθηκε ότι στο Telegram προστέθηκαν 3 εκατομμύρια νέοι χρήστες σε διάστημα 24 ωρών κατά τη διάρκεια μιας "διαρροής" χρηστών του Facebook, συμπεριλαμβανομένου και του WhatsApp.

3.3 Σε ποιες χώρες είναι πιο δημοφιλές το Telegram;

Η δέσμευση του Telegram για την προστασία των δεδομένων των χρηστών της σημαίνει ότι τα δημογραφικά δεδομένα για τους χρήστες της είναι σχετικά περιορισμένα.

Αυτό που γνωρίζουμε έως τώρα είναι ότι σε μερικές χώρες, το Telegram είναι η πιο δημοφιλής εφαρμογή ανταλλαγής μηνυμάτων. Οι χώρες αυτές είναι η Αιθιοπία, το Ιράν και το Ουζμπεκιστάν.



Εικόνα 1. Δημοφιλία του Telegram ανά τον κόσμο

Το 2016, υπολογίστηκε ότι υπήρχαν 20 εκατομμύρια χρήστες του Telegram στο Ιράν. Το BBC είχε ανέφερε ότι αριθμός αυτός έφτασε τους 50 εκατομμύρια χρήστες τον Απρίλιο του 2018. Σε αυτό το σημείο, η εφαρμογή Telegram απαγορεύτηκε στη χώρα, όμως παρά την απαγόρευση οι χρήστες μπορούν να έχουν ακόμα πρόσβαση στην εφαρμογή μέσω VPN.

Το περιοδικό Wired ανέφερε ότι για πολλούς Ιρανοί, το Telegram έχει αντικαταστήσει πλήρως το διαδίκτυο μίας και είναι το μόνο μέσω το οποίο προσφέρει αφιltrάριστη πρόσβαση. Όπως προαναφέραμε οι Ιρανοί χρήστες του διαδικτύου, είναι εξοικειωμένοι σε μεγάλο βαθμό με την χρήση του VPN. Το 2017 ο διευθύνων σύμβουλος της Telegram, Pavel Durov, ανακοίνωσε ότι χορηγούσε επιχορηγήσεις bitcoin σε άτομα και οργανισμούς που διαχειριζόταν VPN.

Ένα χρόνο μετά την απαγόρευση του Telegram, τα ελεγχόμενα από το κράτος μέσα ενημέρωσης ανέφεραν ότι οι άνθρωποι επέστρεφαν στο Telegram, μια κίνηση που φαίνεται να υποδηλώνει ένα κύμα αποδοκιμασίας. Η διείσδυση του Telegram στο Ιράν εκτιμάται περίπου στο 56%. Το ποσοστό αυτό είναι κάτω του 60% το οποίο ήταν το ποσοστό προ της απαγόρευσης, αλλά αρκετά πάνω από το 47% το οποίο ήταν το ποσοστό στα αρχικά βήματα της εφαρμογής στην χώρα.

Έχει παρατηρηθεί επίσης ραγδαία αύξηση της δημοτικότητας του Telegram και σε άλλες χώρες σε στιγμές που οι χρήστες αναζητούσαν μια εναλλακτική πηγή πληροφόρησης. Για παράδειγμα, το 2014 στη Νότια Κορέα, μετά την ανακοίνωση της κυβέρνησης για καταστολή του

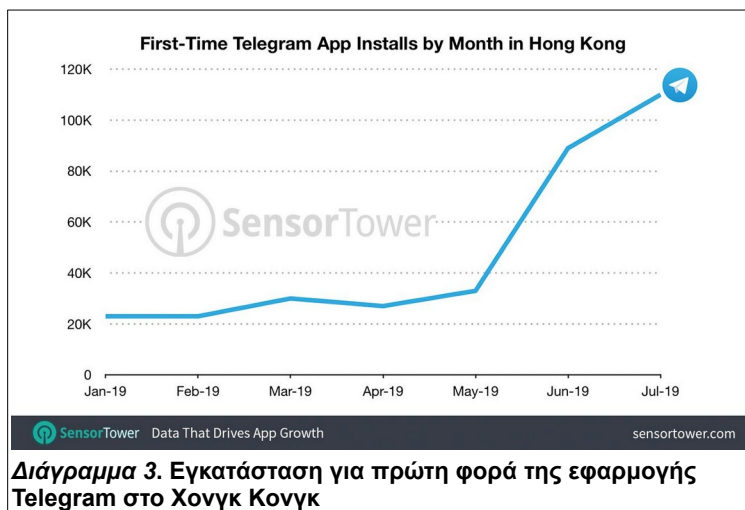
λόγου που θεωρείται προσβλητικός για τον πρόεδρο ή για ειδήσεις που βασίζονται σε «φήμες», αυτό είχε ως αποτέλεσμα μια τεράστια ροή χρηστών προς στο Telegram. Υπολογίζεται ότι σε διάστημα 7 ημερών περίπου 1,5 εκατομμύριο νέοι χρήστες εγγράφηκαν στην εφαρμογή.

Στη Βραζιλία σε μια περίοδο σύντομης απαγορεύσεις της εφαρμογής WhatsApp παρατηρήθηκε μεγάλος αριθμός χρηστών να συρρέει στο Telegram σε αναζήτηση μιας εναλλακτικής υπηρεσίας ανταλλαγής μηνυμάτων. Η απαγόρευση έλαβε χώρα το 2015 για μια χρονική περίοδο 48 ωρών και ήταν τόσο μεγάλος ο αριθμός των Βραζιλιάνων που έτρεξαν να εγγραφούν στο Telegram και σε τόσο σύντομο χρονικό ώστε το Telegram να αντιμετωπίσει προβλήματα με την αποστολή των κωδικών ασφαλείας στους χρήστες αυτούς. Συνολικά, 5,7 εκατομμύρια νέοι χρήστες εγγράφηκαν στο Telegram σε μια μόνο ημέρα.

Μια άλλη απαγόρευση διάρκειας 72 ωρών τον επόμενο χρόνο έστειλε άλλους 7 εκατομμύρια Βραζιλιάνους χρήστες στο Telegram σε μια μόλις ημέρα. Η απαγόρευση οφειλόταν στην εφαρμογή πλήρους κρυπτογράφησης από άκρο σε άκρο στην εφαρμογή WhatsApp πράγμα που καθιστούσε αδύνατη τη συμμόρφωση τους εταιρίας με μια δικαστική απόφαση που ήθελε να είναι οι συνομιλίες της εφαρμογής στη διάθεση των εισαγγελείων. Παρόλα αυτά θα πρέπει να σημειωθεί ότι στο σημείο αυτό της διεύθυνση του Telegram σε ποσοστό 18% (MUA) στην Βραζιλίας δεν ήταν ακόμα ανταγωνιστική ως προς το ποσοστό 96% του WhatsApp's και το ποσοστό 82% του Facebook Messenger.

Η τελευταία μεγάλη αύξηση αριθμού χρηστών του Telegram παρατηρήθηκε στο Χονγκ Κονγκ. Οι πολιτικές αναταραχές που άρχισαν τον Ιούλιο του 2019 σημείωσαν αύξηση 323% σε ετήσια βάση στην εγκατάσταση της εφαρμογής από χρήστες που εγγράφονταν στην εφαρμογή για πρώτη φορά. Συνολικά, στην εφαρμογή πρόστέθηκαν περίπου 110.000 νέοι χρήστες στο Χονγκ Κονγκ, σύμφωνα με τα στατιστικά στοιχεία του Sensor Tower. Και αυτό συνέβει παρά το γεγονός ότι το Telegram απαγορεύεται επισήμως στην Κίνα, γεγονός που σημαίνει ότι η χρήση του απαιτεί VPN.

Την πρώτη εβδομάδα του Αυγούστου του 2019, στο Telegram πρόστέθηκαν 41.000 νέοι χρήστες στο Χονγκ Κονγκ ποσοστό αυξημένο σε ποσοστό 100% σε σύγκριση με την πρώτη εβδομάδα του Ιουνίου του 2019 (βλ. *Διάγραμμα 3.*).



Το Telegram διαδραμάτισε καθοριστικό ρόλο στην οργάνωση των διαμαρτυριών που έγιναν στην χώρα. Στα τέλη του Αυγούστου του 2019, το Telegram ανακοίνωσε ότι επεξεργάζονταν μία κίνηση που θα επέτρεπε στους χρήστες να αποκρύπτουν τον αριθμό του τηλεφώνου τους. Με αυτόν τον τρόπο θα ήταν πιο δύσκολο για τις αρχές να εντοπίζουν τους συμμετέχοντες σε συνομιλίες ομάδας αναζητώντας τηλεφωνικούς αριθμούς. Η κίνηση αυτή έγινε μετά από προειδοποιήσεις ειδικών σχετικά με ένα πιθανό ελάττωμα στην εφαρμογή που θα επέτρεπε στις μυστικές υπηρεσίες πληροφοριών της χώρας να εντοπίζουν άτομα που εμπλέκονται στις μαζικές διαδηλώσεις. Αξίζει να αναφέρουμε ότι στα γκρουπ του Telegram μπορούν να εγγραφούν έως και 200.000 χρήστες.

Εκτός των χωρών που αναφέραμε, από τον χάρτη της Εικόνας 1. συμπεραίνουμε ότι το Telegram είναι επίσης μια αρκετά δημοφιλής εφαρμογή τόσο στην Αιθιοπία όπως και στο Ουζμπεκιστάν.

3.4 Στατιστικά στοιχεία για τα κανάλια (channel) του Telegram

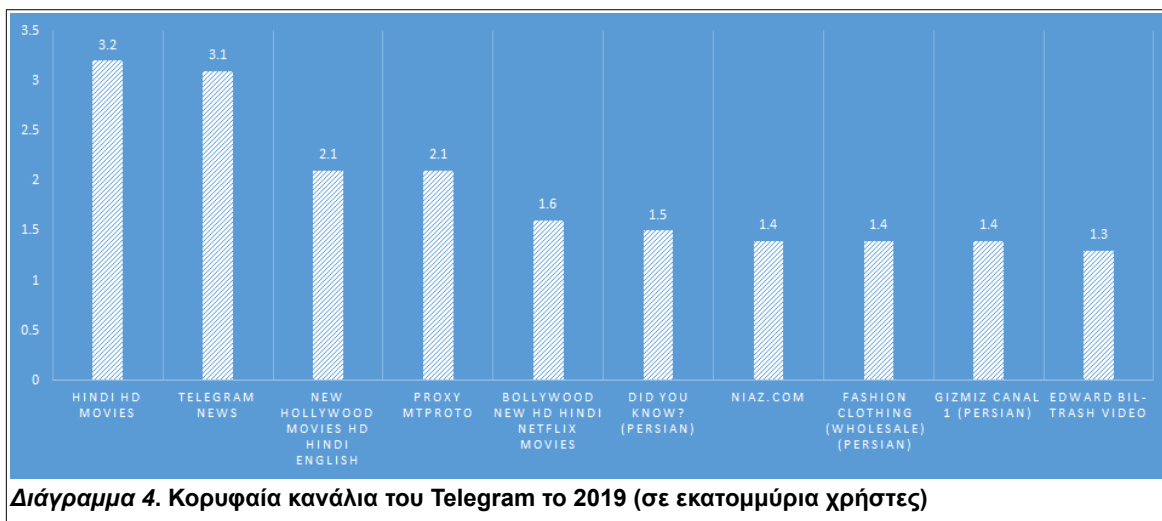
Τα κανάλια του Telegram είναι στην ουσία ομάδες μηνυμάτων μονής κατεύθυνσης όπου ένας διαχειριστής μπορεί να στέλνει μηνύματα σε όλα τα μέλη του καναλιού ταυτόχρονα. Δεν υπάρχει κάποιο όριο στον αριθμό των μελών που είναι εγγεγραμμένοι σε κάποιο κανάλι.

Παρόλο που τα κανάλια αυτά δεν είναι διαδραστικά, οι χρήστες έχουν την δυνατότητα να εγκαταστήσουν κάποια bots σε αυτά ή και κάποιες άλλες λειτουργίες οι οποίες επιτρέπουν στα κανάλια να έχουν μια πρόσθετη λειτουργικότητα. Τέτοιες λειτουργίες είναι για παράδειγμα η εγκατάσταση bots για feedbacks, η δημιουργία δημοσκοπήσεων, η χρέωση συνδρομών για συμμετοχή σε ένα ιδιωτικό κανάλι κ.α..

Όπως και στο YouTube, το κορυφαίο κανάλι σε πλήθος συνδρομητών στο Telegram εξυπηρετεί την αγορά της Ινδίας και χωρών με μεγάλο πληθυσμό Ινδών. Μέχρι τον Νοέμβριο του 2019 το κανάλι HINDI HD MOVIES του Telegram μετρούσε 3,2 εκατομμύρια συνδρομητές. Είναι πολύ πιθανό ότι τουλάχιστον ένα μέρος αυτού του κοινού κατοικεί στο Ιράν. Τα κανάλια New Hollywood Movies HD Hindi English (2.1 million) και Bollywood New HD Hindi Netflix Movies (1.6 million) εξυπηρετούν την ίδια αγορά.

Το επίσημο κανάλι ειδήσεων του Telegram (@Telegram) είναι το δεύτερο δημοφιλέστερο κανάλι, όχι κατά πολύ πίσω από το κανάλι HINDI HD MOVIES, με 3,1 εκατομμύρια συνδρομητές. Ο διακομιστής μεσολάβησης (Proxy Server) MTPProto υπολογίζεται ότι έχει περίπου 2,1 εκατομμύρια χρήστες, και είναι χρήσιμο εργαλείο για όσους επιθυμούν να αποκτήσουν πρόσβαση σε περιεχόμενο που έχει αποκλειστεί στη χώρα τους.

Τα υπόλοιπα από τα 10 κορυφαία κανάλια του Telegram είναι κυρίως κανάλια στην περσική γλώσσα με περιεχόμενο σχετικό με την εκπαίδευση (Did You Know?), την ψυχαγωγία (NiazCom, Gizmiz Canal 1) και το shopping (Fashion Clothing). Και η δεκάδα κλείνει με ένα ρώσικο κανάλι ψυχαγωγίας, το Edward Bil-Trash video. Σημειωτέον ότι την ρώσικη γλώσσα ομιλεί επίσης ένα σημαντικό ποσοστό των κατοίκων του Ουζμπεκιστάν. Στο *Διάγραμμα 4*. που παρατίθεται βλέπουμε τα 10 κορυφαία κανάλια με βάση τον αριθμό συνδρομητών έως και τον Σεπτέμβριο του 2019.



Κεφάλαιο 4. WhatsApp

Το WhatsApp είναι μια εφαρμογή ανταλλαγής μηνυμάτων για smartphones που δημιουργήθηκε το 2009 από δύο πρώην υπαλλήλους του Yahoo, τον Brian Acton και τον Jan Koum.

Ο Koum μετά την αγορά ενός iPhone στις αρχές του 2009, αντιλήφθηκε τις δυνατότητες που προσφέρονταν με την κυκλοφορία της εφαρμογής μέσω του Apple App Store. Η εφαρμογή επικεντρώθηκε αρχικά στην κοινοποίηση κατάστασης του χρήστη. Όπως αποκάλυψε ο ίδιος ο Koum, κίνητρο για την ιδέα αυτή ήταν η επιθυμία του να μην χάνει εισερχόμενες κλήσεις του τις ώρες που βρίσκονταν στο γυμναστήριο.

Η εφαρμογή, που αναπτύχθηκε από τον Igor Solomennikov, αξιοποίησε τη νέα λειτουργία Push Ειδοποιήσεων (Push Notifications) της Apple για την ενημέρωση των δικτύων χρηστών με ενημερώσεις κατάστασης (Status Updates). Οι χρήστες την δυνατότητα αυτή την χρησιμοποίησαν σαν να ήταν μια υπηρεσία άμεσων μηνυμάτων, βάζοντας τις βάσεις για την εξέλιξη της εφαρμογής. Η έκδοση 2.0 του WhatsApp ενσωμάτωσε τη λειτουργία ανταλλαγής μηνυμάτων κάτι το οποίο έμελλε να είναι καθοριστικό για το μέλλον της εφαρμογής.

Το WhatsApp ήταν η μοναδική εφαρμογή την εποχή εκείνη, εκτός από το BBM της Blackberry, που παρείχε στους χρήστες της δωρεάν παροχή υπηρεσίας μηνυμάτων. Το μόνο που έπρεπε να κάνουν οι χρήστες ήταν να εγγραφούν στην εφαρμογή αποκλειστικά και μόνο με τον αριθμό του τηλεφώνου τους. Η επιτυχία της έκδοσης 2.0 του WhatsApp ήταν άμεση, φτάνοντας πολύ γρήγορα τους 250.000 ενεργούς χρήστες. Με τα δεδομένα αυτά ο Acton κατάφερε να πείσει έναν στενό κύκλο συναδέλφων του, πρώην υπαλλήλων της Yahoo, να συνεισφέρουν και να χρηματοδοτήσουν με 250.000 δολάρια στα αρχικά βήματα της εφαρμογής.

Αργότερα το ίδιο έτος το WhatsApp ενσωμάτωσε τα μηνύματα πολυμέσων και το επόμενο έτος λάνσαρε την λειτουργία του και για τις συσκευές που υποστήριζαν λειτουργικό Android. Από το σημείο εκείνο και μετά, η δημοφιλία του WhatsApp άρχισε να αυξάνεται κατακόρυφα με αποτέλεσμα να κατατάσσεται στις τρεις πιο δημοφιλείς εφαρμογές τόσο στο iOS όσο και στο Android. Μέχρι τον Οκτώβριο του 2011, αποστέλλονταν μέσω του WhatsApp ένα δισεκατομμύριο μηνύματα ημερησίως. Στις αρχές του 2013 το WhatsApp διέθετε πάνω από 200 εκατομμύρια ενεργούς χρήστες, και παρόλο του ότι ο αριθμός αυτός είναι αρκετά εντυπωσιακός, δύσκολα συγκρίνεται με τους 1,5 δισ. ενεργούς μηνιαίους χρήστες που είχε το WhatsApp μέχρι το τέλος του 2017.

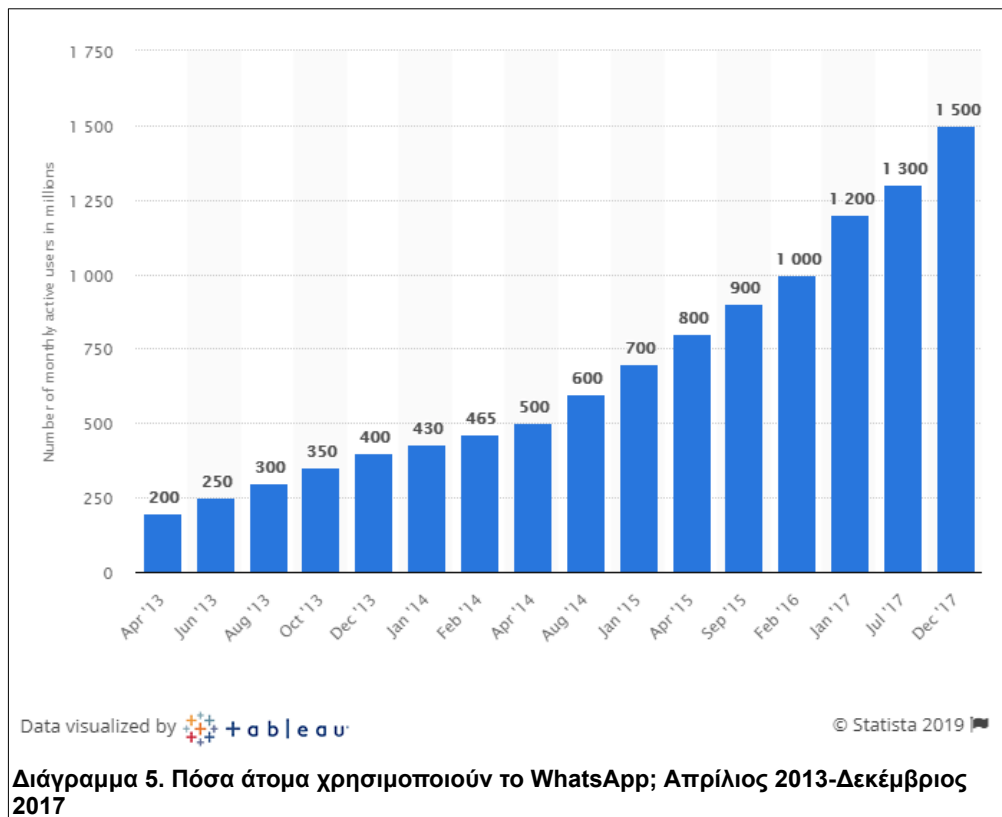
Τη χρονική εκείνη περίοδο δεν ήταν μόνο οι εταιρείες επιχειρηματικών κεφαλαίων που ενδιαφερόντουσαν για τις υπηρεσίες ανταλλαγής μηνυμάτων. Το Facebook έβλεπε το WhatsApp ως μια πιθανή απειλή και ως εκ τούτου αποφάσισε να λάβει δράση. Τον Φεβρουάριο του 2014, το Facebook προχώρησε στην εξαγορά του WhatsApp με κόστος τα 19 δισεκατομμύρια δολάρια, γεγονός το οποίο αποτελεί μια από τις μεγαλύτερες εξαγορές στην ιστορία στον χώρο της τεχνολογίας.

Με την εξαγορά αυτή προκλήθηκε κάποια διαμάχη βάση των ανησυχιών που υπήρχαν για τον λόγο του ότι η μητρική εταιρεία (Facebook) θα είχε πρόσβαση στα δεδομένα που διακινούνται στο WhatsApp, παρά τις αρχικές υποσχέσεις που έκαναν λόγο για αυξημένα επίπεδα κρυπτογράφησης. Για αυτόν ακριβώς τον λόγο και οι δύο ιδρυτές του WhatsApp αποφάσισαν να εγκαταλείψουν την εταιρεία.

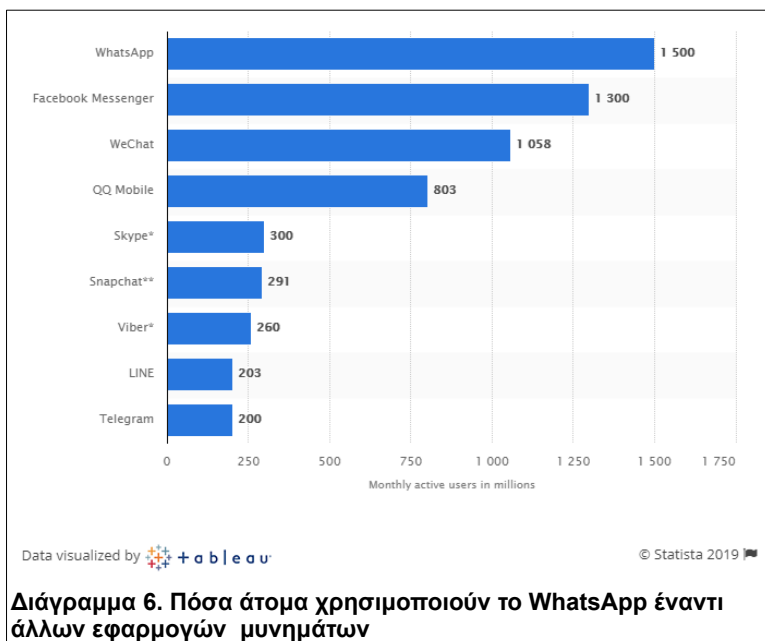
Παρόλα αυτά το WhatsApp παρέμεινε ο ηγέτης της αγοράς στον τομέα και πλέον, στην εφαρμογή έχουν ενσωματωθεί επιπλέον χαρακτηριστικά όπως είναι για παράδειγμα οι φωνητικές κλήσεις, οι βιντεοκλήσεις, οι ομαδικές κλήσεις (βίντεο και φωνή) και τέλος το λανσάρισμα στις αρχές του 2018 της υπηρεσίας WhatsApp Business.

4.1 Στατιστικά στοιχεία χρηστών του WhatsApp

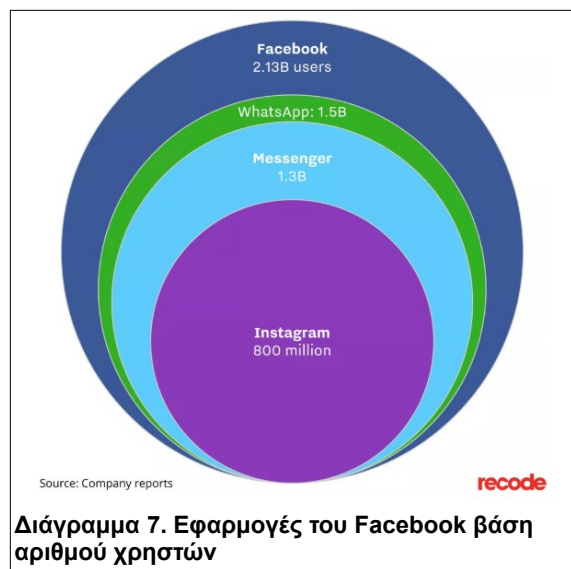
Τα τελευταία επίσημα στατιστικά στοιχεία του WhatsApp (που χρονολογούνται από το 4ο τρίμηνο του 2017) κατέγραψαν 1,5 δισ. μηνιαίους χρήστες σε περισσότερες από 180 χώρες. Αυτό δείχνει μια εξαιρετική δεκαετία για την εταιρία, με συνεχόμενη αύξηση των χρηστών της. Την περίοδο της εξαγοράς της εταιρίας από το Facebook, η εφαρμογή μετρούσε γύρω στους 500 εκατομμύρια χρήστες.



Σύμφωνα με τα τελευταία διαθέσιμα στατιστικά στοιχεία, του Οκτώβριου του 2018, το WhatsApp κατατάσσεται στην πρώτη θέση της κατάταξης των παγκόσμιων εφαρμογών messenger, με περίπου 200 εκατομμύρια περισσότερους χρήστες από το Facebook Messenger που κατατάσσεται στην δεύτερη θέση (βλ. Διάγραμμα 6.).



Επίσης τα στατιστικά αυτά στοιχεία κατατάσσουν το WhatsApp στην δεύτερη θέση σε αριθμό χρηστών με βάση των αριθμό των χρηστών που χρησιμοποιούν εφαρμογές που ανήκουν στο Facebook (βλ. Διάγραμμα 7.).



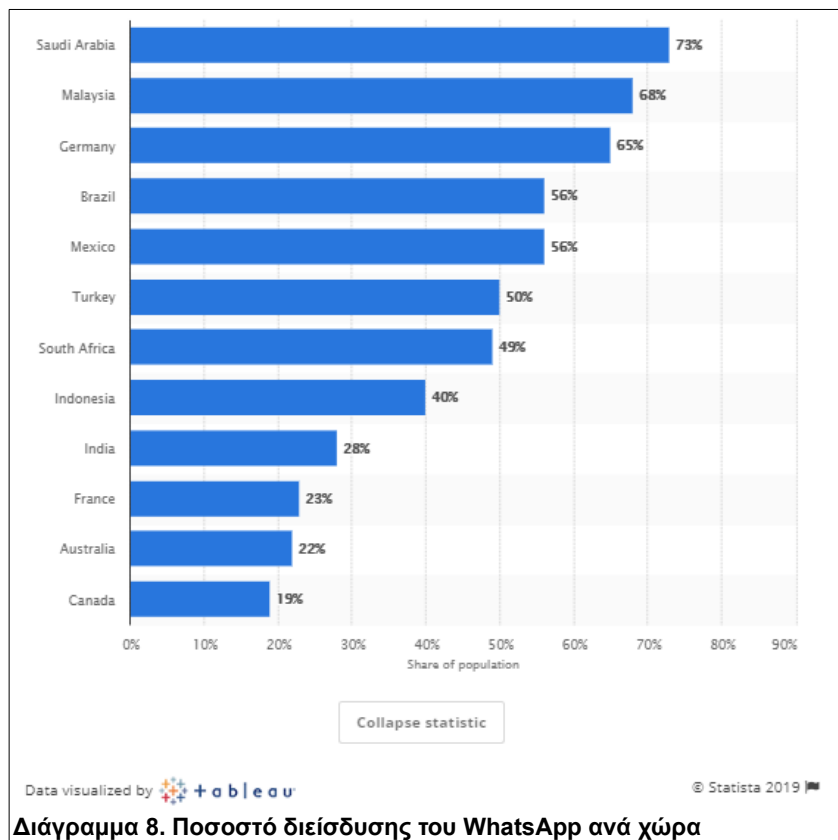
4.2 Χρήστες του WhatsApp ανά χώρα

Στην Εικόνα 2. βλέπουμε τις πιο δημοφιλείς εφαρμογές ανταλλαγής μηνυμάτων σε όλο τον κόσμο, με βάση την κατάταξή τους στο Google Play Store, κάτι που αφορά τις συσκευές Android. Όπως είναι σαφές από το πράσινο χρώμα, το WhatsApp είναι η κυρίαρχη εφαρμογή ανταλλαγής μηνυμάτων σε όλο τον κόσμο. Σύμφωνα με τα στοιχεία αυτά, βλέπουμε ότι είναι ελάχιστες οι χώρες στις οποίες το WhatsApp δεν είναι στην κορυφή, αλλά στις λίγες αυτές χώρες περιλαμβάνονται πολύ σημαντικές αγορές όπως είναι οι ΗΠΑ, η Αυστραλία και το μεγαλύτερο μέρος της Ασίας.



Εικόνα 2. Δημοφιλία του WhatsApp ανά τον κόσμο

Σύμφωνα με τα στατιστικά στοιχεία του WhatsApp που δημοσιεύονται στο Statista, η διείσδυση της εφαρμογής στην αγορά της Ινδίας είναι σχετικά χαμηλή. Εάν εξετάσουμε τους χρήστες του WhatsApp ανά χώρα με όρους διείσδυσης στην αγορά, μπορούμε να βρούμε τα υψηλότερα επίπεδα στη Γερμανία με 65%, τη Βραζιλία με 56%, τη Μαλαισία με 68% και τη Σαουδική Αραβία στην κορυφή της λίστας με 73% διείσδυση αγοράς, που σημαίνει ότι στα 100 άτομα που έχουν smartphone στην Σαουδική Αραβία οι 73 χρησιμοποιούν το WhatsApp. Τα στοιχεία αυτά αφορούν το τρίτο τρίμηνο του 2017 (βλ. Διάγραμμα 8.).



4.3 WhatsApp Groups

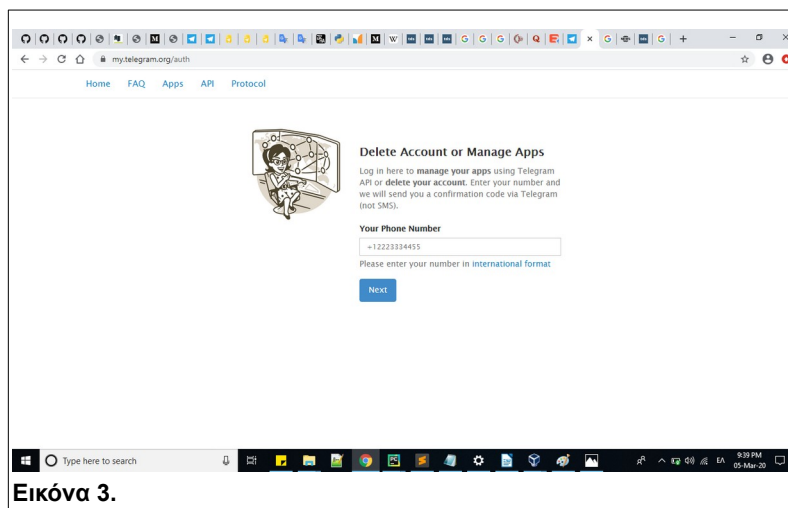
Τα groups του WhatsApp είναι ομαδικές συνομιλίες αποτελούμενες από χρήστες του WhatsApp. Επικεφαλής σε κάθε group είναι οι διαχειριστές τους. Ο διαχειριστής μπορεί να προσθέσει και να αφαιρέσει άτομα από την ομάδα ή ακόμα και να τους περιορίζει από το να κάνουν άλλα πράγματα. Εάν ένας διαχειριστής θέλει να προσθέσει ένα νέο μέλος στην ομάδα του, μπορεί να το κάνει χωρίς να χρειάζεται να έχει αποθηκευμένο τον αριθμό επικοινωνίας του νέου μέλους που θέλει να προσθέσει.

Ένας άλλος τρόπος για έναν χρήστη να γίνει μέλος σε μία ομάδα είναι τα invitation links. Σε περίπτωση που κάποιος έχει λάβει ένα invitation link ή έχει βρει ένα invitation link από κάποια ιστοσελίδα μπορεί να γίνει μέλος σε ένα group με την προϋπόθεση ότι εκείνη την χρονική στιγμή το group θα έχει λιγότερα από 256 άτομα, δεδομένου ότι κάθε group του WhatsApp έχει ένα μέγιστο όριο 256 ατόμων.

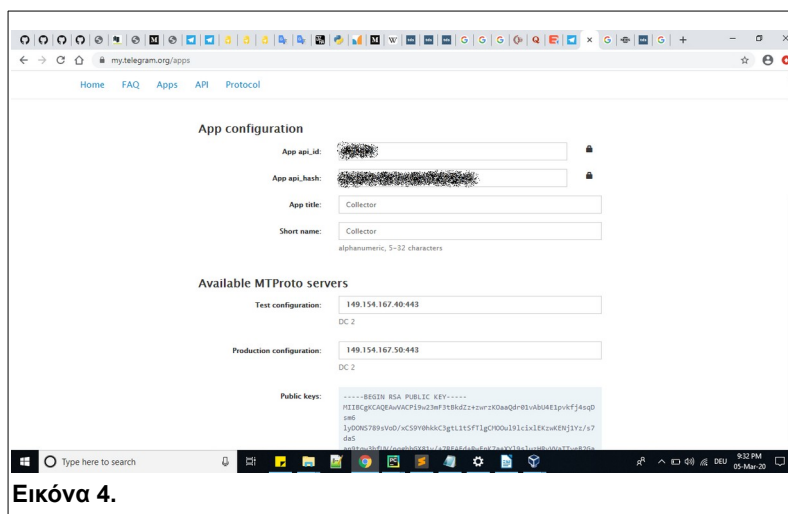
Εάν ο διαχειριστής θέλει να σταματήσει την λειτουργία του group, μπορεί να το κάνει ανά πάσα στιγμή. Το ίδιο δεν συμβαίνει με τα απλά μέλη του group. Τα απλά μέλη το μόνο που μπορούν να κάνουν είναι να διαγραφούν από το group όποτε αυτοί το θελήσουν. Όταν ένα μέλος βγαίνει από την ομάδα, εμφανίζεται ένα μήνυμα στα άλλα μέλη ότι το άτομο έχει εγκαταλείψει την ομάδα.

Κεφάλαιο 5. Συλλογή και ανάλυση δεδομένων Telegram

Για την συλλογή δεδομένων από το Telegram θα χρησιμοποιήσουμε την Telethon, μια βιβλιοθήκη της Python για το API του Telegram. Για να αλληλεπιδράσουμε με το Telegram API θα χρειαστεί να αποκτήσουμε ένα `api_id` και ένα `api_hash`. Αυτό το πετυχαίνουμε ακολουθώντας τις οδηγίες από τον επίσημο ιστότοπο: https://core.telegram.org/api/obtaining_api_id, Screenshots από τα βήματα που ακολουθήθηκαν βλέπουμε στην Εικόνα 3 και Εικόνα 4. Τα στοιχεία αυτά τα αποθήκευσα στο `settings.ini` (Script 1.). Για ευνοήτους λόγους τα πεδία των στοιχείων θα αποκρυφτούν. Όπου υπάρχει το σύμβολο αστεράκι (*) αντικαθιστούμε τα στοιχεία μας. Θα χρειαστούμε επίσης και ένα token το οποίο θα το πάρουμε αφού επικοινωνήσουμε με τον BotFather όπως βλέπουμε στην Εικόνα 5. Με την εντολή `/newbot` ξεκινάμε έναν διάλογο με τον BotFather και αφού μας ζητήσει να του δώσουμε ένα name και ένα username για το bot μας, μας στέλνει το token.



Εικόνα 3.

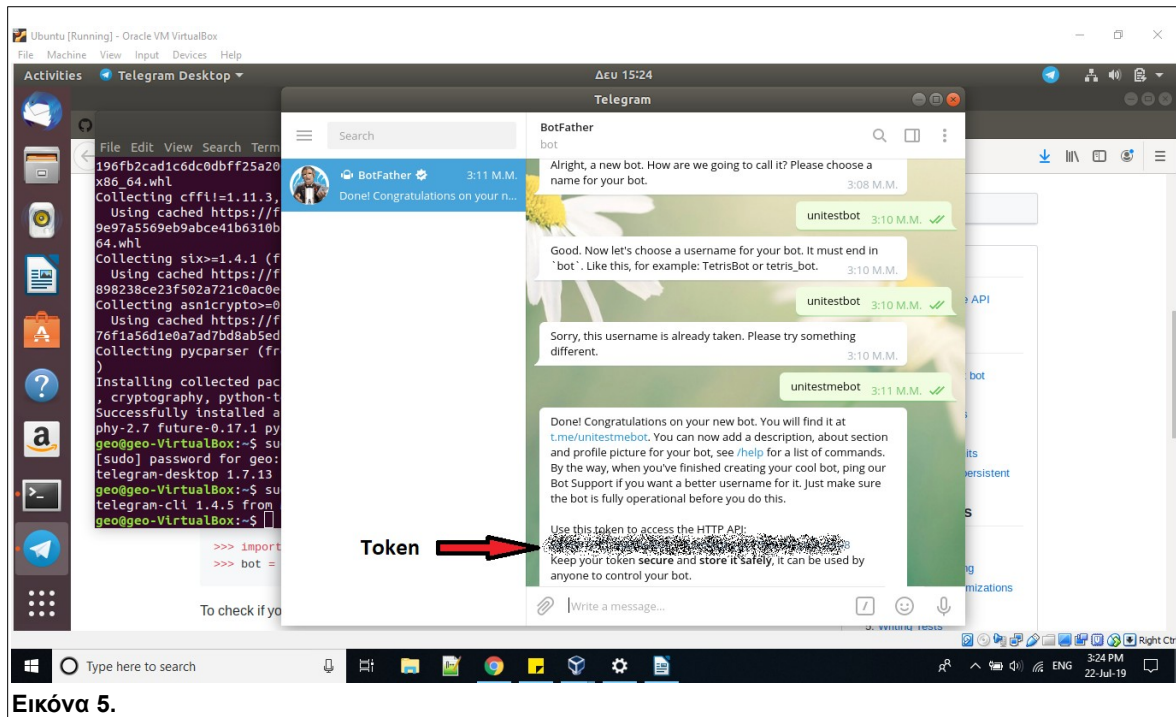


Εικόνα 4.

```
[telegram]
phone = *****
username = *****
token =*****

[telegram-api]
app_id = *****
hash = *****
```

Script 1. settings.ini



Εικόνα 5.

Στο PyCharm τρέχουμε το Script 2. (collector.py) το οποίο μας βοηθάει να συλλέξουμε τα μηνύματα.

```

from telethon.sync import TelegramClient
from telethon.tl.functions.messages import GetDialogsRequest, GetHistoryRequest
from telethon.tl.types import InputPeerEmpty, PeerChannel, InputPeerChannel, PeerChat, InputPeerChat,
PeerUser, \
    InputPeerUser
from telethon.errors.rpcerrorlist import SessionPasswordNeededError
import configparser
from utils import files_management
from telethon.tl.types.messages import Messages

# Fortwsh credentials apo to settings.ini opu exoume apothikeysi ta stoixeia mas
config = configparser.ConfigParser()
config.read('settings.ini')

api_id = config['telegram-api']['app_id']
api_hash = config['telegram-api']['hash']

phone = config['telegram']['phone']
username = config['telegram']['username']

# Dhmiourgia client kai syndesi
client = TelegramClient(username, api_id, api_hash)
client.connect()

# Authentication gia na mporesoyme na kanoume calls sto service
if not client.is_user_authorized():
    client.send_code_request(phone)
    try:
        client.sign_in(phone, input('Enter the code: '))
    except SessionPasswordNeededError:
        client.sign_in(password=input('Password: '))

def get_messages(number_of_messages = 100):
    """Methodos gia the sylogh mynhmatws apo ta group poy anhkei o xrhsths.
    H parametros number_of_messages orizei ton megisto aritho twn mynhmatwn
    pou tha sullexthoun apo kathe group. An ayth den oristei diaforetika apo
    to xrhsth, tote syllegei ta 100 pio prosfate mhnymata"""

    dialogs = client(GetDialogsRequest(
        offset_date=None,
        offset_id=0,
        offset_peer=InputPeerEmpty(),
        limit= 1000,
        hash= 7
    ))

```

```

chats = {}

for c in dialogs.chats:
    chats[c.id] = c

for d in dialogs.dialogs:
    peer = d.peer

    if isinstance(peer, PeerChannel):
        id = peer.channel_id
        channel = chats[id]
        name = channel.title

    elif isinstance(peer, PeerChat):
        id = peer.chat_id
        group = chats[id]
        name = group.title

    else:
        continue

print("Getting messages. Depending on the number of messages, this make take a while...")
messages = client.get_messages(name, number_of_messages)
print("Messages collected from channel ", name, " :", len(messages))
#Eggrafi minimaton ston fakelo "data"
files_management.export_messages_to_file("data", name, messages)

# Klhsh ths methodoy. H parakatw grammh tha prepei na mpei se comment an o
# programmatisths kanei import to arxeio se allh klassh. Sta plaisia ths ptyxiakhs,
# H sullogh egine anexarthta.
get_messages(100000)

```

Script 2. collector.py

Αφού συλλέξαμε τα δεδομένα, αρχικά ορίζουμε ότι κάθε μήνυμα του Telegram θα είναι ένα document. Στο Script 3. (document.py) βλέπουμε την κλάση Document. Τα values τις κλάσης είναι ο αποστολέας, η ημερομηνία – ώρα αποστολής και το κείμενο του μηνύματος.

```

class Document:

    def __init__(self, sender, datetime, text):
        """
        Arxikpoihs ths domhs arxeioy pou xrhsimopoihtai gia thn analysi.
        Kathe mynhma einai ena document.
        Ena document perigrafete apo ta pedia apostoleos, wra kai mera kai periexomeno
        """
        self.sender = sender
        self.datetime = datetime
        self.text = text

```

Script 3. document.py

Στο Script 4. (word_cloud_helper.py) ορίζουμε μια μέθοδο με την βοήθεια της οποίας θα πάρουμε την απεικόνιση των Word Clouds.

```
import matplotlib.pyplot as plt
from PIL import Image
from wordcloud import WordCloud

def save_word_cloud(word_frequencies, file_name):
    wc = WordCloud()
    wc.generate_from_frequencies(word_frequencies) #Paragei mia lista me tuples opou to proto melos einai
    enas oros, kai to deytero h syxnothta emfanishs tou p.x [('java', 15), ('scala', 12), ('spring', 12)]

    # Kanei display ta paragomena images
    plt.imshow(wc, interpolation='bilinear')
    plt.axis("off")
    plt.savefig(file_name+'.png', bbox_inches='tight')
```

Script 4. word_cloud_helper.py

Το Script 5. (text_processing.py) μας βοηθάει με την επεξεργασία των κειμένων των μηνυμάτων που συλλέξαμε.

```
import re
import string
from urllib.parse import urlparse
from nltk import stem, tokenize
from sklearn.feature_extraction import stop_words
import string
import nltk
from nltk.stem import WordNetLemmatizer

# TODO: ucomment tin prota for a pou tha treksei
# nltk.download('wordnet') #einai ena megalos dwrean lexical database to opoio prwsferei dinatotites
# limatopoihsis

wordnet_lemmatizer = WordNetLemmatizer() #Einai mia methodos pou kanei lhmata apo tis lekseis pou tou
# dinw... Edw to kataxwroume se mia metavliti gia na tin kaloume pio eykola
stemmer = stem.PorterStemmer() #LancasterStemmer() Yparxoun 3 To Porter einai to pio palio alla
# xrisimopoihtai syxna, Snowball(Porter2), to Lancaster (Paice-Husk) einai to pio epithetiko mporei na
# odigeisei se lathoi
stopwords = set(stop_words.ENGLISH_STOP_WORDS)
exclude = set(string.punctuation).union(stopwords) #To set pairnei mia list, ena tuple h ena dictionary kai
# epistrefei to perioxomeno tous (elements) sorted kai non-repeated. To union se dyo set() epistrefei to
# perioxomeno pou exoun ta dyo set xwris na epanalambanei ta element an ayto yparxei kai sta 2 set... To
# epistrefei mono mia fora
exclude.add("http") # To exclude einai ena set me string punctuation kai stopwords. To add() prwstheti
# set ayto mia leksi. An i leksei pou kanoume add yparxei idi, tote den tin prwstheti
exclude.add("https")
exclude.add("www")
exclude.add("com")
```

```

def is_URL(s):
    # Dedomenoy enos string to kanei parse kai epistrefei an einai url
    try:
        result = urlparse(s)
        return all([result.scheme, result.netloc, result.path]) # to all epistrefei True an ola ta paidia einai gemata
    except:
        return False

def normalize(s):
    return ' '.join(normalised_tokens(s)) # epistrefei tis lekseis pou pairnei apo tin lista pou epistrefei i
    # synartisi normalized_tokens(s) xwris komma metaksi tous diladi oles tis lekseis me ena keno metaksi tous

def normalised_tokens(s):
    words = tokenize.casual_tokenize(s.lower())
    final_words = []
    urls = []
    for word in words:
        if not is_URL(s):
            final_words.extend(tokenize.wordpunct_tokenize(word.lower().strip()))
        else:
            urls.extend(word)
            # final_words.append(word)
    return [stemmer.stem(wordnet_lemmatizer.lemmatize(w, pos="v")) for w in final_words if w not in
    # exclude] # kanei lemmatize kai epistrefei apo tis lekseis pou to lemma tous einai rima to rima pos= "v"
    # diladi verb px goes, go, went epistrefoun go

""" To input ths methodou tha prepei na einai *Document* datatype """
def process(document):
    # Apo th domh document, pairnw to pedio text kai to kanw lowercase
    text = document.text.lower()
    normalised = normalize(text) # Apo tin synartisi pou eftiaksa pio panw, epistrefei tis lekseis pou pairnei
    # apo tin lista pou epistrefei i synartisi normalized_tokens(s) xwris komma metaksi tous diladi oles tis lekseis
    # me ena keno metaksi tous
    # text = re.sub(r"[{}]" .format(string.punctuation), " ", text)
    return normalised

```

Script 5. text_proccesing.py

Με την βοήθεια του Script 6. (files_management.py) παίρνουμε τα documents από τα files με τις συνομιλίες του κάθε group. Όπως προαναφέραμε κάθε μήνυμα το ορίζουμε ως ένα document.

```

import errno
import json
import os

```

```

from data_types.document import Document

excluded_logs = {}

def get_files(root_directory):
    """
    Helper method gia na epistrafoyn ola ta arxeia me thn katalhxh log
    poy periexontai sto root directory. Ta arxeia log periexoyn ta mhnymata
    pou syllegontai. Gia kathe arxeio to full path epistrefetai. Kapoia logs
    periexoyn plhrofories ki oxi dedomena (excluded_logs). Ayta ta arxeia
    den epistrefontai.
    """
    files = []
    # r=root, d=directories, f = files
    for r, d, f in os.walk(root_directory):
        for file in f:
            """files pou periexoun text pliroforia me katali3i .log"""
            if file.endswith('.log') and file not in excluded_logs:
                files.append(os.path.join(r, file))

    return files

def get_documents(files):
    """
    Methodos gia thn anagnwsh twn arxeiwn poy periexoyn dedomena ki epistrofh
    sulloghs dict, opou to kleidi einai to onoma toy arxeioy kai to value
    einai array apo domes Document poy paristanoynta ta mhnymata sto arxeio.
    """
    documents = dict()
    print(files)
    file_counter = 0
    total_files = len(files)
    for file in files:
        file_counter += 1
        print("Reading document ", file_counter, " from ", total_files)
        documents_in_file = []
        messages = load_data(file)
        for id in messages:
            d = Document(messages[id]['sender'], messages[id]['datetime'], messages[id]['text']) #dimiourgo ena
            document d
            documents_in_file.append(d)

        documents[file] = documents_in_file
    return documents

```

```

def export_messages_to_file(directory, channel_name, messages):
    """
    Η parametros *directory* είναι ένα string του folder όπου
    θα αποθηκευθούν οι πληροφορίες που συλλεχθούν από το Telegram.

    Το channel_name είναι το όνομα του αρχείου και messages είναι η λίστα
    των μηνυμάτων που επιστρέφει το Telegram API
    """
    try:
        os.makedirs(directory)
    except OSError as e:
        if e.errno != errno.EEXIST:
            raise
    # Antikathistoyme special characters sto onoma toy arxeioy
    filename = directory + "/" + channel_name.replace("/", "") + ".log"
    # Η δομή data φιλοξενεί τα αρχεία που έχουμε για το channel.
    data = {}
    # Αν υπάρχει ήδη το αρχείο στο directory (έχουμε ήδη μηνύματα για αυτό το channel)
    if os.path.isfile(filename):
        # Διαβάζουμε ότι μηνύματα έχουμε ήδη
        data = load_data(filename)
    # Ανοίγουμε το αρχείο για εγγραφή
    file = open(filename, "w")
    # Και για κάθε μήνυμα
    for message in messages:
        # αν αυτό δεν είναι κάποιο μήνυμα ενημερωσε για
        # ενέργειες που πρέπει να κάνει ο χρήστης στην οθόνη
        if message.action:
            continue
        # Και δεν υπάρχει ήδη στα δεδομένα
        if str(message.id) not in data:
            # Το προσθέτουμε στα δεδομένα
            data[message.id] = dict(sender=message.from_id, datetime=str(message.date),
                                   text=message.text.replace("\n", "\t"))
    # Στις συνειδήσεις, γράφουμε όλα τα δεδομένα στο αρχείο
    # (kanontas overwrite ta perioxomena toy arxeioy)
    json.dump(data, file)
    file.flush()
    file.close()

def load_data(filename):
    file = open(filename, 'r')
    try:
        data = json.load(file)
    except:
        data = {}
    file.close()

```



```

return data

def update_progress(progress):
    print('\r[{}] {}%'.format('#' * (progress / 10), progress))

```

Script 6. files_management.py

Στο Script 7. (corpusCreator.py) ορίζουμε την κλάση CorpusCreator η οποία μας επιστρέφει ένα array με όλα τα documents ανεξαρτήτου του chat του οποίου ανήκουν.

```

from utils import files_management

class CorpusCreator:

    def __init__(self, directory):
        """
        Αρχικοποιήσh της δημιουργίας του corpus.
        Η parametros *directory* πρέπει να είναι ένα string με το path του directory
        που περιέχει τα data που συλλεχθήκαν
        """

        self.data_directory = directory
        self.corpus = []
        self.documents_per_chat = dict()

    def create(self):
        """
        Δημιουργεία του corpus για το training με vasi τα documents
        """
        all_logs = files_management.get_files(self.data_directory)
        self.documents_per_chat = files_management.get_documents(all_logs)
        for file in self.documents_per_chat:
            self.corpus.extend(self.documents_per_chat[file])

    def get(self):
        return self.corpus

```

Script 7. corpusCreator.py

Στο Script 8. (Kmeans.py) βλέπουμε την κλάση KmeansClustering όπου με την βοήθεια του αλγορίθμου K-means κάνουμε το clustering.

```

from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.cluster import KMeans
from preprocessing import text_processing
from utils import word_cloud_helper

class KMeansClustering:

    def __init__(self):

```

```

self.k = 0
self.vectorizer = None
self.clusters = None

def compute(self, corpus, clusters):
    """
    corpus: domh arxeiwn poy xrhsimopoiountai gia thn ekpaideysi
    clusters: o arithoms twn clusters poy dhmioyrgoyntai
    """
    self.k = clusters
    # Xrhsimopoiyme tf-idf gia na paristanoyme ta arxeia san vectors apo lexeis
    self.vectorizer = TfidfVectorizer(preprocessor=text_processing.process,
                                     encoding='utf-16', stop_words='english')
    # Gia ta arxeia sto corpus, ta metatrepoyme se vectors
    tfidf = self.vectorizer.fit_transform(corpus)
    # Kai 3ekiname ton K-means++ kanontas 100 to poly epanalysis gia na dhmioyrgsoyme ta cluster
    self.clusters = KMeans(n_clusters=self.k, init='k-means++', max_iter=100, n_init=1).fit(tfidf)

def predict_cluster(self, document):
    """Dedomenoy enos kainoyrioy mhnymatos (document),
    an den exoyme dhmioyrgsh ta clusters, emfanizoyme mhnyma"""
    if self.clusters is None:
        print("ERROR: clusters have not been computed.")
    else:
        """
        Alliws pairnoyme to vector toy mhnymatos,
        opws aytos ginetai project sto corpus (dhladh gia
        tis lexeis poy exoyme sto corpus) kai ypologizoyme
        se poio cluster brisketai pio konta o vector
        """
        return self.clusters.predict(self.vectorizer.transform([document]))

def print_top_terms(self, top_terms):
    # Me dedomeno ton arithmo top_terms
    print("Top terms per cluster:")
    # Pairnoyme ta kentra twn clusters kai taksinomoyme
    # tis apostaseis twn orwn
    order_centroids = self.clusters.cluster_centers_.argsort()[:, :-1]
    # Epishs apo to vector tou corpus pairnoyme ta onomata twn orwn (lexeis)
    terms = self.vectorizer.get_feature_names()
    # Kai gia kathe cluster
    for i in range(self.k):
        print("Cluster %d:\n" % i, end="")
        # Gia kathe deikth sta top terms
        for ind in order_centroids[i, :top_terms]:
            # Ektypwnoyme ton oro
            print(' %s' % terms[ind], end="")
        print()

```

```

def get_word_cloud(self):
    centroids = self.clusters.cluster_centers_.argsort()[:, :-1]
    terms = self.vectorizer.get_feature_names()
    # Omois me print top terms, alla pairnoyme to varos tis kathe
    # lexeis gia toys top 30 oroyis ana cluster max, kai ftiaxnoyme to world cloud
    for i in range(self.k):
        print("Cluster %d:\n plotting" % i, end="")
        word_cloud_terms = {}
        for ind in centroids[i, :30]:
            if self.clusters.cluster_centers_[i, ind] == 0.0:
                break
            word_cloud_terms[terms[ind]] = self.clusters.cluster_centers_[i, ind]
        word_cloud_helper.save_word_cloud(word_cloud_terms, "images/cluster_"+str(i))

```

Script 8. Kmeans.py

Στο Script 9. (LDA.py) βλέπουμε την κλάση LDATopicExtraction όπου με την βοήθεια του LDA κάνουμε το topic extraction.

```

from sklearn.decomposition import LatentDirichletAllocation
from sklearn.feature_extraction.text import CountVectorizer

from preprocessing import text_processing
from utils import word_cloud_helper

class LDATopicExtraction:

    def __init__(self):
        self.vectorizer = None
        self.lda = None

    """Dedomenou tou corpus me ta 'Document' kai ton arithmo twn topics ypologizei ta topics"""
    def compute(self, corpus, no_features, no_topics):
        self.vectorizer = CountVectorizer(preprocessor=text_processing.process,
                                         max_df=0.95, min_df=2, max_features=no_features,
                                         stop_words='english')
        tf = self.vectorizer.fit_transform(corpus)
        self.lda = LatentDirichletAllocation(n_components=no_topics, max_iter=20, learning_method='online',
                                           learning_offset=50., random_state=0).fit(tf)

    """Kanei predict gia to se poio topic anoikei to document"""
    def predict_topic(self, document):
        if self.lda is None:
            print("ERROR: topics have not been computed.")
        else:
            topics_prediction = self.lda.transform(self.vectorizer.transform([document]))
            result = {}

```

```

    for topic_idx, topic in enumerate(topics_prediction[0]):
        index = "topic_" + str(topic_idx)
        result[index] = topic
    return result

"""Ektypwnei tous top terms"""
def print_topics(self, top_terms):
    feature_names = self.vectorizer.get_feature_names()
    model = self.lda
    for topic_idx, topic in enumerate(model.components_):
        print("=====")
        print("Topic %d:" % (topic_idx))
        print("\n".join([feature_names[i] + ":" + str(round(topic[i], 2))
                          for i in topic.argsort()[:-top_terms - 1:-1]]))
        print("=====")

"""Epistrefei to word cloud"""
def get_word_cloud(self):
    feature_names = self.vectorizer.get_feature_names()
    model = self.lda
    for topic_idx, topic in enumerate(model.components_):
        word_cloud_terms = {}
        sorted_terms = topic.argsort()[:-30 - 1:-1]
        for i in sorted_terms:
            if topic[i] > 0:
                word_cloud_terms[feature_names[i]] = topic[i]
            else:
                break
        word_cloud_helper.save_word_cloud(word_cloud_terms, "images/LDA/topic_" + str(topic_idx))

```

Script 9. LDA.py

Τρέχουμε το Script 10. (tests.py) για να πάρουμε τα τελικά μας αποτελέσματα (clusters) με την βοήθεια του αλγόριθμου K-means .

```

from data_types.document import Document
from preprocessing import corpusCreator
from preprocessing import KMeans

# Diavazei ta documents
from utils import files_management

corpus = corpusCreator.CorporusCreator('data')
corpus.create()

corpus_documents = corpus.get()

# Dimiourgei ta clusters - 12 clusters stin periptosi mas
kmeans = KMeans.KMeansClustering()
kmeans.compute(corpus_documents, 12)

```

```

# Ektypwnei tous top 10 orous ana cluster
kmeans.print_top_terms(10)

# Dhmiourgei ena wordCloud gia to kathe cluster kai ta apothikeyei stous fakelous
kmeans.get_word_cloud()

# Vriskei to cluster enos test document
test_document = Document(sender="test", datetime="", text="Watch this youtube video")
print(kmeans.predict_cluster(test_document))

# Gia ola ta chats, taytopoiei ta clusters apo kathe document sto chat kai
# ektypwnei ta clusters me ton arithmo tw'n document se kathe cluster
i=0
for chat in corpus.documents_per_chat:
    clusters = {}
    i += 1
    for document in corpus.documents_per_chat[chat]:
        predicted_cluster = kmeans.predict_cluster(document)[0]
        if predicted_cluster in clusters:
            clusters[predicted_cluster] += 1
        else:
            clusters[predicted_cluster] = 1
    print("Clusters of chat ",i,"/",len(corpus.documents_per_chat),"(", chat,"): ", clusters)

```

Script 10. tests.py

Και τέλος τρέχουμε το Script 11. (test_lda.py) για να πάρουμε τα τελικά μας αποτελέσματα (topics) με την βοήθεια του LDA .

```

from data_types.document import Document
from preprocessing import corpusCreator
from preprocessing import text_processing

from preprocessing.LDA import LDATopicExtraction

# Dimiourgia tou corpus
corpus = corpusCreator.CorporusCreator("data")
corpus.create()

corpus_documents = corpus.get()
# Dimiourgei ta topics - 13 topics stin periptosi mas
lda = LDATopicExtraction()
lda.compute(corpus_documents, no_features=1000, no_topics=13)
# Ektypwnei tous top 10 orous ana topic
lda.print_topics(10)
# Dhmiourgei ena wordCloud gia to kathe topic kai ta apothikeyei ston fakelo images/LDA
lda.get_word_cloud()

#test_document = Document(sender="test", datetime="", text="Watch this youtube video")

```

```

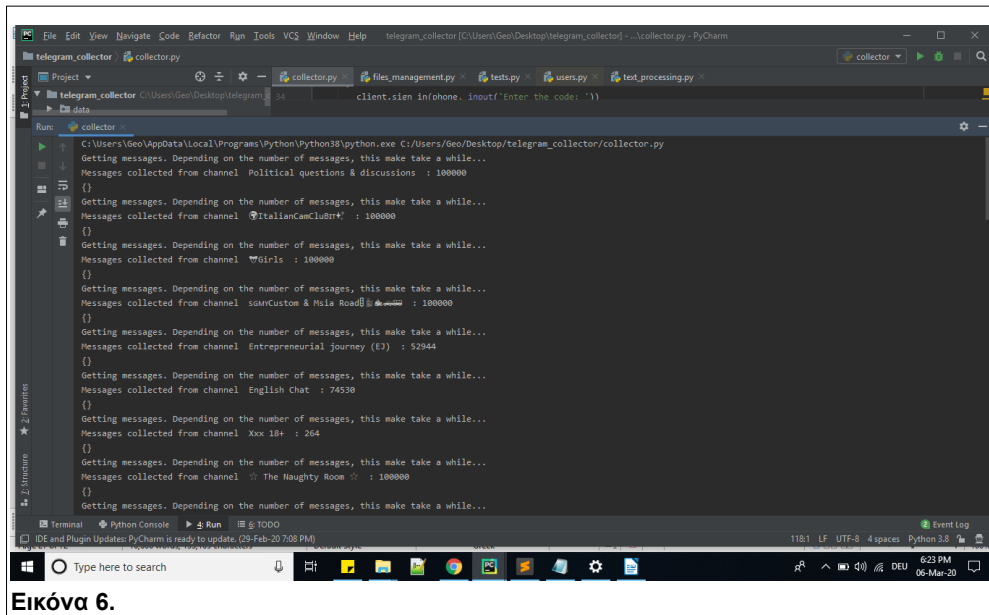
#print(lda.predict_topic(test_document))

i=0
for chat in corpus.documents_per_chat:
    topics = {}
    total = 0
    i += 1
    for document in corpus.documents_per_chat[chat]:
        predicted_topics = lda.predict_topic(document)
        for topic in predicted_topics:
            if topic in topics:
                topics[topic] += predicted_topics[topic]
            else:
                topics[topic] = predicted_topics[topic]
        total += 1
    if total > 0:
        for topic in topics:
            topics[topic] /= total
        print("Chat: ", i, "/", len(corpus.documents_per_chat), "(", chat, "): ")
        print("Topics: ", topics)

```

Script 11. test_lda.py

Αφού τρέξουμε με την βοήθεια του PyCharm το script collector.py για να συλλέξουμε τα δεδομένα (μηνύματα) από το Telegram, παίρνουμε το παρακάτω output (Εικόνα 6.). το οποίο μας λέει ότι "τραβάει" τα δεδομένα από το Telegram και τα αποθηκεύει σε αρχεία στον φάκελο data.



```

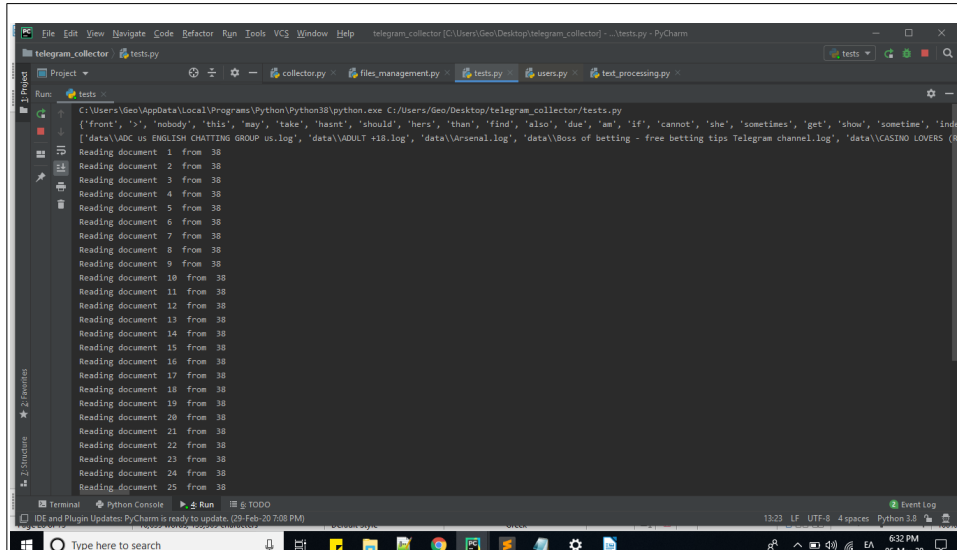
Run: collector
C:\Users\Geo\AppData\Local\Programs\Python\Python38\python.exe C:/Users/Geo/Desktop/telegram_collector/collector.py
Getting messages. Depending on the number of messages, this make take a while...
Messages collected from channel Political questions & discussions : 100000
{}
Getting messages. Depending on the number of messages, this make take a while...
Messages collected from channel ItalianCamCluDr* : 100000
{}
Getting messages. Depending on the number of messages, this make take a while...
Messages collected from channel Girls : 100000
{}
Getting messages. Depending on the number of messages, this make take a while...
Messages collected from channel somCustom & Ns1a Road : 100000
{}
Getting messages. Depending on the number of messages, this make take a while...
Messages collected from channel Entrepreneurial Journey (EJ) : 52944
{}
Getting messages. Depending on the number of messages, this make take a while...
Messages collected from channel English Chat : 74530
{}
Getting messages. Depending on the number of messages, this make take a while...
Messages collected from channel Xxx 18+ : 264
{}
Getting messages. Depending on the number of messages, this make take a while...
Messages collected from channel The Naughty Room : 100000
{}
getting messages. Depending on the number of messages, this make take a while...

```

Εικόνα 6.

5.1 Ανάλυση δεδομένων με τον αλγόριθμο K-means (clustering)

Στην συνέχεια με την βοήθεια του PyCharm τρέχουμε το script tests.py όπου κάνουμε χρήση του αλγορίθμου K-means. Αρχικά παίρνουμε το παρακάτω output, το οποίο μας λέει ότι διαβάζει τα αρχεία όπου είναι αποθηκευμένα τα μηνύματα (Εικόνα 7.).

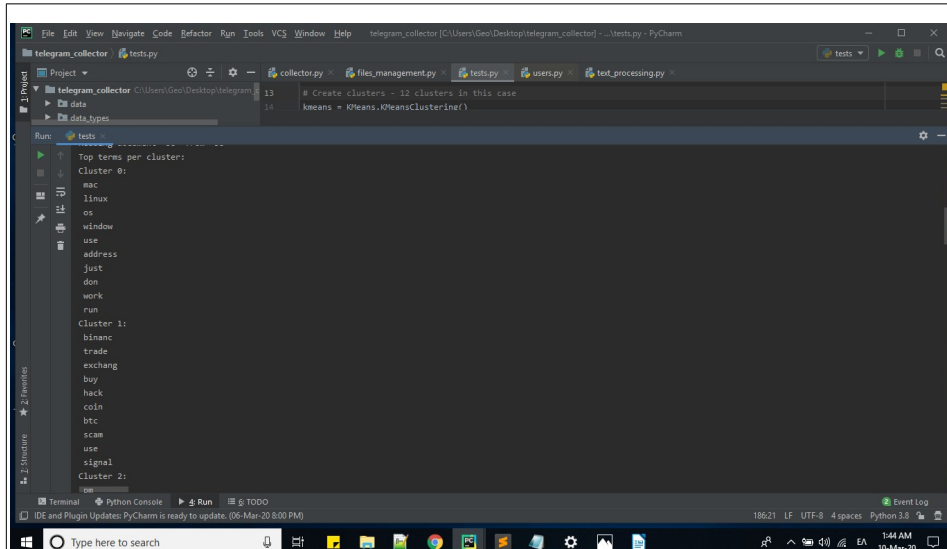


```

C:\Users\Geo\AppData\Local\Programs\Python\Python38\python.exe C:/Users/Geo/Desktop/telegram_collector/tests.py
{'front', '!', 'nobody', 'this', 'may', 'take', 'hasnt', 'should', 'hers', 'than', 'find', 'also', 'due', 'am', 'if', 'cannot', 'she', 'sometimes', 'get', 'show', 'sometime', 'ind
[data\\ADC us ENGLISH CHATTING GROUP us.log', 'data\\ADULT +18.log', 'data\\Arsenal.log', 'data\\Boss of betting - free betting tips Telegram channel.log', 'data\\CASINO LOVERS (
Reading document 1 from 38
Reading document 2 from 38
Reading document 3 from 38
Reading document 4 from 38
Reading document 5 from 38
Reading document 6 from 38
Reading document 7 from 38
Reading document 8 from 38
Reading document 9 from 38
Reading document 10 from 38
Reading document 11 from 38
Reading document 12 from 38
Reading document 13 from 38
Reading document 14 from 38
Reading document 15 from 38
Reading document 16 from 38
Reading document 17 from 38
Reading document 18 from 38
Reading document 19 from 38
Reading document 20 from 38
Reading document 21 from 38
Reading document 22 from 38
Reading document 23 from 38
Reading document 24 from 38
Reading document 25 from 38
  
```

Εικόνα 7.

Στην συνέχεια του output βλέπουμε τα top terms για το κάθε cluster (Εικόνα 8.).



```

# Create clusters - 12 clusters in this case
kmeans = KMeans(n_clusters=12)
# Top terms per cluster:
Cluster 0:
mac
linux
os
window
use
address
just
don
work
run
Cluster 1:
binance
trade
exchange
buy
hack
coin
btc
scam
use
signal
Cluster 2:
ipn
  
```

Εικόνα 8.

Και τέλος το τελευταίο στοιχείο που μας δίνει το output είναι η πληροφορία για το πόσα μηνύματα από κάθε group chat ανήκουν σε καθένα ένα από τα clusters (Εικόνα 9.).

```

# Create clusters - 12 clusters in this case
kmeans = KMeans(KMeansClusterIn())

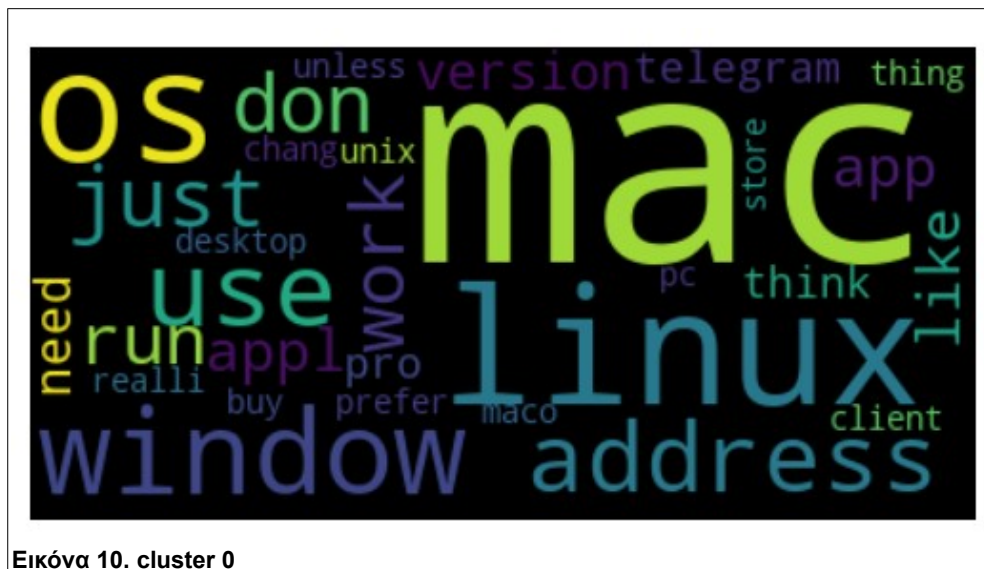
Cluster 0:
plottingCluster 1:
plottingCluster 2:
plottingCluster 3:
plottingCluster 4:
plottingCluster 5:
plottingCluster 6:
plottingCluster 7:
plottingCluster 8:
plottingCluster 9:
plottingCluster 10:
plottingCluster 11:
plotting[7]

Clusters of chat 1 / 38 ( data\ADC us ENGLISH CHATTING GROUP us.log ): {7: 97600, 11: 751, 10: 274, 5: 592, 8: 195, 2: 192, 6: 305, 9: 17}
Clusters of chat 2 / 38 ( data\ADULT +18.log ): {7: 17}
Clusters of chat 3 / 38 ( data\Arsenal.log ): {7: 14169, 10: 11, 6: 7, 2: 1, 9: 1, 8: 10}
Clusters of chat 4 / 38 ( data\Boss of betting - free betting tips Telegram channel.log ): {7: 226}
Clusters of chat 5 / 38 ( data\CASINO LOVERS (ROULETTE - Funtarget)Cricket Betting ids available.log ): {7: 1336, 2: 11, 5: 13, 10: 2, 8: 1}
Clusters of chat 6 / 38 ( data\Chelsea.log ): {7: 3332, 10: 4, 11: 1, 6: 1}
Clusters of chat 7 / 38 ( data\Crypto Groups.log ): {7: 17}
Clusters of chat 8 / 38 ( data\Dating Corner.log ): {7: 12139, 11: 177, 5: 35, 2: 16, 9: 7, 10: 6, 6: 1}
Clusters of chat 9 / 38 ( data\Dating corner.log ): {7: 768, 11: 37, 5: 17, 9: 1, 10: 1, 2: 3}
Clusters of chat 10 / 38 ( data\English Chat.log ): {7: 70393, 10: 349, 5: 244, 2: 141, 11: 433, 6: 349, 9: 16, 8: 186}
Clusters of chat 11 / 38 ( data\English Conversation.log ): {7: 97422, 8: 375, 5: 304, 6: 425, 10: 455, 11: 586, 2: 92, 9: 70, 4: 3}
Clusters of chat 12 / 38 ( data\Entrepreneurial Journey (E3) log ): {7: 31266, 6: 104, 10: 167, 2: 218, 5: 161, 8: 36, 11: 30, 0: 26, 4: 17, 9: 5, 1: 3}

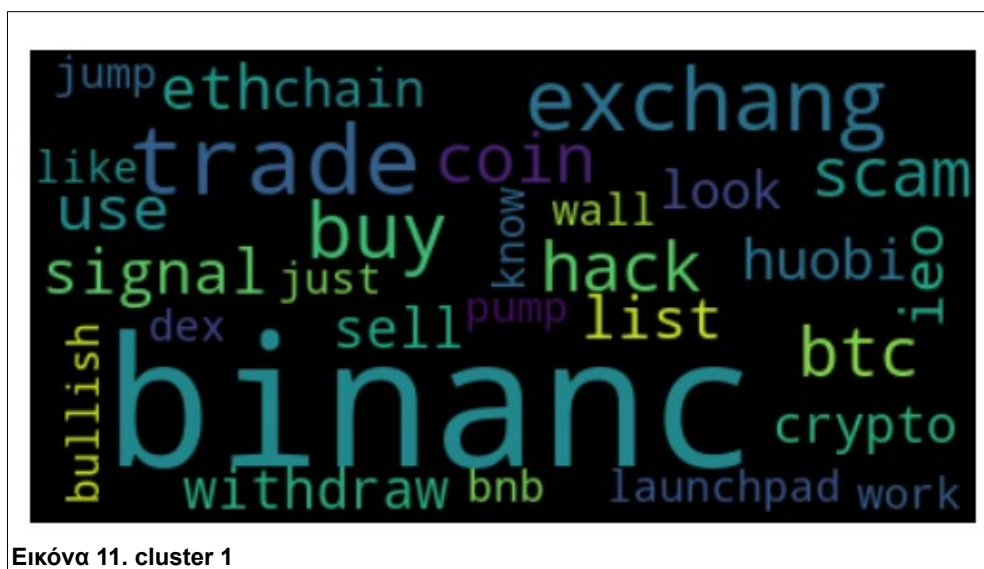
```

Εικόνα 9.

Για παράδειγμα το output “ **Clusters of chat 1 / 38 (data\ADC us ENGLISH CHATTING GROUP us.log): {7: 97600, 11: 751, 10: 274, 5: 592, 8: 195, 2: 192, 6: 305, 9: 17}** “ μας λέει ότι από το συγκεκριμένο group με όνομα “ ADC us ENGLISH CHATTING GROUP us ” 97600 μηνύματα ανήκουν στο cluster 7, 751 μηνύματα ανήκουν στο cluster 11, 274 μηνύματα ανήκουν στο cluster 10, 592 μηνύματα ανήκουν στο cluster 5, 195 μηνύματα ανήκουν στο cluster 8, 192 μηνύματα ανήκουν στο cluster 2, 305 μηνύματα ανήκουν στο cluster 6 και τέλος 17 μηνύματα ανήκουν στο cluster 9. Τα συνολικά αποτελέσματα τα βλέπουμε στον Πίνακα 1. που παρατίθεται, ενώ στις εικόνες 10 – 21 βλέπουμε τα wordclouds τα οποία είναι αποθηκευμένα στον φάκελο images/Kmeans. Στις εικόνες αυτές βλέπουμε τα top terms για το κάθε cluster και από τα top terms μπορούμε να βγάλουμε ένα συμπέρασμα για το τι είδους μηνύματα περιέχει το κάθε chat.



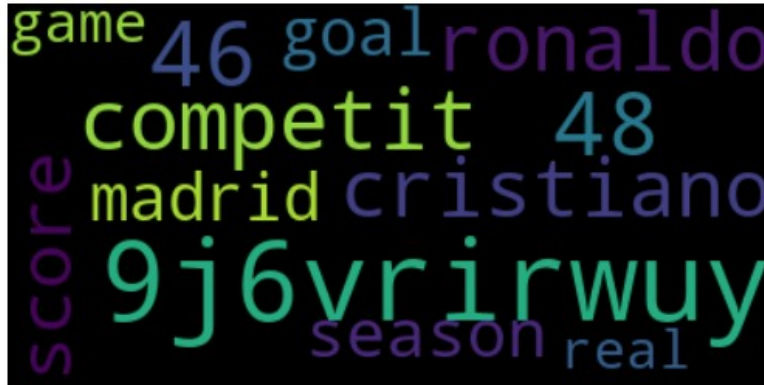
Εικόνα 10. cluster 0



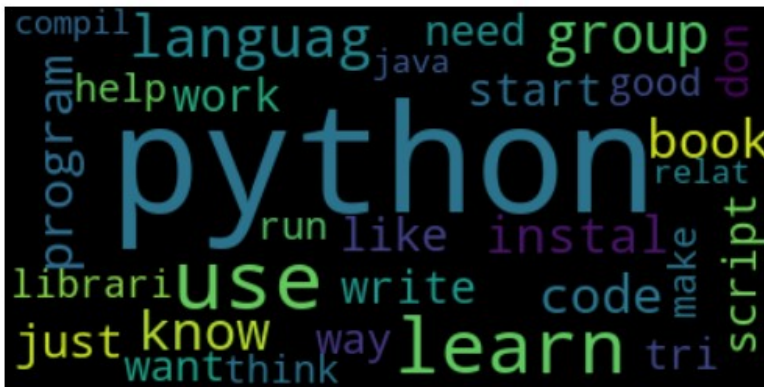
Εικόνα 11. cluster 1



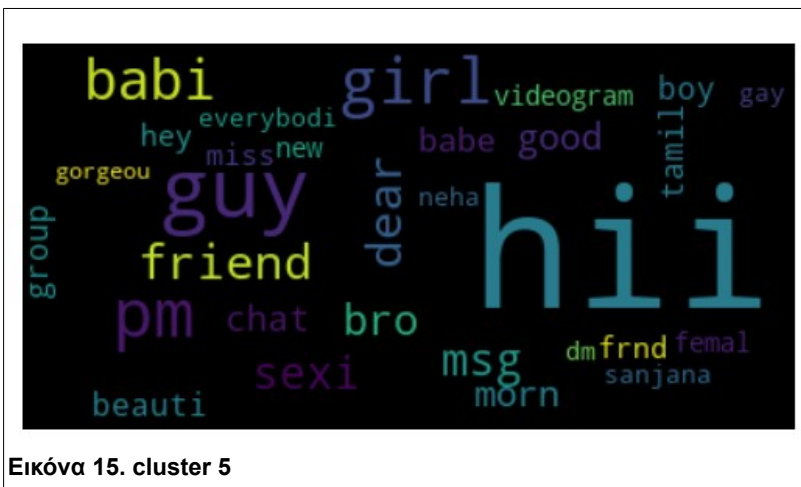
Εικόνα 12. cluster 2



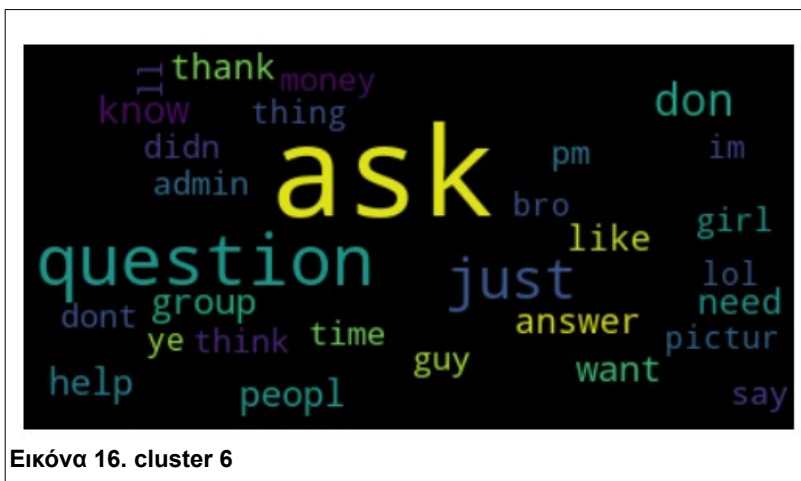
Εικόνα 13. cluster 3



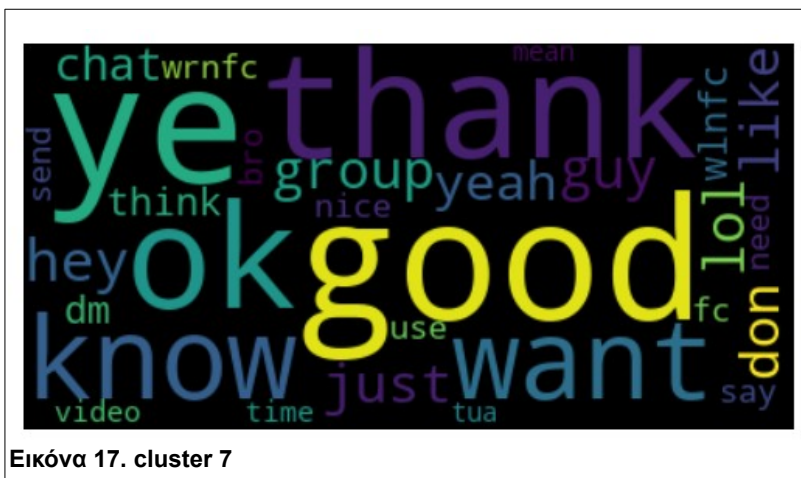
Εικόνα 14. cluster 4



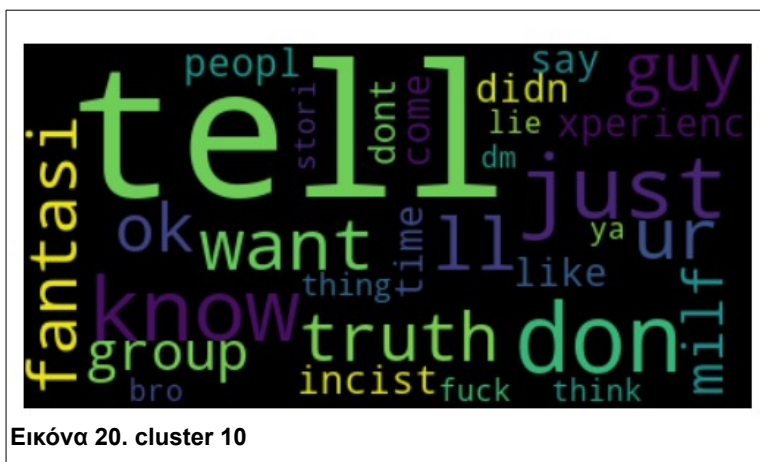
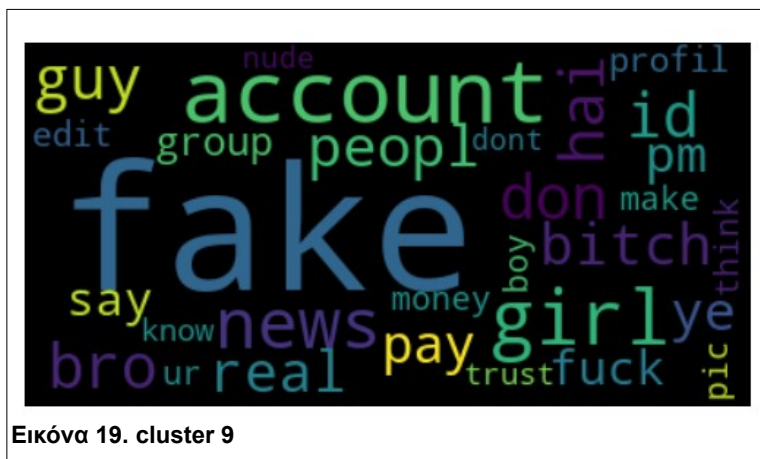
Εικόνα 15. cluster 5

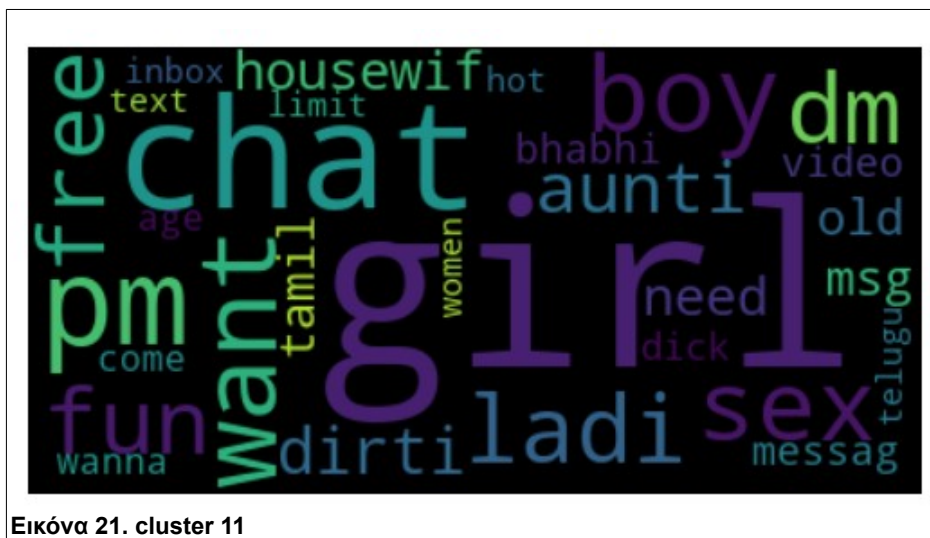


Εικόνα 16. cluster 6



Εικόνα 17. cluster 7





Ακολουθεί ο Πίνακας 1.

Clusters of chat 1 / 38 (data\ADC us ENGLISH CHATTING GROUP us.log): {7: 97600, 11: 751, 10: 274, 5: 592, 8: 195, 2: 192, 6: 305, 9: 17}

Clusters of chat 2 / 38 (data\ADULT +18.log): {7: 17}

Clusters of chat 3 / 38 (data\Arsenal.log): {7: 14169, 10: 11, 6: 7, 2: 1, 9: 1, 8: 10}

Clusters of chat 4 / 38 (data\Boss of betting - free betting tips Telegram channel.log): {7: 226}

Clusters of chat 5 / 38 (data\CASINO LOVERS (ROULLETE - Funtarget)Cricket Betting ids available.log): {7: 1336, 2: 11, 5: 13, 10: 2, 8: 1}

Clusters of chat 6 / 38 (data\Chelsea .log): {7: 3332, 10: 4, 11: 1, 6: 1}

Clusters of chat 7 / 38 (data\Crypto Groups.log): {7: 17}

Clusters of chat 8 / 38 (data\Dating Corner.log): {7: 12139, 11: 177, 5: 35, 2: 16, 9: 7, 10: 6, 6: 1}

Clusters of chat 9 / 38 (data\Dating corner .log): {7: 768, 11: 37, 5: 17, 9: 1, 10: 1, 2: 3}

Clusters of chat 10 / 38 (data\English Chat.log): {7: 70393, 10: 349, 5: 244, 2: 141, 11: 433, 6: 349, 9: 16, 8: 186}

Clusters of chat 11 / 38 (data\English Conversation.log): {7: 97422, 8: 375, 5: 304, 6: 425, 10: 455, 11: 586, 2: 92, 9: 70, 4: 3}

Clusters of chat 12 / 38 (data\Entrepreneurial journey (EJ).log): {7: 31296, 6: 104, 10: 167, 2: 218, 5: 161, 8: 36, 11: 30, 9: 26, 4: 17, 0: 5, 1: 3}

Clusters of chat 13 / 38 (data\EthTrader (under renovation).log): {7: 73272, 6: 252, 2: 65, 9: 80, 1: 245, 8: 140, 10: 344, 5: 11, 11: 64, 4: 10, 0: 4}

Clusters of chat 14 / 38 (data\FC Barcelona Channel.log): {7: 1679}






Clusters of chat 15 / 38 (data\Football Channel .log): {7: 409}

Clusters of chat 16 / 38 (data\Football Daily News.log): {7: 91830, 10: 30, 8: 6, 6: 18, 2: 1, 9: 1, 3: 1}

Clusters of chat 17 / 38 (data\FREE ADULT DATING SERVICE.log): {11: 5, 7: 28}

Clusters of chat 18 / 38 (data\Furry chat 16+.log): {7: 25897, 6: 38, 8: 36, 5: 27, 10: 68, 11: 33, 0: 3, 2: 50, 9: 1}

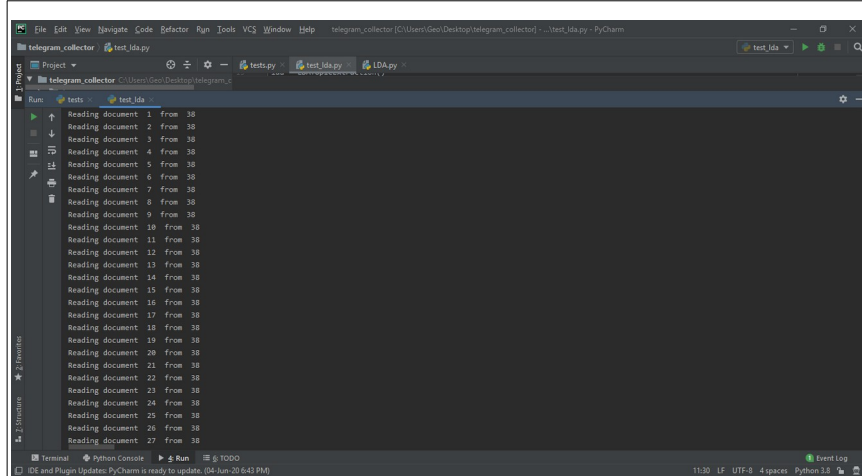
Clusters of chat 19 / 38 (data\Geeks Chat.log): {7: 97490, 6: 885, 2: 165, 8: 166, 10: 517, 9: 65, 4: 347, 5: 123, 0: 140, 11: 85, 1: 10}

Clusters of chat 20 / 38 (data\Hack+.log): {7: 2630, 4: 3}
 Clusters of chat 21 / 38 (data\ICO Speaks.log): {7: 85148, 5: 93, 2: 940, 10: 406, 9: 146, 8: 229, 6: 549, 11: 96, 1: 425, 4: 7}
 Clusters of chat 22 / 38 (data\IT Security Talk.log): {7: 44851, 4: 40, 6: 303, 10: 179, 5: 27, 11: 11, 2: 30, 8: 119, 0: 76, 9: 17}
 Clusters of chat 23 / 38 (data\Motorist Community.log): {7: 13012, 2: 63, 6: 98, 8: 14, 10: 25, 9: 4, 0: 1}
 Clusters of chat 24 / 38 (data\Mountain Hawk Predictions Tennis Football.log): {7: 341}
 Clusters of chat 25 / 38 (data\Political questions & discussions.log): {7: 96990, 10: 1160, 6: 715, 2: 116, 8: 155, 11: 196, 5: 212, 9: 239, 1: 1, 4: 1, 0: 1}
 Clusters of chat 26 / 38 (data\Python.log): {7: 84055, 6: 787, 10: 456, 2: 124, 4: 5052, 8: 121, 0: 62, 5: 43, 9: 26, 11: 46, 1: 3}
 Clusters of chat 27 / 38 (data\Real Madrid C.F..log): {7: 15093, 10: 16, 6: 21, 11: 1}
 Clusters of chat 28 / 38 (data\SteadyLah Network ™.log): {7: 3239}
 Clusters of chat 29 / 38 (data\Telegram movies.log): {7: 1824, 0: 1}
 Clusters of chat 30 / 38 (data\Text Share.log): {}
 Clusters of chat 31 / 38 (data\Work At Your Place.log): {7: 817, 10: 5, 8: 2, 5: 1, 11: 1, 2: 4}
 Clusters of chat 32 / 38 (data\Xxx 18+.log): {7: 263}
 Clusters of chat 33 / 38 (data\☆ The Naughty Room ☆.log): {2: 7294, 7: 75445, 10: 153, 5: 2030, 11: 13589, 9: 611, 8: 94, 6: 119}
 Clusters of chat 34 / 38 (data\H A N G O U T.log): {7: 92749, 11: 3120, 2: 1592, 6: 453, 8: 212, 5: 1022, 10: 386, 9: 184, 4: 5, 0: 1}
 Clusters of chat 35 / 38 (data\sgmyCustom & Msia Road   .log): {7: 83166, 2: 57, 8: 86, 6: 292, 10: 76, 9: 17, 11: 20, 5: 25, 0: 4}
 Clusters of chat 36 / 38 (data\🌐ItalianCamCluBir .log): {7: 68656, 11: 21645, 2: 7193, 9: 1079, 8: 47, 5: 1016, 10: 131, 6: 202, 4: 1}
 Clusters of chat 37 / 38 (data\👩‍🎤ACTRESS ASS MUSEUM™👩‍🎤.log): {7: 80522, 11: 5575, 9: 683, 2: 7229, 6: 93, 5: 146, 10: 122, 8: 377, 4: 1}
 Clusters of chat 38 / 38 (data\👩‍🎤Girls.log): {7: 66730, 5: 649, 9: 483, 2: 11899, 11: 19867, 8: 22, 10: 281, 6: 66}

Πίνακας 1. Τελικά αποτελέσματα Telegram με K-means

5.2 Ανάλυση δεδομένων με LDA (Topic Extraction)

Αφού αναλύσαμε τα δεδομένα που συλλέξαμε με την βοήθεια του αλγορίθμου K-means και πήραμε τα παραπάνω αποτελέσματα του clustering, στο PyCharm τρέχουμε το script `test_lda.py` για να αναλύσουμε τα δεδομένα μας αυτή τη φορά με την βοήθεια του LDA κάνοντας Topic Extraction. Τρέχοντας το script `test_lda.py` παίρνουμε αρχικά το παρακάτω ούτρυτ, το οποίο μας λέει ότι διαβάζει τα αρχεία όπου είναι αποθηκευμένα τα μηνύματα (Εικόνα 22.).



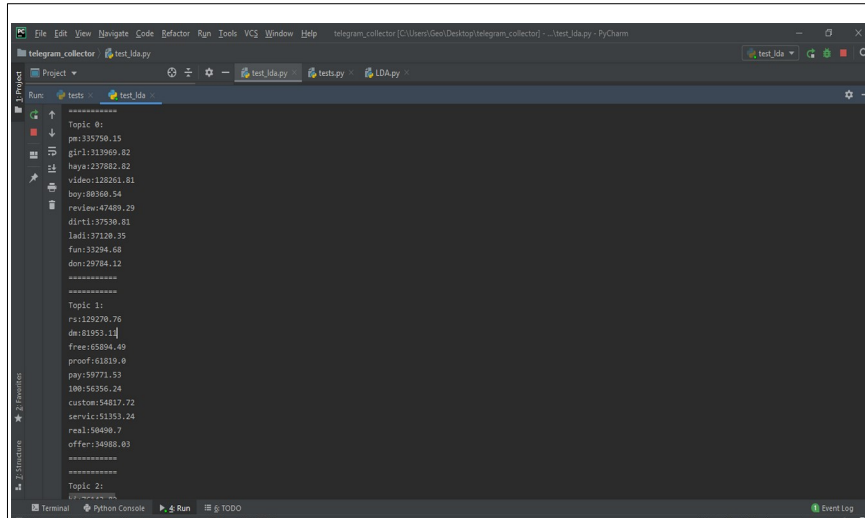
```

Reading document 1 from telegram_1.txt
Reading document 2 from telegram_2.txt
Reading document 3 from telegram_3.txt
Reading document 4 from telegram_4.txt
Reading document 5 from telegram_5.txt
Reading document 6 from telegram_6.txt
Reading document 7 from telegram_7.txt
Reading document 8 from telegram_8.txt
Reading document 9 from telegram_9.txt
Reading document 10 from telegram_10.txt
Reading document 11 from telegram_11.txt
Reading document 12 from telegram_12.txt
Reading document 13 from telegram_13.txt
Reading document 14 from telegram_14.txt
Reading document 15 from telegram_15.txt
Reading document 16 from telegram_16.txt
Reading document 17 from telegram_17.txt
Reading document 18 from telegram_18.txt
Reading document 19 from telegram_19.txt
Reading document 20 from telegram_20.txt
Reading document 21 from telegram_21.txt
Reading document 22 from telegram_22.txt
Reading document 23 from telegram_23.txt
Reading document 24 from telegram_24.txt
Reading document 25 from telegram_25.txt
Reading document 26 from telegram_26.txt
Reading document 27 from telegram_27.txt

```

Εικόνα 22.

Στην συνέχεια του ούτρυτ βλέπουμε τα top terms για το κάθε topic (Εικόνα 23.).



```

=====
Topic 0:
pe:335750.15
girl:313969.82
haya:237882.82
video:118261.81
boy:88360.54
review:47489.29
dir:137930.81
lad:137129.35
fun:33294.68
don:29784.12
=====
Topic 1:
rs:129270.76
de:81953.14
free:65894.49
proof:61819.0
pay:59771.53
100:56350.24
cost:54817.72
servic:51353.24
real:50480.7
offer:34888.83
=====
Topic 2:

```

Εικόνα 23.

Και τέλος το τελευταίο στοιχείο που μας δίνει το ούτρυτ είναι για το κάθε group chat το ποσοστό του κάθε topic όπου κατατάσσεται το group αυτό (Εικόνα 24.).

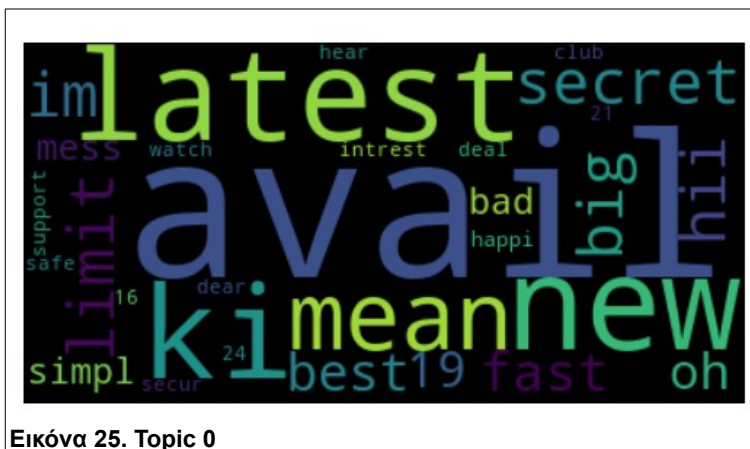
```

Topics: {'topic_0': 0.1616013635078895, 'topic_1': 0.1560079641894828, 'topic_2': 0.17028117658829878, 'topic_3': 0.1550075162418596, 'topic_4': 0.17755261687649432, 'topic_5': 0.17
Chat: data\ADULT +18.log
Topics: {'topic_0': 0.17146703490660994, 'topic_1': 0.3980314699092582, 'topic_2': 0.10227978951551436, 'topic_3': 0.09901190055335145, 'topic_4': 0.08267196717033862, 'topic_5': 0.
Chat: data\Arsenal.log
Topics: {'topic_0': 0.12489642084188689, 'topic_1': 0.33812081490938145, 'topic_2': 0.10627189766480857, 'topic_3': 0.18013317841209073, 'topic_4': 0.10625445876130393, 'topic_5': 0
Chat: data\Boss of betting - free betting tips Telegram channel.log
Topics: {'topic_0': 0.12663159570429847, 'topic_1': 0.3666217545220687, 'topic_2': 0.08403553585469947, 'topic_3': 0.2198745002846462, 'topic_4': 0.07037847907802942, 'topic_5': 0.1
Chat: data\CASINO LOVERS (ROULETTE - Funtarget)Cricket Betting ids available.log
Topics: {'topic_0': 0.18198874321264628, 'topic_1': 0.16682919092696347, 'topic_2': 0.14694618714851648, 'topic_3': 0.21455171499067457, 'topic_4': 0.16904830354772847, 'topic_5': 0
Chat: data\Chelsea.log
Topics: {'topic_0': 0.07199707748841568, 'topic_1': 0.3916006286303685, 'topic_2': 0.052202631465871976, 'topic_3': 0.35824399623031793, 'topic_4': 0.04595578638910563, 'topic_5': 0
Chat: data\Crypto Groups.log
Topics: {'topic_0': 0.02890845018268385, 'topic_1': 0.38309646242231354, 'topic_2': 0.1058710048084064, 'topic_3': 0.35085703221516007, 'topic_4': 0.02329347734071698, 'topic_5': 0.
Chat: data\Dating Corner.log
Topics: {'topic_0': 0.07135014688788859, 'topic_1': 0.31597023697927024, 'topic_2': 0.1010218436659091, 'topic_3': 0.07440255841308199, 'topic_4': 0.083659082430579375, 'topic_5': 0.
Chat: data\English Chat.log
Topics: {'topic_0': 0.17026690294665633, 'topic_1': 0.1712368532446749, 'topic_2': 0.1551973364342517, 'topic_3': 0.16522591569951253, 'topic_4': 0.17101660406984803, 'topic_5': 0.1
Chat: data\English Conversation.log
Topics: {'topic_0': 0.16677674310500792, 'topic_1': 0.1805629051296366, 'topic_2': 0.15923543888359679, 'topic_3': 0.1707291780024529, 'topic_4': 0.161126587344863, 'topic_5': 0.161
Chat: data\Entrepreneurial journey (E).log
Topics: {'topic_0': 0.1444555209433865, 'topic_1': 0.24342908082983303, 'topic_2': 0.1322242402186143, 'topic_3': 0.20967961518071034, 'topic_4': 0.1359248540050272, 'topic_5': 0.13
Chat: data\EthTrader (under renovation).log
Topics: {'topic_0': 0.15803264216169755, 'topic_1': 0.26986345907131276, 'topic_2': 0.11044804748983102, 'topic_3': 0.1743830566979008, 'topic_4': 0.1375209008197611, 'topic_5': 0.1
Chat: data\FC Barcelona Channel.log
Topics: {'topic_0': 0.11728150913040546, 'topic_1': 0.3747861518462875, 'topic_2': 0.1348836563718045, 'topic_3': 0.14472612727062434, 'topic_4': 0.10187925369808938, 'topic_5': 0.1
Chat: data\Football Channel.log

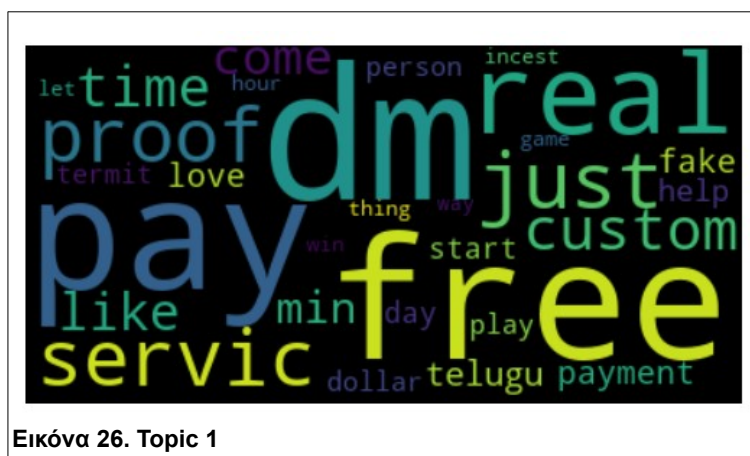
```

Εικόνα 24.

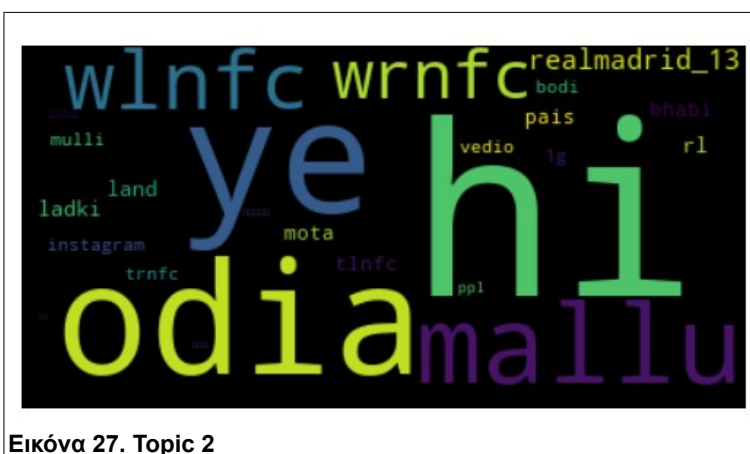
Για παράδειγμα το output “Chat: 1 / 38 (data\ADC us ENGLISH CHATTING GROUP us.log): Topics: {‘topic_0’: 0.07098561242912704, ‘topic_1’: 0.07108862425843088, ‘topic_2’: 0.09914134298343501, ‘topic_3’: 0.0747991509916663, ‘topic_4’: 0.069744492122947679, ‘topic_5’: 0.100046727103669, ‘topic_6’: 0.08624667786579822, ‘topic_7’: 0.06762214830270062, ‘topic_8’: 0.06840506807688701, ‘topic_9’: 0.06980454076510392, ‘topic_10’: 0.07417241617497863, ‘topic_11’: 0.07332614664081709, ‘topic_12’: 0.07461662317747526}“ μας λέει ότι στο συγκεκριμένο chat με όνομα “ ADC us ENGLISH CHATTING GROUP us ” το θέμα συζήτησης των μελών του αντιστοιχεί σε ποσοστό 7% στο Topic_0, 7.1% στο Topic_1, 9.91% στο Topic_2, 7.47% στο Topic_3, 6.97% στο Topic_4, 10% στο Topic_5, 8.62% στο Topic_6, 6.76% στο Topic_7, 6.84% στο Topic_8, 6.98% στο Topic_9, 7.41% στο Topic_10, 7.33% στο Topic_11 και 7.46% στο Topic_12. Τα συνολικά αποτελέσματα τα βλέπουμε στον Πίνακα 2. που παρατίθεται, ενώ στις εικόνες 25 – 37 βλέπουμε τα wordclouds τα οποία είναι αποθηκευμένα στον φάκελο images/LDA. Στις εικόνες αυτές βλέπουμε τα top terms για το κάθε Topic και από τα top terms μπορούμε να βγάλουμε ένα συμπέρασμα για το τι είδους μηνύματα περιέχει το κάθε chat.



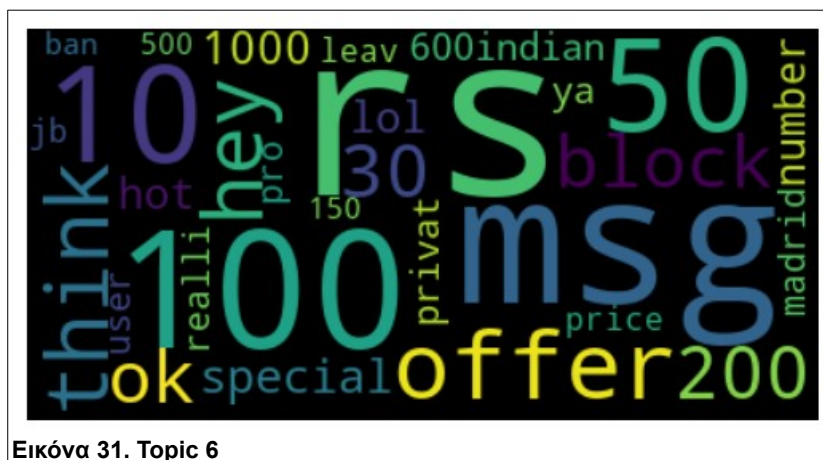
Εικόνα 25. Topic 0



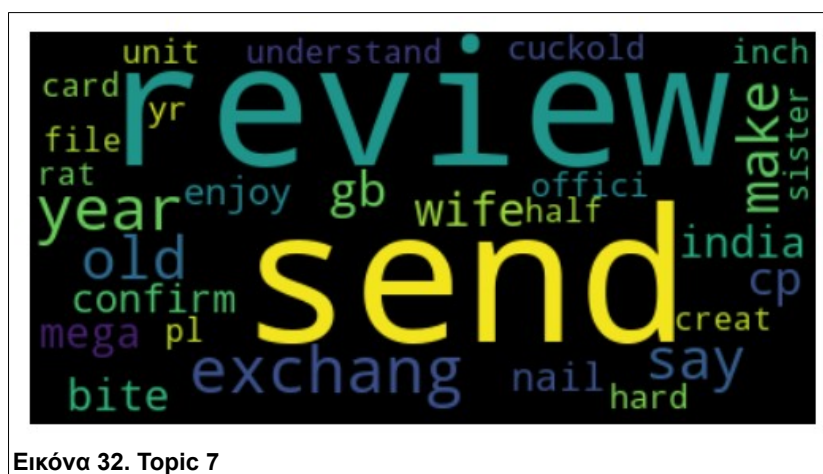
Εικόνα 26. Topic 1



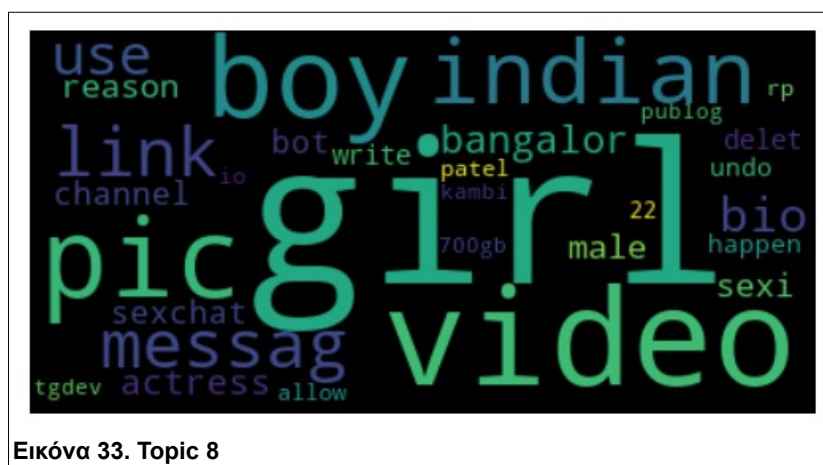
Εικόνα 27. Topic 2



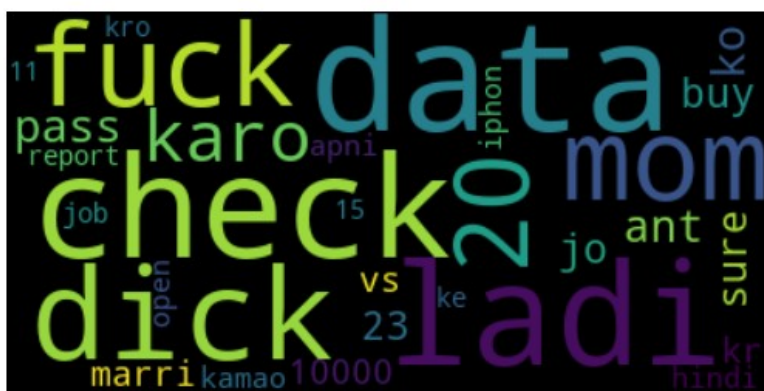
Εικόνα 31. Topic 6



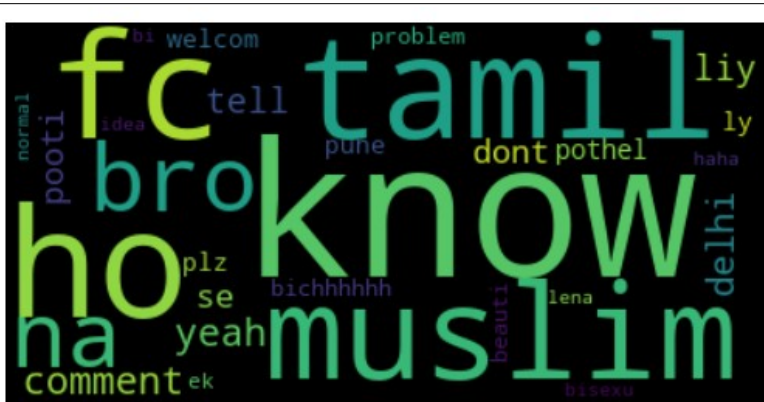
Εικόνα 32. Topic 7



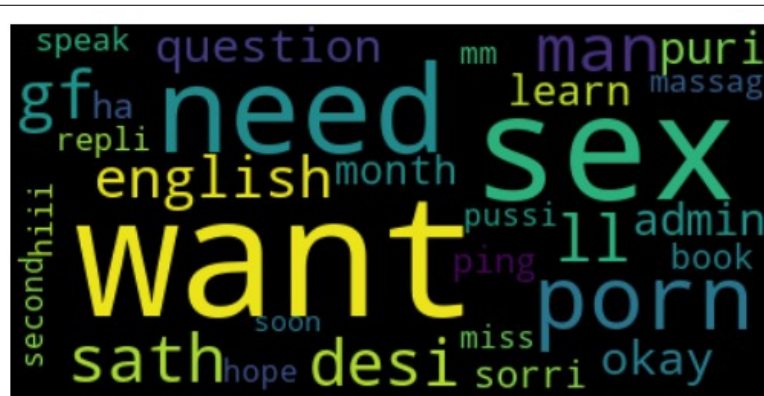
Εικόνα 33. Topic 8



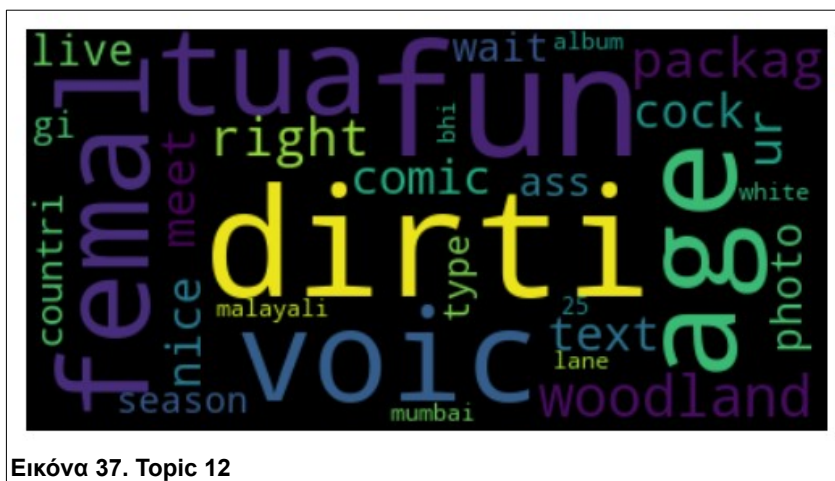
Εικόνα 34. Topic 9



Εικόνα 35. Topic 10



Εικόνα 36. Topic 11



Ακολουθεί ο Πίνακας 2.

```

Chat: 1 / 38 ( data\ADC us ENGLISH CHATTING GROUP us.log ):
Topics: {'topic_0': 0.07098561242912704, 'topic_1': 0.07108862425843088, 'topic_2':
0.09914134298343501, 'topic_3': 0.0747991509916663, 'topic_4': 0.06974492122947679, 'topic_5':
0.100046727103669, 'topic_6': 0.08624667786579822, 'topic_7': 0.06762214830270062, 'topic_8':
0.06840506807688701, 'topic_9': 0.06980454076510392, 'topic_10': 0.07417241617497863, 'topic_11':
0.07332614664081709, 'topic_12': 0.07461662317747526}
Chat: 2 / 38 ( data\ADULT +18.log ):
Topics: {'topic_0': 0.04987789990102501, 'topic_1': 0.07480787177000138, 'topic_2':
0.033631401278461606, 'topic_3': 0.3312037636339405, 'topic_4': 0.08657257775015954, 'topic_5':
0.043435322839185074, 'topic_6': 0.04343532283158836, 'topic_7': 0.03755296990776024, 'topic_8':
0.066964734607198, 'topic_9': 0.09660992595323567, 'topic_10': 0.04343532284367145, 'topic_11':
0.05884148540531179, 'topic_12': 0.03363140127846134}
Chat: 3 / 38 ( data\Arsenal.log ):
Topics: {'topic_0': 0.0814843549413025, 'topic_1': 0.23871492896207375, 'topic_2':
0.04312638563494991, 'topic_3': 0.06697895400776557, 'topic_4': 0.05809112897357207, 'topic_5':
0.05850021161848666, 'topic_6': 0.06874422216127163, 'topic_7': 0.07738190792266382, 'topic_8':
0.050534888800992064, 'topic_9': 0.07959757240877215, 'topic_10': 0.05311900327386256, 'topic_11':
0.06057218353517985, 'topic_12': 0.06315425775910606}
Chat: 4 / 38 ( data\Boss of betting - free betting tips Telegram channel.log ):
Topics: {'topic_0': 0.10186872555493239, 'topic_1': 0.2024541746680143, 'topic_2':
0.03826958782883951, 'topic_3': 0.1253809079943482, 'topic_4': 0.04278741286497573, 'topic_5':
0.043315022636149746, 'topic_6': 0.0864507889406551, 'topic_7': 0.08667237029566253, 'topic_8':
0.039924190928220096, 'topic_9': 0.0748936929317911, 'topic_10': 0.03837718111957357, 'topic_11':
0.0767128909790453, 'topic_12': 0.04289305325779232}
Chat: 5 / 38 ( data\CASINO LOVERS (ROULLETE - Funtarget)Cricket Betting ids available.log ):
Topics: {'topic_0': 0.08763554230277705, 'topic_1': 0.07863068748779041, 'topic_2':
0.061001326134048954, 'topic_3': 0.10645625161501836, 'topic_4': 0.07829317431464003, 'topic_5':
0.05583136194064564, 'topic_6': 0.12546801235032426, 'topic_7': 0.06252274155185723, 'topic_8':

```

```

0.05561472811333214, 'topic_9': 0.09489576537251727, 'topic_10': 0.07038252332460729, 'topic_11':
0.06239161401620147, 'topic_12': 0.060876271476246184}
Chat: 6 / 38 ( data\Chelsea (C).log ):
Topics: {'topic_0': 0.07631136974413831, 'topic_1': 0.2026270441433415, 'topic_2':
0.009981664964741371, 'topic_3': 0.23740173846922522, 'topic_4': 0.13418178390688715, 'topic_5':
0.025840415479211388, 'topic_6': 0.061881504632588236, 'topic_7': 0.06949152083779218, 'topic_8':
0.023567543726598687, 'topic_9': 0.06413053820764467, 'topic_10': 0.02095553882441893, 'topic_11':
0.03554938205493903, 'topic_12': 0.038079955008473734}
Chat: 7 / 38 ( data\Crypto Groups.log ):
Topics: {'topic_0': 0.15529496019999317, 'topic_1': 0.10051657879641342, 'topic_2':
0.01075083543938384, 'topic_3': 0.3801176356466371, 'topic_4': 0.03436729348926933, 'topic_5':
0.010750835439383316, 'topic_6': 0.010750835439383297, 'topic_7': 0.05435247607204896, 'topic_8':
0.08850634824001738, 'topic_9': 0.07510882233427993, 'topic_10': 0.025959408995358372, 'topic_11':
0.04277313446844846, 'topic_12': 0.010750835439383497}
Chat: 8 / 38 ( data\Dating Corner.log ):
Topics: {'topic_0': 0.035654549708569024, 'topic_1': 0.037511483850605806, 'topic_2':
0.06644455168987205, 'topic_3': 0.29420026667477295, 'topic_4': 0.03227268982076992, 'topic_5':
0.031240011242745235, 'topic_6': 0.037060894453657074, 'topic_7': 0.029618609758109114, 'topic_8':
0.32000572880425593, 'topic_9': 0.029935272727602363, 'topic_10': 0.028110721612046113, 'topic_11':
0.030397762942805957, 'topic_12': 0.02754745671418658}
Chat: 9 / 38 ( data\Dating corner♥.log ):
Topics: {'topic_0': 0.04963658584154082, 'topic_1': 0.04975464184094847, 'topic_2':
0.07291372140712678, 'topic_3': 0.11157286346085074, 'topic_4': 0.04334104743476219, 'topic_5':
0.05788025380182947, 'topic_6': 0.1134887361819596, 'topic_7': 0.041550294863004524, 'topic_8':
0.0638225908537663, 'topic_9': 0.04250865689422659, 'topic_10': 0.16219164472362074, 'topic_11':
0.056383037027172137, 'topic_12': 0.13495592566919518}
Chat: 10 / 38 ( data\English Chat.log ):
Topics: {'topic_0': 0.0706721175234561, 'topic_1': 0.08217337407839612, 'topic_2':
0.08334292750106458, 'topic_3': 0.07626675441028873, 'topic_4': 0.07268119858039099, 'topic_5':
0.0863691829341635, 'topic_6': 0.09655989421485466, 'topic_7': 0.06838263439324237, 'topic_8':
0.06620290842233446, 'topic_9': 0.07028050163913768, 'topic_10': 0.07641800115448652, 'topic_11':
0.07672560254120783, 'topic_12': 0.07392490260633401}
Chat: 11 / 38 ( data\English Conversation.log ):
Topics: {'topic_0': 0.07656891346193252, 'topic_1': 0.09224773853251621, 'topic_2':
0.08225173358892891, 'topic_3': 0.07621685884940052, 'topic_4': 0.07181489768463666, 'topic_5':
0.07950332415080702, 'topic_6': 0.08807212922584212, 'topic_7': 0.07152265014848146, 'topic_8':
0.061376609864059684, 'topic_9': 0.07592588739488547, 'topic_10': 0.0763485010598971, 'topic_11':
0.07182988311055019, 'topic_12': 0.07632087292746068}
Chat: 12 / 38 ( data\Entrepreneurial journey (EJ).log ):
Topics: {'topic_0': 0.06762609521601173, 'topic_1': 0.1501993694488715, 'topic_2':
0.07014486429542464, 'topic_3': 0.10396742915670341, 'topic_4': 0.06471416147695784, 'topic_5':
0.07647946258558996, 'topic_6': 0.08277125398726104, 'topic_7': 0.06839859789074719, 'topic_8':
0.05421801193440344, 'topic_9': 0.08025262746041273, 'topic_10': 0.05910327854063491, 'topic_11':
0.0682739771730928, 'topic_12': 0.053850870833720116}
Chat: 13 / 38 ( data\EthTrader (under renovation).log ):
Topics: {'topic_0': 0.09843396861737931, 'topic_1': 0.13516853993116154, 'topic_2':
0.05071266314700416, 'topic_3': 0.07913073506393828, 'topic_4': 0.0644897142217084, 'topic_5':
0.06677200441812697, 'topic_6': 0.10166118698726197, 'topic_7': 0.07370079759176028, 'topic_8':

```

```
0.06086925152773749, 'topic_9': 0.07807731197721418, 'topic_10': 0.0630873681847241, 'topic_11':
0.06600696914297265, 'topic_12': 0.0618894891884154}
Chat: 14 / 38 ( data\FC Barcelona Channel.log ):
Topics: {'topic_0': 0.08073973392806465, 'topic_1': 0.17974881786984126, 'topic_2':
0.0437186648213184, 'topic_3': 0.05544026050701292, 'topic_4': 0.05072871683764152, 'topic_5':
0.047742451898734, 'topic_6': 0.05762913329179291, 'topic_7': 0.0685348499956157, 'topic_8':
0.07852159927990429, 'topic_9': 0.1684124508621611, 'topic_10': 0.05961096566843697, 'topic_11':
0.052753356556027246, 'topic_12': 0.05641899848345619}
Chat: 15 / 38 ( data\Football Channel 🏆.log ):
Topics: {'topic_0': 0.09425887299891487, 'topic_1': 0.23394260465639996, 'topic_2':
0.036940889436279815, 'topic_3': 0.057003764994984836, 'topic_4': 0.054762135012550554, 'topic_5':
0.039857143855178245, 'topic_6': 0.06975095832310962, 'topic_7': 0.06599632215505338, 'topic_8':
0.05005910939802533, 'topic_9': 0.12952194684008056, 'topic_10': 0.04400418747044912, 'topic_11':
0.06551420652546344, 'topic_12': 0.05838785833351108}
Chat: 16 / 38 ( data\Football Daily News.log ):
Topics: {'topic_0': 0.09308932288759986, 'topic_1': 0.2729913824428006, 'topic_2':
0.017279941625512506, 'topic_3': 0.06378374402002883, 'topic_4': 0.047569210584673775, 'topic_5':
0.043506491799518966, 'topic_6': 0.06418604040336937, 'topic_7': 0.09667578688411478, 'topic_8':
0.03590026755209714, 'topic_9': 0.11346256445657352, 'topic_10': 0.041712192721869594, 'topic_11':
0.055237655621312255, 'topic_12': 0.054605398999997}
Chat: 17 / 38 ( data\FREE ADULT DATING SERVICE.log ):
Topics: {'topic_0': 0.04862637362573425, 'topic_1': 0.17723387721776207, 'topic_2':
0.04711122211372783, 'topic_3': 0.06504051502438825, 'topic_4': 0.04862637362871727, 'topic_5':
0.04711122210946365, 'topic_6': 0.09209679208743533, 'topic_7': 0.041050616050616794, 'topic_8':
0.11035076036143465, 'topic_9': 0.04685869684260333, 'topic_10': 0.041050616050616884, 'topic_11':
0.1542901543326045, 'topic_12': 0.08055278055489506}
Chat: 18 / 38 ( data\Furry chat 16+.log ):
Topics: {'topic_0': 0.0834593693074548, 'topic_1': 0.10876686887760503, 'topic_2': 0.0647088412475006,
'topic_3': 0.07186464379192822, 'topic_4': 0.07678048452696594, 'topic_5': 0.07575458195841293,
'topic_6': 0.0905670126046306, 'topic_7': 0.07118465012237075, 'topic_8': 0.06389313806388143,
'topic_9': 0.07556618829183201, 'topic_10': 0.07028549676872296, 'topic_11': 0.07322810034327867,
'topic_12': 0.07394062409533937}
Chat: 19 / 38 ( data\Geeks Chat.log ):
Topics: {'topic_0': 0.07884906840478824, 'topic_1': 0.12933888358291257, 'topic_2':
0.05207462907106658, 'topic_3': 0.0856402633100968, 'topic_4': 0.07261460986486958, 'topic_5':
0.058728334187497414, 'topic_6': 0.0935196619054667, 'topic_7': 0.07326819643464508, 'topic_8':
0.07239546798018993, 'topic_9': 0.08448143629279428, 'topic_10': 0.06510747588058632, 'topic_11':
0.07150936029932259, 'topic_12': 0.0624726127851618}
Chat: 20 / 38 ( data\Hack+.log ):
Topics: {'topic_0': 0.07899324724421423, 'topic_1': 0.07888143032162734, 'topic_2':
0.07440074306011538, 'topic_3': 0.07785896353434225, 'topic_4': 0.07679993161874721, 'topic_5':
0.07549318092047501, 'topic_6': 0.07721619890115353, 'topic_7': 0.07581507913620562, 'topic_8':
0.07821444159729217, 'topic_9': 0.07877477241626185, 'topic_10': 0.07566770572629815, 'topic_11':
0.07579920217801128, 'topic_12': 0.07608510334522818}
Chat: 21 / 38 ( data\ICO Speaks.log ):
Topics: {'topic_0': 0.08054663164437197, 'topic_1': 0.13258398534567328, 'topic_2':
0.0640637071610087, 'topic_3': 0.0820708223698164, 'topic_4': 0.06489545520078427, 'topic_5':
0.10308377373490668, 'topic_6': 0.0951366662475309, 'topic_7': 0.0634066675244357, 'topic_8':
```

0.06478097853642499, 'topic_9': 0.06626491229254018, 'topic_10': 0.06000359302598842, 'topic_11': 0.0669593007486548, 'topic_12': 0.05620350616737324}

Chat: 22 / 38 (data\IT Security Talk.log):

Topics: {'topic_0': 0.08450391654790627, 'topic_1': 0.1456985478319673, 'topic_2': 0.046275777125092184, 'topic_3': 0.08080806794390596, 'topic_4': 0.06986257418323615, 'topic_5': 0.0565543195711263, 'topic_6': 0.09546754189853528, 'topic_7': 0.074070428000943, 'topic_8': 0.06793283616489804, 'topic_9': 0.08613012454408027, 'topic_10': 0.06636459155013823, 'topic_11': 0.06665663262022739, 'topic_12': 0.05967464201761135}

Chat: 23 / 38 (data\Motorist Community.log):

Topics: {'topic_0': 0.06546961308076943, 'topic_1': 0.12006527750422744, 'topic_2': 0.046591624512254556, 'topic_3': 0.0920843155538202, 'topic_4': 0.06540570394847149, 'topic_5': 0.05924079150494492, 'topic_6': 0.09979170626198142, 'topic_7': 0.07216827580881296, 'topic_8': 0.04815944567291508, 'topic_9': 0.12120224182301632, 'topic_10': 0.05907582096958229, 'topic_11': 0.05577425335523896, 'topic_12': 0.09497093000396041}

Chat: 24 / 38 (data\Mountain Hawk Predictions Tennis Football.log):

Topics: {'topic_0': 0.05537235620882344, 'topic_1': 0.2282609083289777, 'topic_2': 0.045384773666333984, 'topic_3': 0.047902679606296364, 'topic_4': 0.07368928307024113, 'topic_5': 0.044078131717686556, 'topic_6': 0.07357344711715105, 'topic_7': 0.047677694739529417, 'topic_8': 0.04306570330949391, 'topic_9': 0.15270989398642526, 'topic_10': 0.06117582718050315, 'topic_11': 0.07410399352801611, 'topic_12': 0.0530053075405218}

Chat: 25 / 38 (data\Political questions & discussions.log):

Topics: {'topic_0': 0.06829785796942496, 'topic_1': 0.11839163998889204, 'topic_2': 0.06281301674969074, 'topic_3': 0.077197700616943, 'topic_4': 0.07167379368420504, 'topic_5': 0.060724476807704736, 'topic_6': 0.09028515268854649, 'topic_7': 0.09244280957317157, 'topic_8': 0.05776375416371471, 'topic_9': 0.0814922925107941, 'topic_10': 0.0766686577883705, 'topic_11': 0.06853702795850578, 'topic_12': 0.0737118194994813}

Chat: 26 / 38 (data\Python.log):

Topics: {'topic_0': 0.07478371665596055, 'topic_1': 0.12338222223934571, 'topic_2': 0.04919450727392428, 'topic_3': 0.08562738918140989, 'topic_4': 0.06827471131547401, 'topic_5': 0.06708718585896435, 'topic_6': 0.09486243528751546, 'topic_7': 0.07586340552529791, 'topic_8': 0.07030432523720212, 'topic_9': 0.08716436446854829, 'topic_10': 0.06483577675290181, 'topic_11': 0.07508684361606423, 'topic_12': 0.06353311658677366}

Chat: 27 / 38 (data\Real Madrid C.F..log):

Topics: {'topic_0': 0.06594357325589212, 'topic_1': 0.2566888362156821, 'topic_2': 0.10832998995827389, 'topic_3': 0.05426510375333086, 'topic_4': 0.041876210988092424, 'topic_5': 0.038985432850268856, 'topic_6': 0.11185083635826966, 'topic_7': 0.0632580635988364, 'topic_8': 0.05157983384048169, 'topic_9': 0.07730152316107161, 'topic_10': 0.032145498805253074, 'topic_11': 0.04986692680532166, 'topic_12': 0.047908170409212504}

Chat: 28 / 38 (data\SteadyLah Network ™.log):

Topics: {'topic_0': 0.05585767409833146, 'topic_1': 0.23269086098808656, 'topic_2': 0.011297981239834332, 'topic_3': 0.10060356787868116, 'topic_4': 0.12976237056665765, 'topic_5': 0.019220506653041255, 'topic_6': 0.08150971051542263, 'topic_7': 0.05227397250686928, 'topic_8': 0.01720732250806319, 'topic_9': 0.16006406104479767, 'topic_10': 0.016524363827957782, 'topic_11': 0.07768665780839347, 'topic_12': 0.04530095036386617}

Chat: 29 / 38 (data\Telegram movies.log):

Topics: {'topic_0': 0.08997983461553818, 'topic_1': 0.09878750683343275, 'topic_2': 0.04592695344781638, 'topic_3': 0.05269405163664073, 'topic_4': 0.04763302920165911, 'topic_5': 0.047117019593174735, 'topic_6': 0.15487710278841696, 'topic_7': 0.1841589430074558, 'topic_8':

0.04781922429057965, 'topic_9': 0.05593296314703611, 'topic_10': 0.04715482907449306, 'topic_11': 0.06145619234494143, 'topic_12': 0.0664623500188233}
 Chat: 30 / 38 (data\Text Share.log):
 Topics: {}
 Chat: 31 / 38 (data\Work At Your Place.log):
 Topics: {'topic_0': 0.07336703877804696, 'topic_1': 0.14313983068150657, 'topic_2': 0.06520366020872082, 'topic_3': 0.08185509366152277, 'topic_4': 0.07928628277497661, 'topic_5': 0.0789924306261777, 'topic_6': 0.08896287309631844, 'topic_7': 0.0675144900398492, 'topic_8': 0.058168907709998094, 'topic_9': 0.06888232205051514, 'topic_10': 0.06009940882658881, 'topic_11': 0.058779906384315954, 'topic_12': 0.07574775516147318}
 Chat: 32 / 38 (data\Xxx 18+.log):
 Topics: {'topic_0': 0.017630618794571045, 'topic_1': 0.35405627610810836, 'topic_2': 0.004907342957308817, 'topic_3': 0.11886799737912024, 'topic_4': 0.00936885818409803, 'topic_5': 0.007933431408820451, 'topic_6': 0.018626766318708743, 'topic_7': 0.24401706215863203, 'topic_8': 0.012828485105469672, 'topic_9': 0.010562049179639576, 'topic_10': 0.006442234475974271, 'topic_11': 0.18246708378925602, 'topic_12': 0.01229179414029165}
 Chat: 33 / 38 (data\☆ The Naughty Room ☆.log):
 Topics: {'topic_0': 0.06356550348904051, 'topic_1': 0.08855146016158043, 'topic_2': 0.09369168533350428, 'topic_3': 0.0741460110786562, 'topic_4': 0.0982017184361448, 'topic_5': 0.06794248007536417, 'topic_6': 0.07398728169475569, 'topic_7': 0.06365563573196541, 'topic_8': 0.10698506851046548, 'topic_9': 0.0710349697964324, 'topic_10': 0.0631760250399456, 'topic_11': 0.07192611035665088, 'topic_12': 0.0631360502951013}
 Chat: 34 / 38 (data\HANGOUT.log):
 Topics: {'topic_0': 0.07496594769142668, 'topic_1': 0.09154234806526558, 'topic_2': 0.08524382209599812, 'topic_3': 0.07350713968664108, 'topic_4': 0.07827448025779377, 'topic_5': 0.07738385572008323, 'topic_6': 0.09307677052443936, 'topic_7': 0.06503387659674427, 'topic_8': 0.07554378870161796, 'topic_9': 0.07257304937377199, 'topic_10': 0.07448539802699883, 'topic_11': 0.06956687503801519, 'topic_12': 0.06880264822069397}
 Chat: 35 / 38 (data\sgm\Custom & Msia Road 📁 📁 📁 📁 📁.log):
 Topics: {'topic_0': 0.05136663083040195, 'topic_1': 0.08733027679272758, 'topic_2': 0.145181516614415, 'topic_3': 0.07720275301883905, 'topic_4': 0.06003902540260905, 'topic_5': 0.0699431236456086, 'topic_6': 0.08392384613635297, 'topic_7': 0.05146686669621509, 'topic_8': 0.04159675200134235, 'topic_9': 0.08784343941008455, 'topic_10': 0.08538951328104306, 'topic_11': 0.04415008959087885, 'topic_12': 0.11456616657901691}
 Chat: 36 / 38 (data\🇮🇹ItalianCamCluBт 🇮🇹.log):
 Topics: {'topic_0': 0.057729868517708775, 'topic_1': 0.11632001597045384, 'topic_2': 0.06449706241299093, 'topic_3': 0.07536439236367956, 'topic_4': 0.10929692478871994, 'topic_5': 0.0519009617838222, 'topic_6': 0.07140161984949675, 'topic_7': 0.06263822847853537, 'topic_8': 0.11796995840629551, 'topic_9': 0.07635405513771153, 'topic_10': 0.04982771797515252, 'topic_11': 0.07557273800987817, 'topic_12': 0.07112645630508738}
 Chat: 37 / 38 (data\👩ACTRESS ASS MUSEUM™👩.log):
 Topics: {'topic_0': 0.06172785494634014, 'topic_1': 0.08905103296180035, 'topic_2': 0.06716952074441014, 'topic_3': 0.07991036982585732, 'topic_4': 0.11103426575295618, 'topic_5': 0.051598375207770164, 'topic_6': 0.07327245576318474, 'topic_7': 0.06690262609936305, 'topic_8': 0.10704496086340425, 'topic_9': 0.07088539687310366, 'topic_10': 0.06994830097631688, 'topic_11': 0.07550265974509746, 'topic_12': 0.07595218023987145}
 Chat: 38 / 38 (data\👩Girls.log):
 Topics: {'topic_0': 0.05276802736008725, 'topic_1': 0.1083626137557598, 'topic_2':

```
0.054506831449693344, 'topic_3': 0.07385870985702951, 'topic_4': 0.1215029324628708, 'topic_5':  
0.04761445459261477, 'topic_6': 0.0773623483169981, 'topic_7': 0.06877383266162507, 'topic_8':  
0.13441601711541254, 'topic_9': 0.0812152764163712, 'topic_10': 0.048502321786856976, 'topic_11':  
0.07021060830047328, 'topic_12': 0.060906025923644476}
```

Πίνακας 2. Τελικά αποτελέσματα Telegram με LDA.

Κεφάλαιο 6. Συλλογή και ανάλυση δεδομένων από το WhatsApp

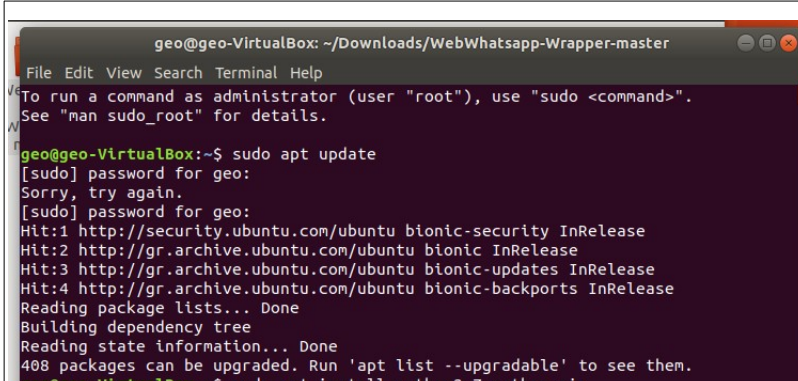
Στο κεφάλαιο αυτό θα παρουσιάσουμε τα βήματα που ακολουθήθηκαν για την συλλογή και την ανάλυση δεδομένων για την εφαρμογή μηνυμάτων WhatsApp. Για την συλλογή των δεδομένων θα χρειαστούμε μια συσκευή κινητής τηλεφωνίας ή ένα tablet με εγκατεστημένη την εφαρμογή WhatsApp καθώς και μια κάρτα SIM, λόγο του όπως αναφέραμε, το WhatsApp χρειάζεται τον αριθμό του χρήστη για την εγγραφή του στην υπηρεσία.

Αρχικά, είναι απαραίτητο να μπορέσουμε να κάνουμε εγγραφή σε μερικά group του WhatsApp. Στην επιλογή τους προσπάθησα να έχω όσο το δυνατόν περισσότερες κατηγορίες των group αυτών και ο αριθμός των group που επιλέχθηκαν ήταν τυχαίος. Σε περίπτωση που κάποιος θα ήθελε να κάνει μια ποιο στοχευμένη έρευνα, θα έπρεπε να προσέξει την κατηγορία των group αυτών και πιθανών τον αριθμό τους ή και την χώρα προέλευσης των χρηστών τους κτλ. Για την εύρεση των group χρησιμοποιήθηκε η μηχανή αναζήτησης Google καθώς και ιστότοποι που περιέχουν συνδέσμους (invitation links) για τα groups. Οι σύνδεσμοι αυτοί έχουν το επίθεμα chat.whatsapp.com.

Αφού κάνουμε την εγγραφή στα group, με την βοήθεια ενός API του οποίου η χρήση είναι ελεύθερη στο Github στον σύνδεσμο <https://github.com/mukulhase/WebWhatsapp-Wrapper> κάνουμε τα ακόλουθα βήματα.

6.1 Εγκατάσταση απαιτούμενων εργαλείων

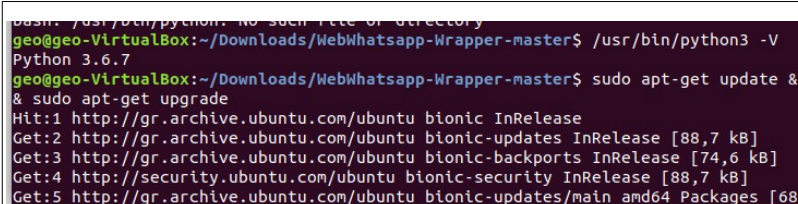
Αρχικά από το terminal του Ubuntu με την εντολή `sudo apt update`, κάνουμε ένα γενικό update για να είμαστε σίγουροι ότι το σύστημα μας είναι πλήρως ενημερωμένο (βλ. Εικόνα 38). Μετά το update με την εντολή `sudo apt-get update && sudo apt-get upgrade` κοιτάμε την έκδοση του Python στο σύστημα μας όπως βλέπουμε στην Εικόνα 39 έχουμε την έκδοση 3.



```
geo@geo-VirtualBox: ~/Downloads/WebWhatsapp-Wrapper-master
File Edit View Search Terminal Help
To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.

geo@geo-VirtualBox:~$ sudo apt update
[sudo] password for geo:
Sorry, try again.
[sudo] password for geo:
Hit:1 http://security.ubuntu.com/ubuntu bionic-security InRelease
Hit:2 http://gr.archive.ubuntu.com/ubuntu bionic InRelease
Hit:3 http://gr.archive.ubuntu.com/ubuntu bionic-updates InRelease
Hit:4 http://gr.archive.ubuntu.com/ubuntu bionic-backports InRelease
Reading package lists... Done
Building dependency tree
Reading state information... Done
408 packages can be upgraded. Run 'apt list --upgradable' to see them.
```

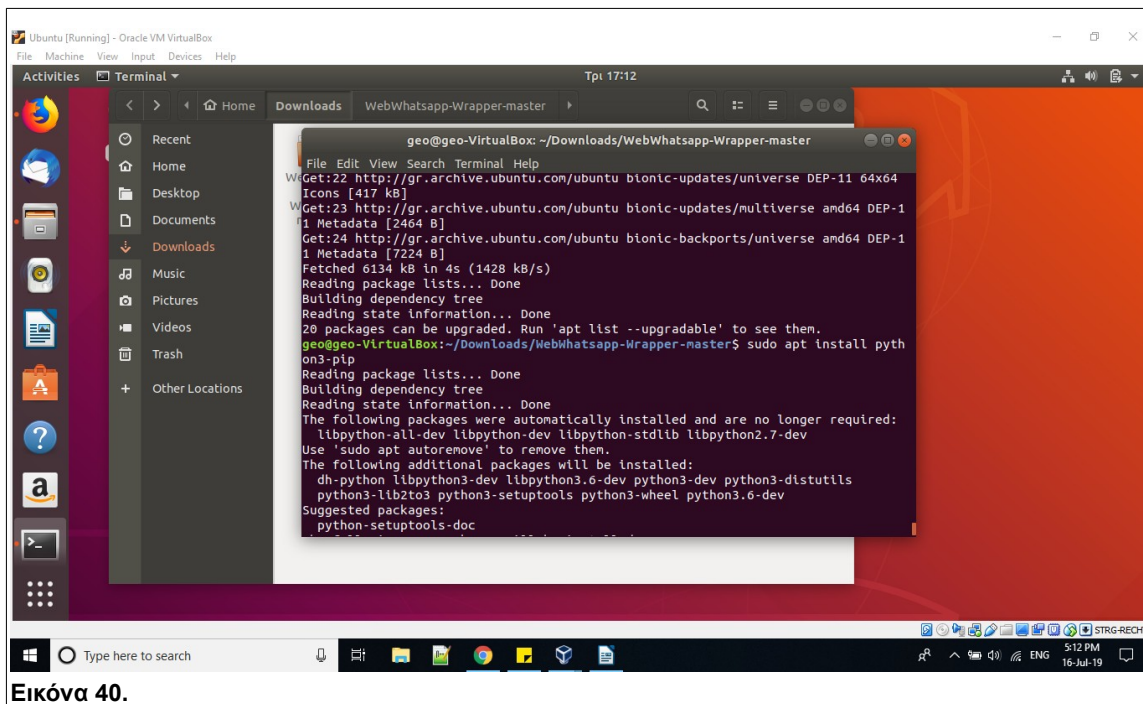
Εικόνα 38.



```
bash: /usr/bin/python: no such file or directory
geo@geo-VirtualBox:~/Downloads/WebWhatsapp-Wrapper-master$ /usr/bin/python3 -V
Python 3.6.7
geo@geo-VirtualBox:~/Downloads/WebWhatsapp-Wrapper-master$ sudo apt-get update &
& sudo apt-get upgrade
Hit:1 http://gr.archive.ubuntu.com/ubuntu bionic InRelease
Get:2 http://gr.archive.ubuntu.com/ubuntu bionic-updates InRelease [88,7 kB]
Get:3 http://gr.archive.ubuntu.com/ubuntu bionic-backports InRelease [74,6 kB]
Get:4 http://security.ubuntu.com/ubuntu bionic-security InRelease [88,7 kB]
Get:5 http://gr.archive.ubuntu.com/ubuntu bionic-updates/main amd64 Packages [68
```

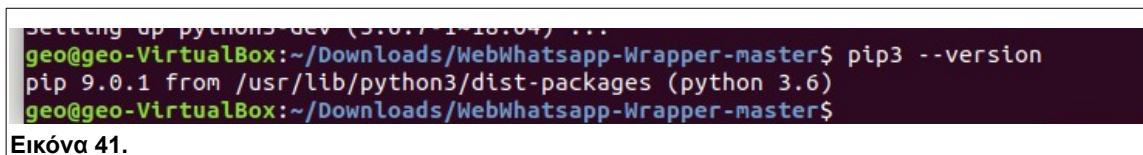
Εικόνα 39.

Με την εντολή `sudo apt install python3-pip` εγκαθιστούμε το `pip` το οποίο είναι ένα `package manager` του Python (βλ. Εικόνα 40.).



Εικόνα 40.

Και επιβεβαιώνουμε την έκδοση του `pip` όπως βλέπουμε στην Εικόνα 41.



Εικόνα 41.

Στο επόμενο βήμα με την εντολή `pip3 install -r requirements.txt` (βλ. Εικόνα 42.) εγκαθιστούμε στο σύστημα μας τα απαιτούμε εργαλεία που χρειαζόμαστε. Το αρχείο `requirements.txt` βρίσκεται στον φάκελο `WebWhatsapp-Wrapper-master` που κατεβάσαμε από το Github.

```

geo@geo-VirtualBox:~/Downloads/WebWhatsapp-Wrapper-master
geo@geo-VirtualBox:~/Downloads/WebWhatsapp-Wrapper-master$ pwd
~/home/geo/Downloads/WebWhatsapp-Wrapper-master
geo@geo-VirtualBox:~/Downloads/WebWhatsapp-Wrapper-master$ pip install -r requirements.txt
bash: /usr/bin/pip: No such file or directory
geo@geo-VirtualBox:~/Downloads/WebWhatsapp-Wrapper-master$ pip3 install -r requirements.txt
Collecting python-dateutil<=2.6.0 (from -r requirements.txt (line 1))
  Downloading https://files.pythonhosted.org/packages/41/17/c62facbfbfd163c7f57f3844689e3a78bae1f403648a6afb1d086d87fbb/python_dateutil-2.8.0-py2.py3-none-any.whl (226kB)
    100% |#####| 235kB 272kB/s
Collecting selenium<=3.4.3 (from -r requirements.txt (line 2))
  Downloading https://files.pythonhosted.org/packages/80/d6/4294f0b4bce4de0abf13e17190289f9d0613b0a44e5dd6a7f5ca98459853/selenium-3.141.0-py2.py3-none-any.whl (904kB)
    100% |#####| 911kB 548kB/s
Collecting six<=1.10.0 (from -r requirements.txt (line 3))
  Downloading https://files.pythonhosted.org/packages/73/fb/00a976f728dd1f6ce898238ce23f502a71c0ac0ecfedb0e0d8c04e9/six-1.12.0-py2.py3-none-any.whl
Collecting python-axolotl (from -r requirements.txt (line 4))
  Downloading https://files.pythonhosted.org/packages/0f/b3/4a0b0e664435691bba36face366cece506d002294c293bd3dc40c2b2ba/python_axolotl-0.2.3-py3-none-any.whl (76kB)

```

Εικόνα 42.

Με την εντολή `pip install webwhatsappi` εγκαθιστούμε ένα package το οποίο χρησιμοποιείται για την παροχή μιας διεπαφής Python (interface) για αλληλεπίδραση με το WhatsApp Web για την αποστολή και λήψη Whatsapp μηνυμάτων (βλ. Εικόνα 43.).

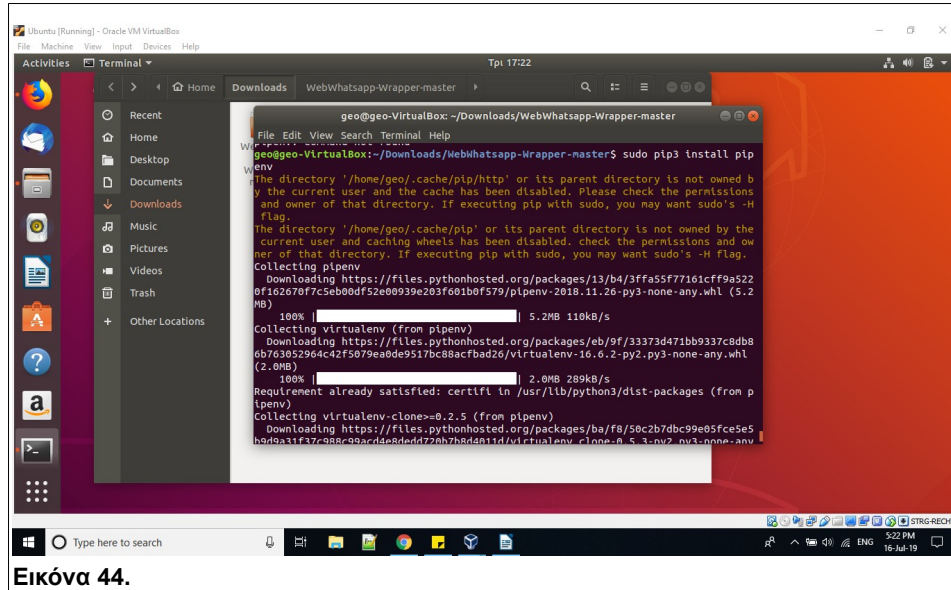
```

geo@geo-VirtualBox:~/Downloads/WebWhatsapp-Wrapper-master
geo@geo-VirtualBox:~/Downloads/WebWhatsapp-Wrapper-master$ pip install webwhatsappi
bash: /usr/bin/pip: No such file or directory
geo@geo-VirtualBox:~/Downloads/WebWhatsapp-Wrapper-master$ pip3 install webwhatsappi
Collecting webwhatsappi
  Downloading https://files.pythonhosted.org/packages/22/fc/5c0d6a59ba58d7920bbd3d86cb953b9325a2ef80e1bb180d5d8edcb22f3/webwhatsappi-2.0.5-py3-none-any.whl (43kB)
    100% |#####| 51kB 811kB/s
Collecting selenium<=3.4.3 (from webwhatsappi)
  Using cached https://files.pythonhosted.org/packages/80/d6/4294f0b4bce4de0abf13e17190289f9d0613b0a44e5dd6a7f5ca98459853/selenium-3.141.0-py2.py3-none-any.whl
Collecting six<=1.10.0 (from webwhatsappi)
  Using cached https://files.pythonhosted.org/packages/73/fb/00a976f728dd1f6ce898238ce23f502a71c0ac0ecfedb0e0d8c04e9/six-1.12.0-py2.py3-none-any.whl
Collecting python-axolotl (from webwhatsappi)
  Using cached https://files.pythonhosted.org/packages/0f/b3/4a0b0e664435691bba36face366cece506d002294c293bd3dc40c2b2ba/python_axolotl-0.2.3-py3-none-any.whl
Collecting cryptography (from webwhatsappi)
  Using cached https://files.pythonhosted.org/packages/07/18/16c55376336b46347a7

```

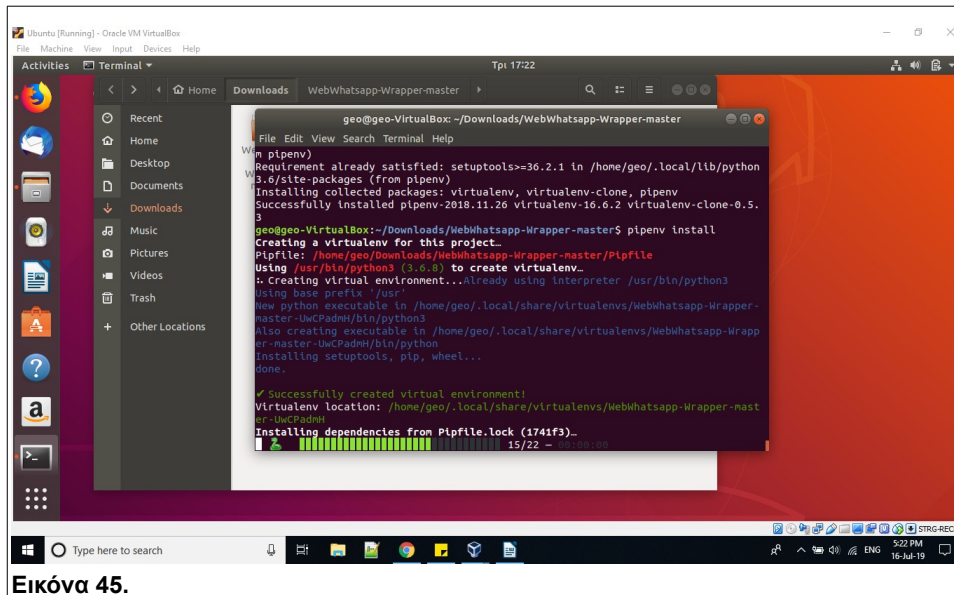
Εικόνα 43.

Επόμενο βήμα είναι η εγκατάσταση του `pipenv` το οποίο είναι ένα εργαλείο για την αυτόματη δημιουργία και διαχείρισης ενός εικονικού περιβάλλοντος. Αυτό το πετυχαίνουμε με την εντολή `sudo pip3 install pipenv` (βλ. Εικόνα 44.).



Εικόνα 44.

Αφού εγκαταστήσαμε το `pipenv` δημιουργούμε ένα εικονικό περιβάλλον για το project μας με την εντολή `pipenv install` (βλ. Εικόνα 45.)



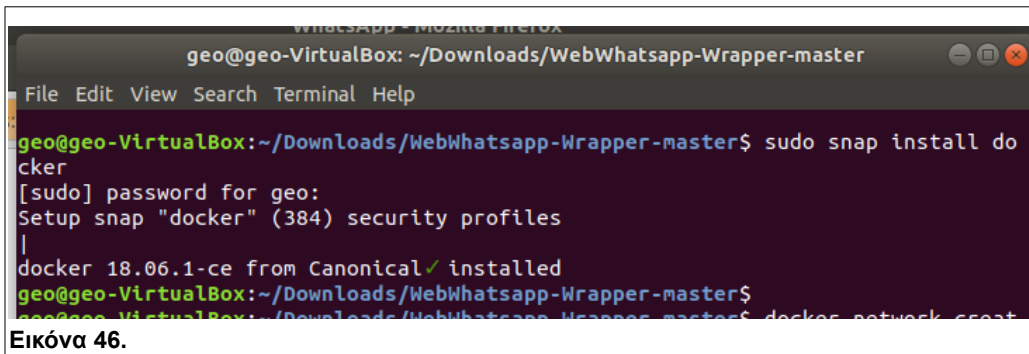
Εικόνα 45.

Καλό θα ήταν να τρέξουμε το Selenium και το webwhatsapp client ως Docker containers. Με αυτόν τον τρόπο θα αποφύγουμε τυχόν προβλήματα εγκατάστασης ή προβλήματα εξάρτησης για το πρότζεκτ μας.

Το Docker είναι ένα σύστημα ανοιχτού κώδικα που διευκολύνει την ασφαλή δημιουργία, ανάπτυξη και εκτέλεση εφαρμογών χρησιμοποιώντας containers. Τα containers επιτρέπουν στους προγραμματιστές να πακετάρουν μια εφαρμογή με όλα τα μέρη που χρειάζεται, όπως για παράδειγμα τις βιβλιοθήκες και άλλες εξαρτήσεις, και να τα μεταφέρουν όλα ως ένα πακέτο.

Αυτά τα πακέτα ή αλλιώς containers λογισμικού θα είναι ικανά να τρέξουν σε οποιοδήποτε άλλο μηχάνημα, ανεξάρτητα από το περιβάλλον και τις προσαρμοσμένες ρυθμίσεις. Το Docker είναι διαθέσιμο για εφαρμογές Linux και Windows.

Με την εντολή `sudo snap install docker` εγκαθιστώ ένα Docker Engine στο σύστημα μου (Εικόνα 46.).



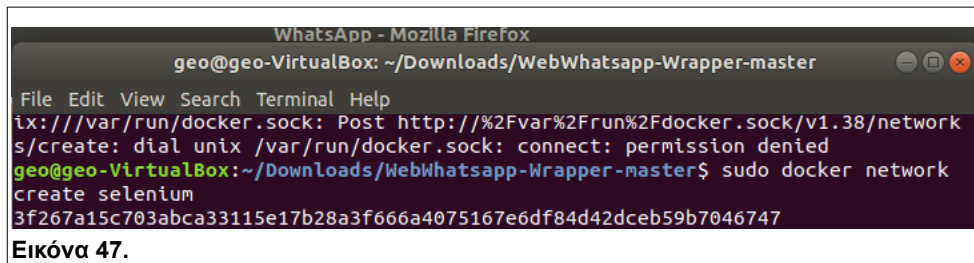
```

geo@geo-VirtualBox: ~/Downloads/WebWhatsapp-Wrapper-master
File Edit View Search Terminal Help
geo@geo-VirtualBox:~/Downloads/WebWhatsapp-Wrapper-master$ sudo snap install docker
[sudo] password for geo:
Setup snap "docker" (384) security profiles
|
docker 18.06.1-ce from Canonical ✓ installed
geo@geo-VirtualBox:~/Downloads/WebWhatsapp-Wrapper-master$

```

Εικόνα 46.

Με την εντολή `docker network create selenium` (Εικόνα 47.) συνδέω στο δίκτυο του Docker το container Selenium και ύστερα το τρέχω με την εντολή `docker run -d -p 4444:4444 -p 5900:5900 --name firefox --network selenium -v /dev/shm:/dev/shm selenium/standalone-firefox-debug:3.14.0-curium` (Εικόνα 48.).

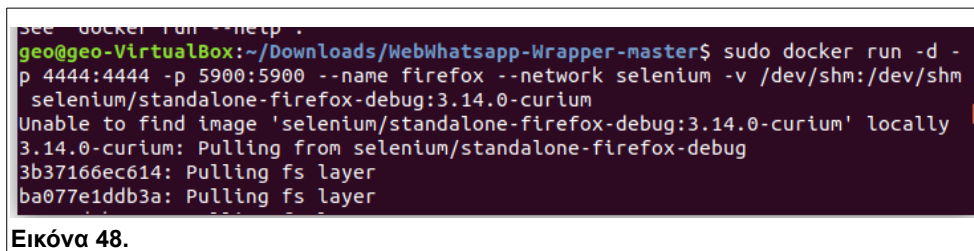


```

WhatsApp - Mozilla Firefox
geo@geo-VirtualBox: ~/Downloads/WebWhatsapp-Wrapper-master
File Edit View Search Terminal Help
ix:///var/run/docker.sock: Post http://%2Fvar%2Frun%2Fdocker.sock/v1.38/networks/create: dial unix /var/run/docker.sock: connect: permission denied
geo@geo-VirtualBox:~/Downloads/WebWhatsapp-Wrapper-master$ sudo docker network create selenium
3f267a15c703abca33115e17b28a3f666a4075167e6df84d42dceb59b7046747

```

Εικόνα 47.



```

See 'docker run --help'.
geo@geo-VirtualBox:~/Downloads/WebWhatsapp-Wrapper-master$ sudo docker run -d -p 4444:4444 -p 5900:5900 --name firefox --network selenium -v /dev/shm:/dev/shm selenium/standalone-firefox-debug:3.14.0-curium
Unable to find image 'selenium/standalone-firefox-debug:3.14.0-curium' locally
3.14.0-curium: Pulling from selenium/standalone-firefox-debug
3b37166ec614: Pulling fs layer
ba077e1ddb3a: Pulling fs layer

```

Εικόνα 48.

Η εντολή `sudo docker build -t webwhatsapi .` (το κενό και η τελεία ανήκουν στην εντολή) χρησιμοποιεί το `dockerfile` για να δημιουργήσει ένα νέο `image` με όλα τα απαιτούμενα πακέτα από το αρχείο `requirements.txt` (Εικόνα 49).

```

permissions denied
geo@geo-VirtualBox:~/Downloads/WebWhatsapp-Wrapper-master$ sudo docker build -t
webwhatsapi .
Sending build context to Docker daemon 611.8kB
Step 1/4 : FROM python:2.7
2.7: Pulling from library/python
5ae19949497e: Pull complete
ed3d96a2798e: Pull complete
f12136850781: Pull complete
1a9ad5d5550b: Pull complete
6f18049a0455: Pull complete

```

Εικόνα 49.

Και τρέχουμε το `container` αυτό με την εντολή `docker run --network selenium -it -e SELENIUM='http://firefox:4444/wd/hub' -v $(pwd):/app webwhatsapi /bin/bash -c "pip install .;pip list;python sample/remote.py"` (Εικόνα 50).

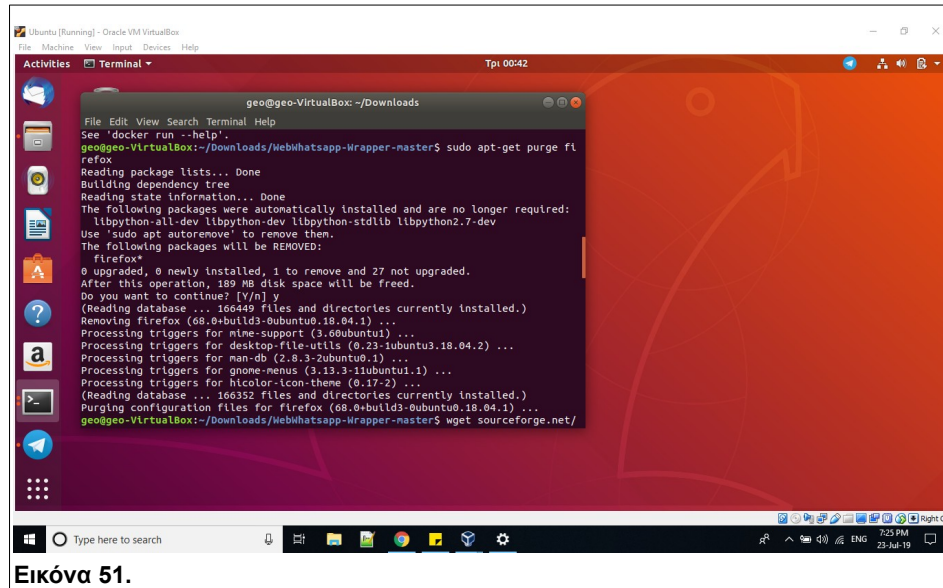
```

geo@geo-VirtualBox:~/Downloads/WebWhatsapp-Wrapper-master$ sudo docker run --ne
twork selenium -it -e SELENIUM='http://firefox:4444/wd/hub' -v $(pwd):/app web
whatsapi /bin/bash -c "pip install .;pip list;python sample/remote.py"
[sudo] password for geo:
DEPRECATION: Python 2.7 will reach the end of its life on January 1st, 2020. Pl
ease upgrade your Python as Python 2.7 won't be maintained after that date. A f
uture version of pip will drop support for Python 2.7.
Processing /app
Requirement already satisfied: python-dateutil>=2.6.0 in /usr/local/lib/python2
.7/site-packages (from webwhatsapi==2.0.3) (2.8.0)
Requirement already satisfied: selenium>=3.4.3 in /usr/local/lib/python2.7/site
-packages (from webwhatsapi==2.0.3) (3.141.0)
Requirement already satisfied: six>=1.10.0 in /usr/local/lib/python2.7/site-pac
kages (from webwhatsapi==2.0.3) (1.12.0)
Requirement already satisfied: python-axolotl in /usr/local/lib/python2.7/site-
packages (from webwhatsapi==2.0.3) (0.2.3)
Requirement already satisfied: cryptography in /usr/local/lib/python2.7/site-pa
ckages (from webwhatsapi==2.0.3) (2.7)
Requirement already satisfied: python-magic in /usr/local/lib/python2.7/site-pa

```

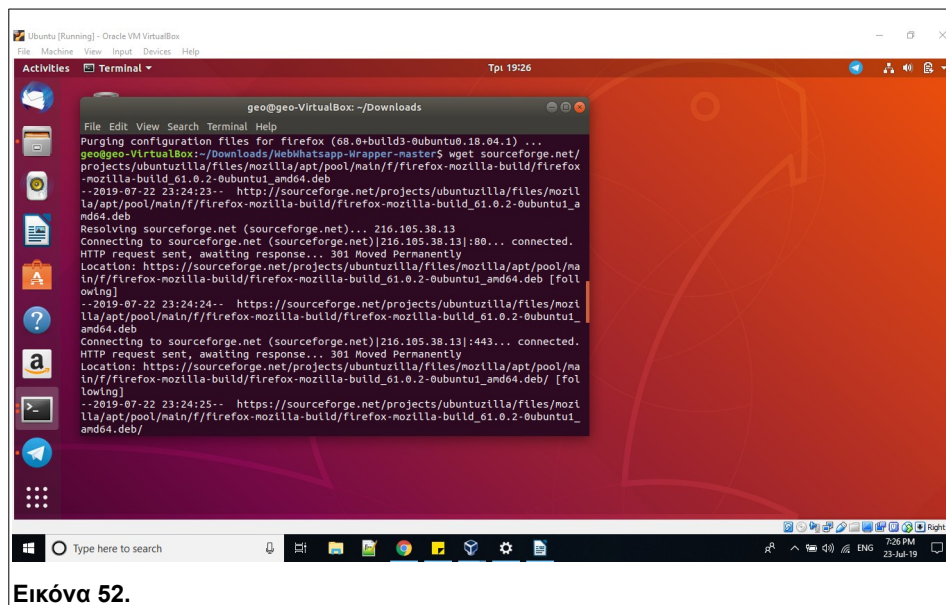
Εικόνα 50.

Λόγο του ότι το συγκεκριμένο πρότζεκτ απαιτεί συγκεκριμένη έκδοσή του Firefox, με την εντολή `sudo apt-get purge firefox` απεγκαθιστώ το Firefox από το σύστημά μου (Εικόνα 51.).



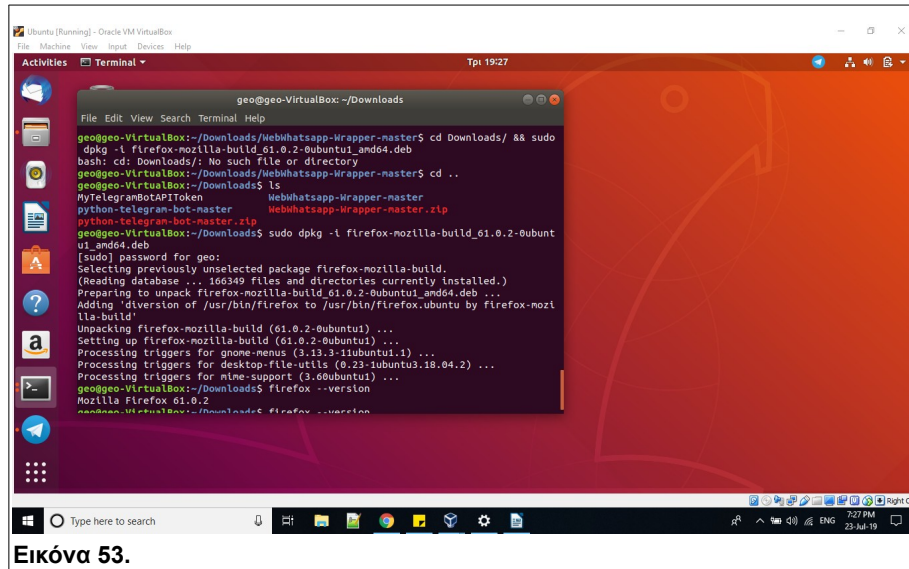
Εικόνα 51.

Με την εντολή: `wget sourceforge.net/projects/ubuntuzilla/files/mozilla/apt/pool/main/f/firefox-mozilla-build/firefox-mozilla-build_61.0.2-0ubuntu1_amd64.deb` κατεβάζω την έκδοση 61.0.2 του Firefox που απαιτεί το πρότζεκτ (Εικόνα 52.).



Εικόνα 52.

Με την εντολή `sudo dpkg -i firefox-mozilla-build_61.0.2-0ubuntu1_amd64.deb` εγκαθιστώ την συγκεκριμένη έκδοση στο σύστημά μου (Εικόνα 53).



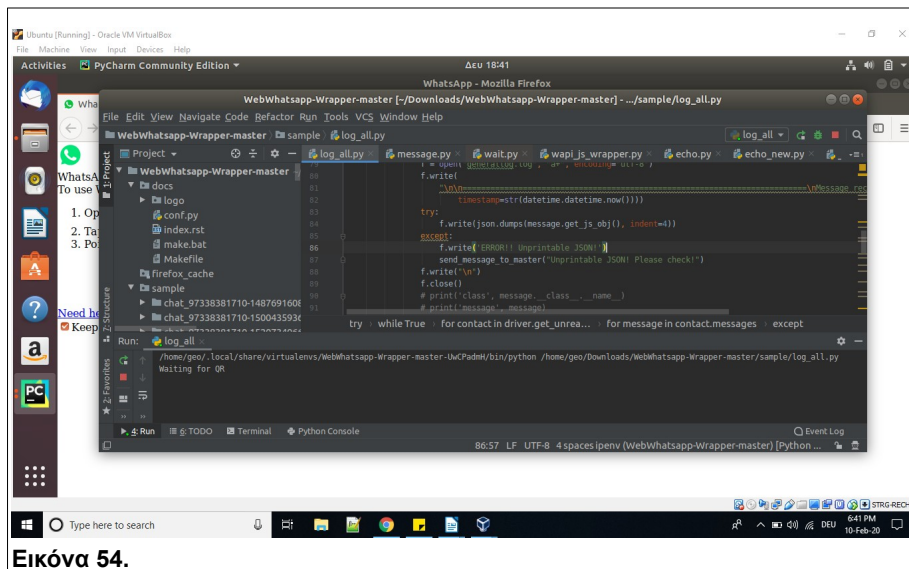
```

geo@geo-VirtualBox:~/Downloads
File Edit View Search Terminal Help
geo@geo-VirtualBox:~/Downloads/WhatsApp-Wrapper-master$ cd Downloads/ && sudo
dpkg -i firefox-mozilla-build_61.0.2-0ubuntu1_amd64.deb
bash: cd: Downloads/: No such file or directory
geo@geo-VirtualBox:~/Downloads$ ls
MyTelegramBotToken      WhatsApp-Wrapper-master
python-telegram-bot-master  WhatsApp-Wrapper-master.zip
python-telegram-bot-master.zip
geo@geo-VirtualBox:~/Downloads$ sudo dpkg -i firefox-mozilla-build_61.0.2-0ubuntu1_amd64.deb
[sudo] password for geo:
Selecting previously unselected package firefox-mozilla-build.
(Reading database ... 106349 files and directories currently installed.)
Preparing to unpack firefox-mozilla-build_61.0.2-0ubuntu1_amd64.deb ...
Adding 'd'version of /usr/bin/firefox to /usr/bin/firefox.ubuntu by firefox-mozilla-build'
Unpacking firefox-mozilla-build (61.0.2-0ubuntu1) ...
Setting up firefox-mozilla-build (61.0.2-0ubuntu1) ...
Processing triggers for gnome-menus (3.13.3-11ubuntu1) ...
Processing triggers for desktop-file-utils (0.23-1ubuntu3.18.04.2) ...
Processing triggers for mime-support (3.0ubuntu1) ...
geo@geo-VirtualBox:~/Downloads$ firefox --version
Mozilla Firefox 61.0.2
  
```

Εικόνα 53.

6.2 Συλλογή δεδομένων

Μετά την εγκατάσταση των εργαλείων τρέχουμε το αρχείο `log_all.py` το οποίο βρίσκεται στο path `WebWhatsapp-Wrapper-master/sample/log_all.py` (Εικόνα 54.).

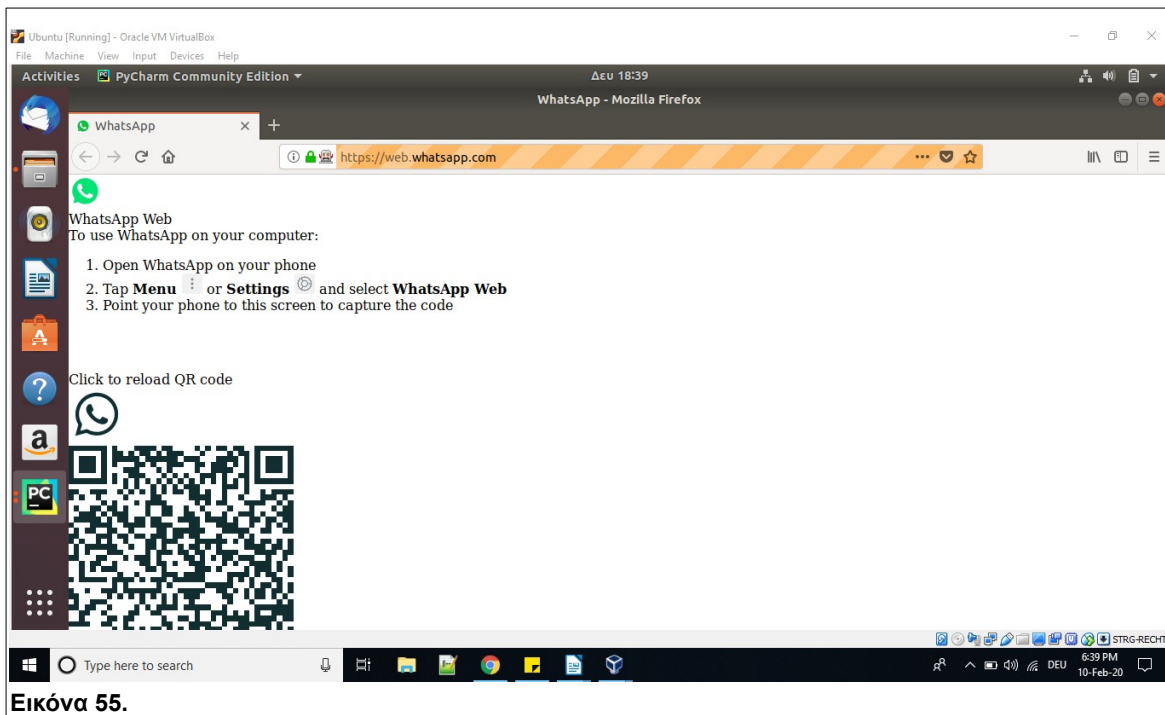


```

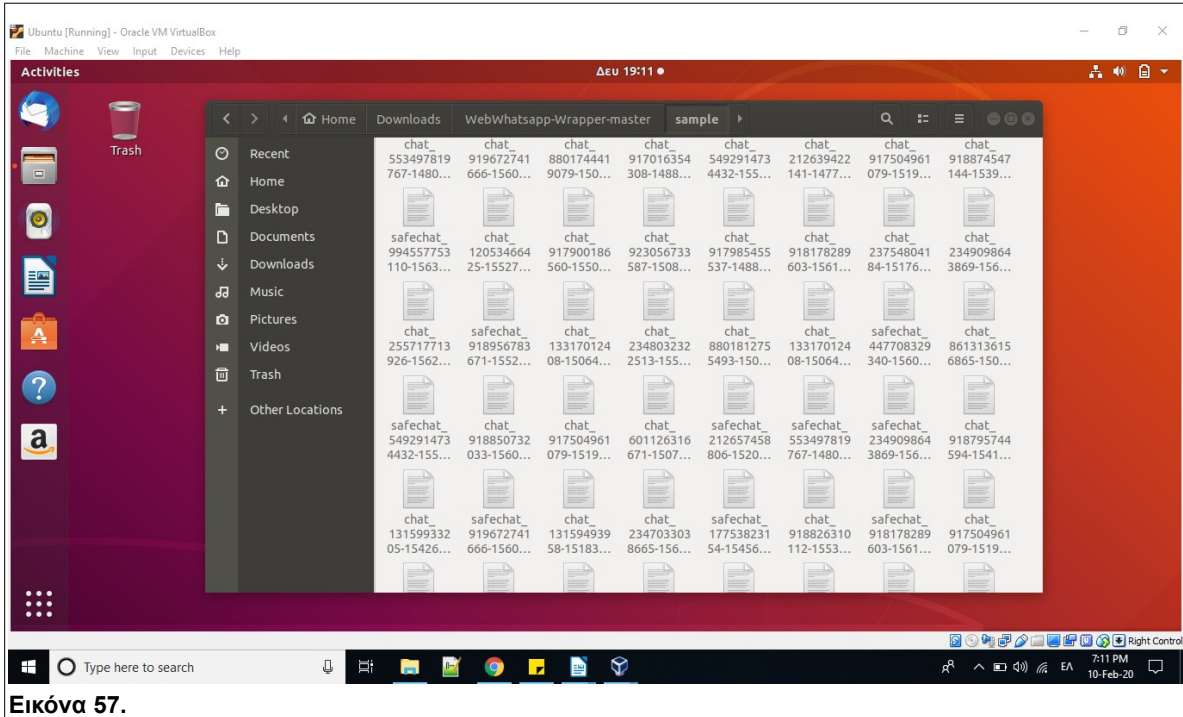
WebWhatsapp-Wrapper-master [-/Downloads/WebWhatsapp-Wrapper-master] -.../sample/log_all.py
File Edit View Navigate Code Refactor Run Tools VCS Window Help
WebWhatsapp-Wrapper-master sample log_all.py
Project
WebWhatsapp-Wrapper-master
  docs
  logo
  1. Op
  2. Ta
  3. Po
  conf.py
  index.rst
  make.bat
  Makefile
  Firefox_cache
  sample
    chat_97338381710-1487691608
    chat_97338381710-1500435924
  2. Structure
  Run: log_all
/home/geo/.local/share/virtualenvs/WebWhatsapp-Wrapper-master-UcPdmi/bin/python /home/geo/Downloads/WebWhatsapp-Wrapper-master/sample/log_all.py
Waiting for QR
86:57 LF UTF-8 4 spaces (penv (WebWhatsapp-Wrapper-master)) [Python ...
  
```

Εικόνα 54.

Μετά μας ανοίγει ένα παράθυρο του Firefox με ένα QR code (Εικόνα 55.) το οποίο σαρώνουμε πηγαίνοντας στο μενού των ρυθμίσεων του WhatsApp επιλέγοντας την επιλογή WhatsApp Web (Εικόνα 56.).



Μετά από τα παραπάνω βήματα αρχίζει η συλλογή των μηνυμάτων στο path `WebWhatsapp-Wrapper-master/sample/` (Εικόνα 57.)



Εικόνα 57.

6.3 Ανάλυση δεδομένων

Αφού συλλέξουμε τα δεδομένα, προχωράμε με την επεξεργασία και ανάλυση τους. Αρχικά ορίζουμε ότι κάθε μήνυμα του WhatsApp θα είναι ένα document. Στο Script 12. (document.py) βλέπουμε την κλάση Document. Τα πεδία τις δομής είναι ο αποστολέας, η ημερομηνία – ώρα αποστολής και το κείμενο του μηνύματος.

```
class Document:

    def __init__(self, sender, datetime, text):
        """
        Arxikpouhsh ths domhs arxeiou pou xrhsimopoihtai gia thn analysi.
        Kathe mynhma einai ena document.
        Ena document perigrafete apo ta pedia apostoleos, wra kai mera kai periexomeno
        """
        self.sender = sender
        self.datetime = datetime
        self.text = text
```

Script 12. document.py

Στο Script 13. (word_cloud_helper.py) ορίζουμε μια συνάρτηση με την βοήθεια της οποίας θα πάρουμε την απεικόνιση των Word Clouds. Το Word Cloud είναι μια τεχνική που χρησιμοποιείται για την απεικόνιση δεδομένων κειμένου στα οποία το μέγεθος κάθε λέξης υποδεικνύει τη συχνότητα ή τη σημασία της. Σημαντικά σημεία δεδομένων κειμένου μπορούν να επισημανθούν με την χρήση ενός σύννεφου λέξεων (Word Cloud). Τα Word Clouds χρησιμοποιούνται ευρέως για την ανάλυση δεδομένων από ιστότοπους κοινωνικών δικτύων.

```
import matplotlib.pyplot as plt
from PIL import Image
from wordcloud import WordCloud

def save_word_cloud(word_frequencies, file_name):
    wc = WordCloud()
    wc.generate_from_frequencies(word_frequencies) #Paragei mia lista me tuples opou to proto melos einai
    enas oros, kai to deytero h syxnothta emfanishs tou p.x [('java', 15), ('scala', 12), ('spring', 12)]

    # Kanei display ta paragomena images
    plt.imshow(wc, interpolation='bilinear')
    plt.axis("off")
    plt.savefig(file_name+'.png', bbox_inches='tight')
```

Script 13. word_cloud_helper.py

Το Script 14. (text_processing.py) μας βοηθάει με την επεξεργασία των κειμένων των μηνυμάτων που συλλέξαμε.

```
import re
import string
from urllib.parse import urlparse
from nltk import stem, tokenize
from sklearn.feature_extraction import stop_words
import string
import nltk
from nltk.stem import WordNetLemmatizer

# TODO: uncomment tin proti for a pou tha treksei
# nltk.download('wordnet') #einai ena megalo dwrean lexical database to opoio prwsferei dinatotites
# limatopoihsis

wordnet_lemmatizer = WordNetLemmatizer() #Einai mia methodos pou kanei lhmata apo tis lekseis pou tou
# dinw... Edw to kataxwroume se mia metavliti gia na tin kaloume pio eykola
stemmer = stem.PorterStemmer() #.LancasterStemmer() Υπαρχουν 3 Το Porter einai to pio palio alla
# xrisimopoihtai syxna, Snowball(Porter2), to Lancaster (Paice-Husk) einai to pio epithetiko mporei na
# odigeisei se lathoi
stopwords = set(stop_words.ENGLISH_STOP_WORDS)
exclude = set(string.punctuation).union(stopwords) #To set pairnei mia list, ena tuple h ena dictionary kai
# epistrefei to periexomeno tous (elements) sorted kai non-repeated. To union se dyo set() epistrefei to
# periexomeno pou exoun ta dyo set xwris na epanalambanei ta element an ayto yparxei kai sta 2 set... To
# epistrefei mono mia fora
exclude.add("http") # To exclude einai ena set me string punctuation kai stopwords. To add() prwsthetei sto
# set ayto mia leksi. An i leksei pou kanoume add yparxei idi, tote den tin prwsthetei
exclude.add("https")
exclude.add("www")
exclude.add("com")

def is_URL(s):
    # Dedomenoy enos string to kanei parse kai epistrefei an einai url
    try:
        result = urlparse(s)
        return all([result.scheme, result.netloc, result.path]) # to all epistrefei True an ola ta paidia einai gemata
    except:
        return False

def normalize(s):
    return ' '.join(normalised_tokens(s)) # epistrefei tis lekseis pou pairnei apo tin lista pou epistrefei i
    # synartisi normalized_tokens(s) xwris komma metaksi tous diladi oles tis lekseis me ena keno metaksi tous

def normalised_tokens(s):
    words = tokenize.casual_tokenize(s.lower())
```

```

final_words = []
urls = []
for word in words:
    if not is_URL(s):
        final_words.extend(tokenize.wordpunct_tokenize(word.lower().strip()))
    else:
        urls.extend(word)
        # final_words.append(word)
return [stemmer.stem(wordnet_lemmatizer.lemmatize(w, pos="v")) for w in final_words if w not in
exclude] # kanei lemmatize kai epistrefei apo tis lekseis pou to lemma tous einai rima to rima pos= "v"
diladi verb px goes, go, went epistrefoun go

""" To input ths methodou tha prepei na einai *Document* datatype """
def process(document):
    # Apo th domh document, pairnw to pedio text kai to kanw lowercase
    text = document.text.lower()
    normalised = normalize(text) # Apo tin synartisi pou eftiaksa pio panw, epistrefei tis lekseis pou pairnei
apo tin lista pou epistrefei i synartisi normalized_tokens(s) xwris komma metaksi tous diladi oles tis lekseis
me ena keno metaksi tous
    # text = re.sub(r"[{}]" .format(string.punctuation), " ", text)
    return normalised

```

Script 14. text_proccesing.py

Με την βοήθεια του Script 15. (files_management.py) παίρνουμε τα documents από τα files με τις συνομιλίες του κάθε group. Όπως προαναφέραμε κάθε μήνυμα το ορίζουμε ως ένα document.

```

import os
from data_types.document import Document

excluded_logs = ['generallog.log', 'geckodriver.log']

def get_files(root_directory):
    """
    Helper method gia na epistrafoyn ola ta arxeia me thn katalhxh log
    pou periexontai sto root directory. Ta arxeia log periexoyn ta mhnymata
    pou syllegontai. Gia kathe arxeio to full path epistrefetai. Kapoia logs
    periexoyn plhrofories ki oxi dedomena (excluded_logs). Ayta ta arxeia
    den epistrefontai.
    """
    files = []
    # r=root, d=directories, f = files
    for r, d, f in os.walk(root_directory):
        for file in f:
            if file.endswith('.log') and file not in excluded_logs:
                files.append(os.path.join(r, file))

    return files

```

```

def get_documents(files):
    """
    Methodos gia thn anagnwsh twn arxeiwn poy periexoun dedomena ki epistrofh
    sulloghs dict, opou to kleidi einai to onoma toy arxeioy kai to value
    einai array apo domes Document poy paristanoun ta mhnymata sto arxeio.
    """
    documents = dict()

    """Epeidh h fortwsh twn mhnymatwn mporei na parei wra, xrhsimopoioume counter
    kai ton synoliko aritho twn arxeiwn gia na ektypwspoyme thn proodo sth consola"""
    file_counter = 0
    total_files = len(files)
    for file in files:
        file_counter +=1
        print("Reading document ", file_counter, " from ", total_files)
        """To documents_in_file apothileyi lista twn document (mhnymatwn) sto arxei
        poy ginetai anagnwsh"""
        documents_in_file = []
        with open(file) as fp:
            line = fp.readline()
            document = None
            # Gia kathe grammh sto arxeio
            while line:
                # An h grammi arxizei me '[' tote ayht antistoixh se ena ena kainourio mhnyma
                if line.startswith('['):
                    # Prin diavasw to epomeno mhnyma, epe3ergazomai oti exw diavasei mexri twra
                    if document != None :
                        sender = None
                        document_text = ""
                        datetime = ""
                        if '|' in document: #An sto string poy perigrafei to arxeio yparxei ayto | to symvolo
                            document_info = document.split(' ')[0].strip() # Meta3y twn symbolwn [...] briskontai
                            plhrofories apostolhs toy mhntmatos (apostoleas, wra klp)
                            document_text = document.split(' ')[1].strip() # Kai to keimeno vrisketai sth
                            synexeia
                            sender = document_info.split(' | ')[0].replace('[', '').strip() #Pairnw ton apostolea tou
                            minimatos poy brisketai prin to '|' sto info kommati ki antikathisto to symbolo [ kai ta kena
                            datetime = document_info.split(' | ')[1].strip() #Pairnw tin imerominia apostolis tou
                            minimatos poy vrisketai meta to '|' ki antikautistw ] kai kena
                        else:
                            # Se periptwsh pou to mhnyma den exei to symvolo '|', tote ayto den exei hmeromhnia
                            sender = document.split(' ')[0].replace('[', '').strip()
                            document_text = document.split(' ')[1].strip()

                        d = Document(sender, datetime, document_text) #dimiourgo mia domh document d
                        documents_in_file.append(d) #Kai kanw to d append stin lista documents_in_file

                    #Afoy exw parei ta stoixeia toy prohgoymenoy mhnymatos, orizw to kainourio

```



```

        document = line
        # Alliws, an h grammh den 3ekina me '[' tote einai synexeia to prohgoymeno mhnymatos
    else:
        document += " "+line
        line = fp.readline()
        # Afoy diavasw ola ta mhnymata sto arxeio, ta apothikew sto dictionary
        documents[file] = documents_in_file

    return documents

def update_progress(progress):
    print("\r[{}] {}%".format('#'*(progress/10), progress))

```

Script 15. files_management.py

Στο Script 16. (corpusCreator.py) ορίζουμε την κλάση CorpusCreator η οποία μας επιστρέφει ένα array με όλα τα documents ανεξαρτήτου του chat του οποίου ανήκουν.

```

from utils import files_management

class CorpusCreator:

    def __init__(self, directory):
        """
        Arxikopoihsh ths dhmiourgias toy corpus.
        H parametros *directory* prepei na einai ena string me to path tou directory
        pou periexei ta data pou syllexthikan
        """

        self.data_directory = directory
        self.corpus = []
        self.documents_per_chat = dict()

    def create(self):
        """
        Dhmiourgeia tou corpus gia to training me vasi ta documents
        """
        all_logs = files_management.get_files(self.data_directory)
        self.documents_per_chat = files_management.get_documents(all_logs)
        for file in self.documents_per_chat:
            self.corpus.extend(self.documents_per_chat[file])

    def get(self):
        return self.corpus

```

Script 16. corpusCreator.py

Στο Script 17. (Kmeans.py) βλέπουμε την κλάση KmeansClustering όπου με την βοήθεια του αλγορίθμου K-means κάνουμε το clustering.

```

from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.cluster import KMeans
from preprocessing import text_processing
from utils import word_cloud_helper

class KMeansClustering:

    def __init__(self):
        self.k = 0
        self.vectorizer = None
        self.clusters = None

    def compute(self, corpus, clusters):
        """
        corpus: domh arxeiwn poy xrhsimopoiouyntai gia thn ekpaideysi
        clusters: o arithmos twn clusters poy dhmiourgoiountai
        """
        self.k = clusters
        # Xrhsimopoiyme tf-idf gia na paristanoyme ta arxia san vectors apo lexeis
        self.vectorizer = TfidfVectorizer(preprocessor=text_processing.process,
                                         encoding='utf-16', stop_words='english')
        # Gia ta arxia sto corpus, ta metatrepoyme se vectors
        tfidf = self.vectorizer.fit_transform(corpus)
        # Kai 3ekiname ton K-means++ kanontas 100 to poly epanalysis gia na dhmiourghsoyme ta cluster
        self.clusters = KMeans(n_clusters=self.k, init='k-means++', max_iter=100, n_init=1).fit(tfidf)

    def predict_cluster(self, document):
        """Dedomenoy enos kainoyrioy mhnymatos (document),
        an den exoyme dhmiourghsh ta clusters, emfanizoyme mhnyma"""
        if self.clusters is None:
            print("ERROR: clusters have not been computed.")
        else:
            """
            Alliws pairnoyme to vector toy mhnymatos,
            opws ayto ginetai project sto corpus (dhladh gia
            tis lexeis poy exoyme sto corpus) kai ypologizoyme
            se poio cluster brisketai pio konta o vector
            """
            return self.clusters.predict(self.vectorizer.transform([document]))

    def print_top_terms(self, top_terms):
        # Me dedomeno ton arithmo top_terms
        print("Top terms per cluster:")
        # Pairnoyme ta kenra twn clusters kai taksinomye
        # tis apostaseis twn orwn
        order_centroids = self.clusters.cluster_centers_.argsort()[:, :-1]
        # Epishs apo to vector tou corpus pairnoyme ta onomata twn orwn (lexeis)

```

```

terms = self.vectorizer.get_feature_names()
# Kai gia kathe cluster
for i in range(self.k):
    print("Cluster %d:\n" % i, end=")
    # Gia kathe deikth sta top terms
    for ind in order_centroids[i, :top_terms]:
        # Ektypwnoyme ton oro
        print(' %s' % terms[ind], end=")
        print()

def get_word_cloud(self):
    centroids = self.clusters.cluster_centers_.argsort()[:, :-1]
    terms = self.vectorizer.get_feature_names()
    # Omois me print top terms, alla pairnoyme to varos tis kathe
    # lexeis gia toys top 30 oroys ana cluster max, kai ftiaxnoyme to world cloud
    for i in range(self.k):
        print("Cluster %d:\n plotting" % i, end=")
        word_cloud_terms = {}
        for ind in centroids[i, :30]:
            if self.clusters.cluster_centers_[i, ind] == 0.0:
                break
            word_cloud_terms[terms[ind]] = self.clusters.cluster_centers_[i, ind]
        word_cloud_helper.save_word_cloud(word_cloud_terms, "images/cluster_"+str(i))

```

Script 17. Kmeans.py

Στο Script 18. (LDA.py) βλέπουμε την κλάση LDATopicExtraction όπου με την βοήθεια του LDA κάνουμε το topic extraction.

```

from sklearn.decomposition import LatentDirichletAllocation
from sklearn.feature_extraction.text import CountVectorizer

from preprocessing import text_processing
from utils import word_cloud_helper

class LDATopicExtraction:

    def __init__(self):
        self.vectorizer = None
        self.lda = None

    """Dedomenou tou corpus me ta 'Document' kai ton arithmo twn topics ypologizei ta topics"""
    def compute(self, corpus, no_features, no_topics):
        self.vectorizer = CountVectorizer(preprocessor=text_processing.process,
                                         max_df=0.95, min_df=2, max_features=no_features,
                                         stop_words='english')
        tf = self.vectorizer.fit_transform(corpus)
        self.lda = LatentDirichletAllocation(n_components=no_topics, max_iter=50, learning_method='online',

```

```

learning_offset=50., random_state=0).fit(tf)

"""Κάνει predict για το σε ποιο topic ανοικει το document"""
def predict_topic(self, document):
    if self.lda is None:
        print("ERROR: topics have not been computed.")
    else:
        topics_prediction = self.lda.transform(self.vectorizer.transform([document]))
        result = {}
        for topic_idx, topic in enumerate(topics_prediction[0]):
            index = "topic_"+str(topic_idx)
            result[index] = topic
        return result

"""Εκτυπώνει τους top terms"""
def print_topics(self, top_terms):
    feature_names = self.vectorizer.get_feature_names()
    model = self.lda
    for topic_idx, topic in enumerate(model.components_):
        print("=====")
        print("Topic %d:" % (topic_idx))
        print("\n".join([feature_names[i]+"."+str(round(topic[i], 2))
                        for i in topic.argsort()[:-top_terms - 1:-1]]))
        print("=====")

"""Επιστρέφει το word cloud"""
def get_word_cloud(self):
    feature_names = self.vectorizer.get_feature_names()
    model = self.lda
    for topic_idx, topic in enumerate(model.components_):
        word_cloud_terms = {}
        sorted_terms = topic.argsort()[:-30 - 1:-1]
        for i in sorted_terms:
            if topic[i] > 0:
                word_cloud_terms[feature_names[i]] = topic[i]
            else:
                break
        word_cloud_helper.save_word_cloud(word_cloud_terms, "images/LDA/topic_"+str(topic_idx))

```

Script 18. LDA.py

Τρέχουμε το Script 19. (tests.py) για να πάρουμε τα τελικά μας αποτελέσματα (clusters) με την βοήθεια του αλγόριθμου K-means .

```

from data_types.document import Document
from preprocessing import corpusCreator
from preprocessing import KMeans

# Diavazei ta documents
from utils import files_management

```

```

corpus = corpusCreator.CorporaCreator('Valte to path opou einai apothikeymena ta data/WebWhatsapp-
Wrapper-master/sample')
corpus.create()

corpus_documents = corpus.get()

# Dimiourgei ta clusters - 15 clusters stin periptosi mas
kmeans = KMeans.KMeansClustering()
kmeans.compute(corpus_documents, 15)

# Ektypwnei tous top 10 orous ana cluster
kmeans.print_top_terms(10)

# Dhmiourgei ena wordCloud gia to kathe cluster kai ta apothikeyei stous fakelous
kmeans.get_word_cloud()

# Vriskei to cluster enos test document
test_document = Document(sender="test", datetime="", text="Watch this youtube video")
print(kmeans.predict_cluster(test_document))

# Gia ola ta chats, taytopoiei ta clusters apo kathe document sto chat kai
# ektypwnei ta clusters me ton arithmo twv document se kathe cluster
i=0
for chat in corpus.documents_per_chat:
    clusters = {}
    i += 1
    for document in corpus.documents_per_chat[chat]:
        predicted_cluster = kmeans.predict_cluster(document)[0]
        if predicted_cluster in clusters:
            clusters[predicted_cluster] += 1
        else:
            clusters[predicted_cluster] = 1
    print("Clusters of chat ",i,"/",len(corpus.documents_per_chat),"(", chat,"): ", clusters)

```

Script 19. tests.py

Και τέλος τρέχουμε το Script 20. (test_lda.py) για να πάρουμε τα τελικά μας αποτελέσματα (topics) με την βοήθεια του LDA .

```

from data_types.document import Document
from preprocessing import corpusCreator
from preprocessing import text_processing

from preprocessing.LDA import LDATopicExtraction

# Dimiourgia tou corpus
corpus = corpusCreator.CorporaCreator(" Valte to path opou einai apothikeymena ta data/WebWhatsapp-
Wrapper-master/sample ")
corpus.create()

```

```
corpus_documents = corpus.get()
# Dimiourgei ta topics - 13 topics stin periptosi mas
lda = LDATopicExtraction()
lda.compute(corpus_documents, no_features=1000, no_topics=13)
# Ektypwnei tous top 10 orous ana topic
lda.print_topics(10)
# Dhmiourgei ena wordCloud gia to kathe topic kai ta apothikeyei ston fakelo images/LDA
lda.get_word_cloud()

#test_document = Document(sender="test", datetime="", text="Watch this youtube video")
#print(lda.predict_topic(test_document))

i=0
for chat in corpus.documents_per_chat:
    topics = {}
    total = 0
    i += 1
    for document in corpus.documents_per_chat[chat]:
        predicted_topics = lda.predict_topic(document)
        for topic in predicted_topics:
            if topic in topics:
                topics[topic] += predicted_topics[topic]
            else:
                topics[topic] = predicted_topics[topic]
        total += 1
    if total > 0:
        for topic in topics:
            topics[topic] /= total
    print("Chat: ", i, "/", len(corpus.documents_per_chat), "(", chat, "): ")
    print("Topics: ", topics)
```

Script 20. test_lda.py

6.3.1 Ανάλυση δεδομένων με τον αλγόριθμο K-means (clustering)

Αφού τρέξουμε με την βοήθεια του PyCharm το script tests.py τότε παίρνουμε το παρακάτω output, το οποίο μας λέει ότι διαβάζει τα αρχεία όπου είναι αποθηκευμένα τα μηνύματα (Εικόνα 58.).

```

File Edit View Navigate Code Refactor Run Tools VCS Window Help george (C:\Users\Geo\Desktop\george\george) - tests.py - PyCharm
george | tests | files_management.py
1 | README | corpusCreator.py | KMeans.py | text_processing.py | word_cloud_helper.py | tests.py | cp1252.py | files_management.py
2 | Run | tests
3 | C:\Users\Geo\AppData\Local\Programs\Python\Python38\python.exe C:\Users\Geo\Desktop\george\tests.py
4 | [nltk_data] Downloading package wordnet to
5 | [nltk_data] C:\Users\Geo\AppData\Roaming\nltk_data...
6 | ('others', 'i', 'whatever', 'seemed', 'otherwise', 'done', 'more', 'yourselves', 'first', 'being', 'three', 'well', 'i', 'thus', 'as', 'same', 'give', 'be', 'i', 'am', 'i', 'herein
7 | [nltk_data] Package wordnet is already up-to-date.
8 | Reading file 1 from 200
9 | Reading file 2 from 200
10 | Reading file 3 from 200
11 | Reading file 4 from 200
12 | Reading file 5 from 200
13 | Reading file 6 from 200
14 | Reading file 7 from 200
15 | Reading file 8 from 200
16 | Reading file 9 from 200
17 | Reading file 10 from 200
18 | Reading file 11 from 200
19 | Reading file 12 from 200
20 | Reading file 13 from 200
21 | Reading file 14 from 200
22 | Reading file 15 from 200
23 | Reading file 16 from 200
24 | Reading file 17 from 200
25 | Reading file 18 from 200
26 | Reading file 19 from 200
27 | Reading file 20 from 200
28 | Run | Python Console
29 | IDE and Plugin Updates: PyCharm is ready to update. (58 minutes ago)
30 | 45:46 LF UTF-8 4 spaces Python 3.8
31 | Type here to search
32 | Eικόνα 58.

```

Εικόνα 58.

Στην συνέχεια το output μας δίνει τα top terms για το κάθε cluster (Εικόνα 59.)

```

File Edit View Navigate Code Refactor Run Tools VCS Window Help george (C:\Users\Geo\Desktop\george\george) - tests.py - PyCharm
george | tests | files_management.py
1 | README | corpusCreator.py | KMeans.py | text_processing.py | word_cloud_helper.py | tests.py | cp1252.py | files_management.py
2 | Run | tests
3 | Top terms per cluster:
4 | Cluster 0:
5 | rule
6 | 45
7 | 16
8 | check
9 | appli
10 | internet
11 | 24
12 | 34
13 | exist
14 | read
15 | 45
16 | type
17 | islam
18 | ye
19 | kind
20 | common
21 | Cluster 1:
22 | 13999823065
23 | 7928
24 | 1559282938
25 | chat_234083222513
26 | impk
27 | Run | Python Console
28 | IDE and Plugin Updates: PyCharm is ready to update. (22-Feb-20 8:03 PM)
29 | 18:26 LF UTF-8 4 spaces Python 3.8
30 | Type here to search
31 | Eικόνα 59.

```

Εικόνα 59.

Και τέλος το τελευταίο στοιχείο που μας δίνει το output είναι η πληροφορία για το πόσα μηνύματα από κάθε group chat ανήκουν σε καθένα ένα από τα clusters (Εικόνα 60.).

```

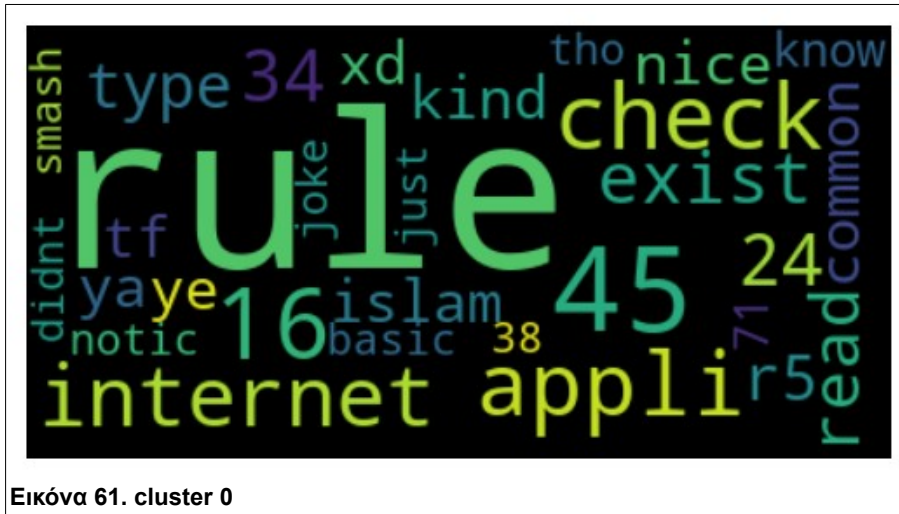
Clusters of chat 1 / 280 ( C:/Users/Geo/Desktop/george/Meibhatsapp-Wrappers-master/sample/chat_12053466425-1552762183@g.us.chat.log ): {2: 544, 5: 29, 13: 3, 7: 7, 4: 1}
Clusters of chat 2 / 280 ( C:/Users/Geo/Desktop/george/Meibhatsapp-Wrappers-master/sample/chat_12064721374-1552734205@g.us.chat.log ): {2: 127, 13: 1, 3: 1}
Clusters of chat 3 / 280 ( C:/Users/Geo/Desktop/george/Meibhatsapp-Wrappers-master/sample/chat_12533139925-1533654308@g.us.chat.log ): {2: 29, 4: 1}
Clusters of chat 4 / 280 ( C:/Users/Geo/Desktop/george/Meibhatsapp-Wrappers-master/sample/chat_13159493958-1518354992@g.us.chat.log ): {2: 1149, 4: 25, 10: 1, 3: 1, 13: 2}
Clusters of chat 5 / 280 ( C:/Users/Geo/Desktop/george/Meibhatsapp-Wrappers-master/sample/chat_13159493958-1518355212@g.us.chat.log ): {2: 391, 7: 42, 5: 41, 4: 4}
Clusters of chat 6 / 280 ( C:/Users/Geo/Desktop/george/Meibhatsapp-Wrappers-master/sample/chat_13159933205-1542629155@g.us.chat.log ): {2: 603, 4: 4}
Clusters of chat 7 / 280 ( C:/Users/Geo/Desktop/george/Meibhatsapp-Wrappers-master/sample/chat_13317012408-1506449192@g.us.chat.log ): {2: 180, 5: 50, 4: 8, 7: 13}
Clusters of chat 8 / 280 ( C:/Users/Geo/Desktop/george/Meibhatsapp-Wrappers-master/sample/chat_13317012408-1506462608@g.us.chat.log ): {2: 457, 5: 12, 4: 16, 10: 1}
Clusters of chat 9 / 280 ( C:/Users/Geo/Desktop/george/Meibhatsapp-Wrappers-master/sample/chat_14087211456-1539971560@g.us.chat.log ): {2: 321, 0: 1, 4: 1}
Clusters of chat 10 / 280 ( C:/Users/Geo/Desktop/george/Meibhatsapp-Wrappers-master/sample/chat_14803950754-1529862422@g.us.chat.log ): {2: 1}
Clusters of chat 11 / 280 ( C:/Users/Geo/Desktop/george/Meibhatsapp-Wrappers-master/sample/chat_17122506796-1545616164@g.us.chat.log ): {2: 101}
Clusters of chat 12 / 280 ( C:/Users/Geo/Desktop/george/Meibhatsapp-Wrappers-master/sample/chat_17753823154-1545637451@g.us.chat.log ): {2: 415, 8: 4, 7: 86, 5: 92}
Clusters of chat 13 / 280 ( C:/Users/Geo/Desktop/george/Meibhatsapp-Wrappers-master/sample/chat_17863435548-1534713807@g.us.chat.log ): {}
Clusters of chat 14 / 280 ( C:/Users/Geo/Desktop/george/Meibhatsapp-Wrappers-master/sample/chat_19793987619@g.us.chat.log ): {}
Clusters of chat 15 / 280 ( C:/Users/Geo/Desktop/george/Meibhatsapp-Wrappers-master/sample/chat_201018051555-1493575339@g.us.chat.log ): {2: 36}
Clusters of chat 16 / 280 ( C:/Users/Geo/Desktop/george/Meibhatsapp-Wrappers-master/sample/chat_212639422141-1477567255@g.us.chat.log ): {2: 422, 4: 5, 5: 78, 7: 86}
Clusters of chat 17 / 280 ( C:/Users/Geo/Desktop/george/Meibhatsapp-Wrappers-master/sample/chat_212657458806-1520152237@g.us.chat.log ): {4: 66, 2: 653, 13: 2, 5: 87, 3: 1, 7: 60}
Clusters of chat 18 / 280 ( C:/Users/Geo/Desktop/george/Meibhatsapp-Wrappers-master/sample/chat_212707719479-1552688726@g.us.chat.log ): {2: 128, 4: 4}
Clusters of chat 19 / 280 ( C:/Users/Geo/Desktop/george/Meibhatsapp-Wrappers-master/sample/chat_22896523148@g.us.chat.log ): {}
Clusters of chat 20 / 280 ( C:/Users/Geo/Desktop/george/Meibhatsapp-Wrappers-master/sample/chat_22965141143@g.us.chat.log ): {}
Clusters of chat 21 / 280 ( C:/Users/Geo/Desktop/george/Meibhatsapp-Wrappers-master/sample/chat_233544217390-1498962626@g.us.chat.log ): {2: 980, 4: 11, 9: 1, 5: 9, 7: 10, 10: 1}
Clusters of chat 22 / 280 ( C:/Users/Geo/Desktop/george/Meibhatsapp-Wrappers-master/sample/chat_2347033838665-1563313466@g.us.chat.log ): {2: 689, 13: 6, 14: 2, 3: 12}
Clusters of chat 23 / 280 ( C:/Users/Geo/Desktop/george/Meibhatsapp-Wrappers-master/sample/chat_2347034328548-1491921470@g.us.chat.log ): {2: 162, 4: 5}
Clusters of chat 24 / 280 ( C:/Users/Geo/Desktop/george/Meibhatsapp-Wrappers-master/sample/chat_234709463201-1510612408@g.us.chat.log ): {2: 147}

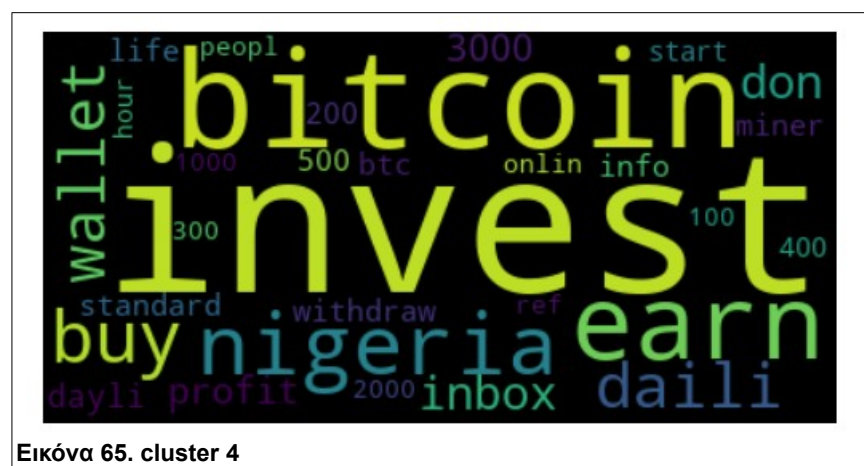
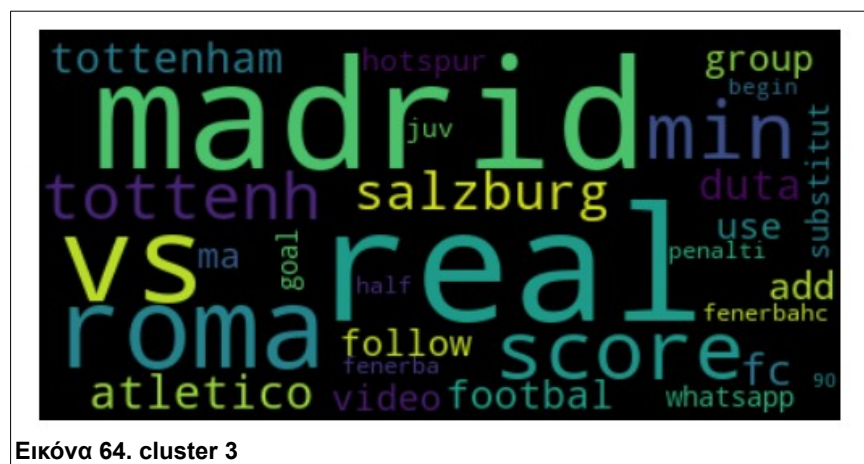
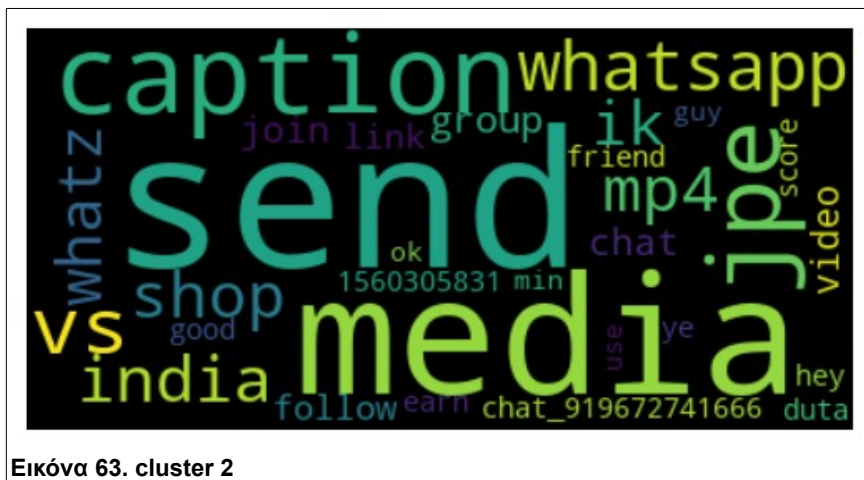
```

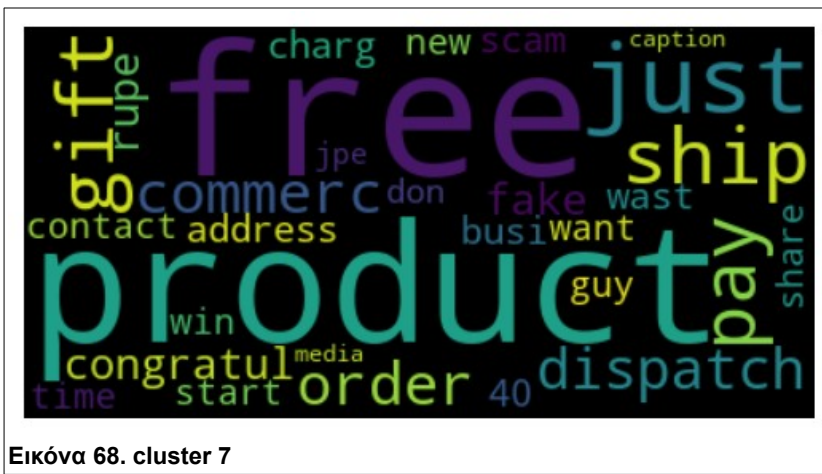
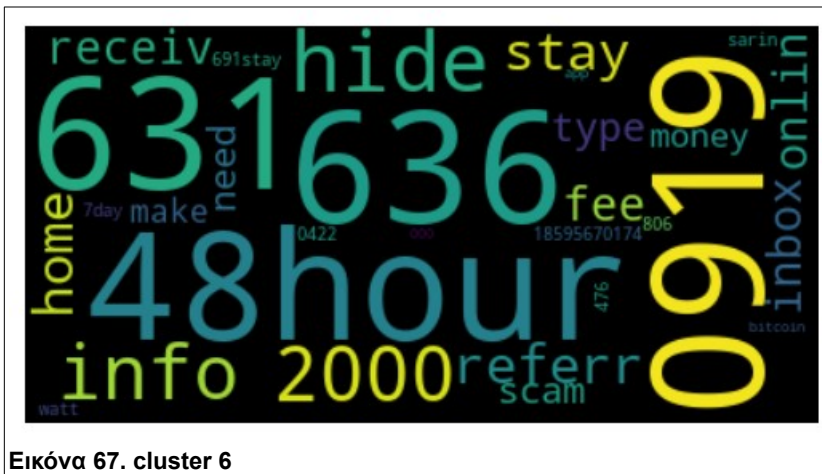
Εικόνα 60.

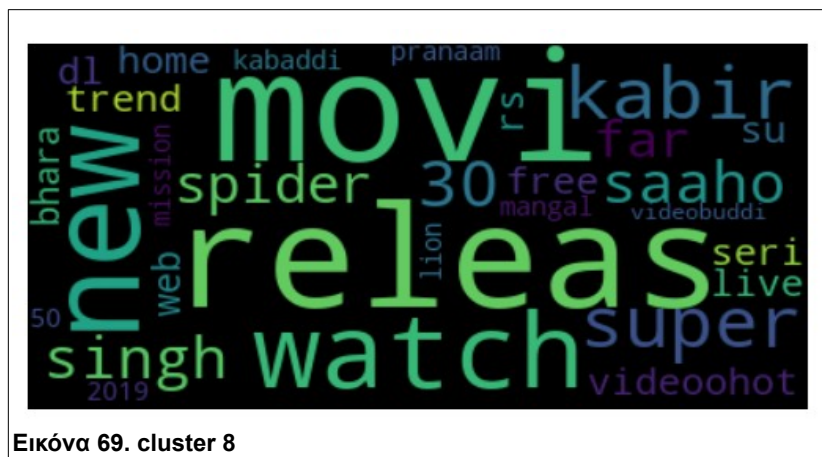
Όπως και στην περίπτωση του Telegram, με τον ίδιο ακριβώς τρόπο διαβάζουμε και το output στην περίπτωση του WhatsApp. Για παράδειγμα από το output (**chat_12053466425-1552762183@g.us.chat.log**): **{12: 42, 14: 19, 2: 457, 11: 5, 10: 8, 9: 29, 1: 19, 13: 4, 8: 1}** βλέπουμε ότι στο 1ο από τα 280 συνολικά chats, 42 μηνύματα ανήκουν στο cluster 12, 19 μηνύματα ανήκουν στο cluster 14, 457 μηνύματα ανήκουν στο cluster 2, 5 μηνύματα στο cluster 11, 8 μηνύματα ανήκουν στο cluster 10, 29 μηνύματα ανήκουν στο cluster 9, 19 μηνύματα ανήκουν στο cluster 1, 4 μηνύματα ανήκουν στο cluster 13 και 1 μήνυμα στο cluster 8 κ.ο.κ. Αναλυτικός πίνακας με όλα τα αποτελέσματα παρατίθεται στον Πίνακα 2.

Αφού τελειώσει η όλη διαδικασία παίρνουμε τις εικόνες με τα wordclouds τα οποία είναι αποθηκευμένα στον φάκελο images/Kmeans. Οι εικόνες αυτές που παρατίθεται μας δίνουν τα top terms για το κάθε cluster και από τα top terms μπορούμε να βγάλουμε ένα συμπέρασμα για το τι είδους μηνύματα περιέχει το κάθε chat.





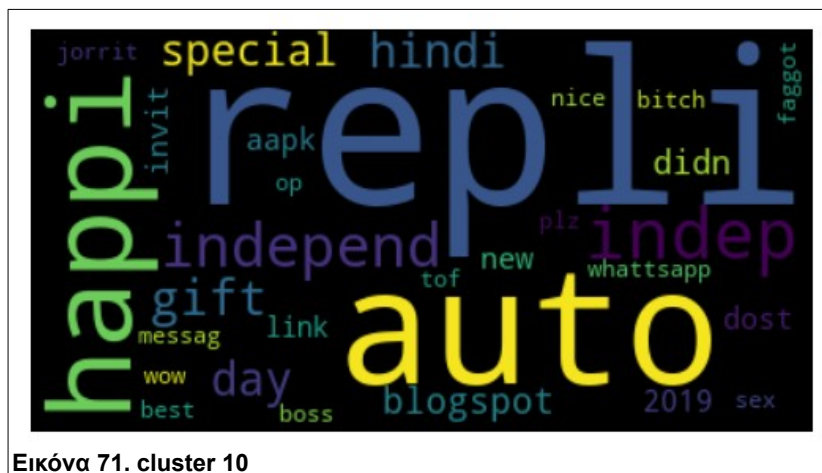




Εικόνα 69. cluster 8



Εικόνα 70. cluster 9



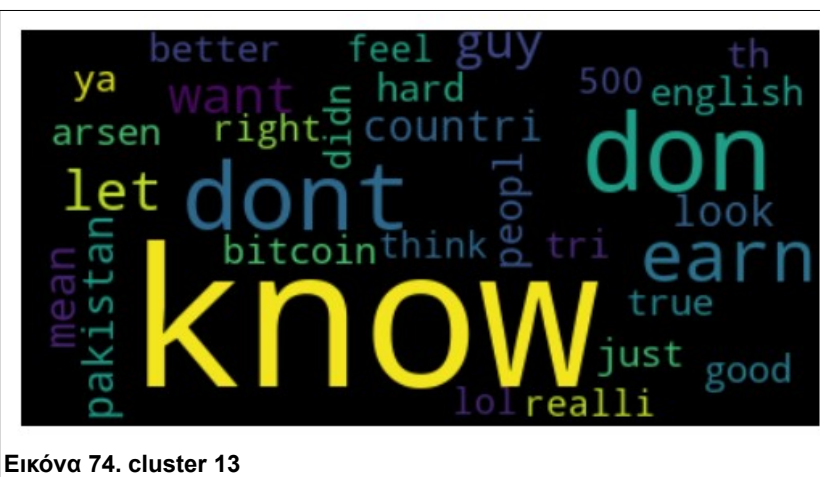
Εικόνα 71. cluster 10



Εικόνα 72. cluster 11



Εικόνα 73. cluster 12



Εικόνα 74. cluster 13


```
chat_19793987619@c.us.chat.log ): {}
Clusters of chat 15 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_201018051555-1493575339@g.us.chat.log ): {2: 32, 12: 1, 13: 3}
Clusters of chat 16 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_212639422141-1477567255@g.us.chat.log ): {2: 401, 11: 2, 12: 23, 14: 3, 9: 78, 1: 84}
Clusters of chat 17 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_212657458806-1520152237@g.us.chat.log ): {2: 616, 12: 23, 8: 4, 10: 2, 4: 2, 14: 13, 1: 121, 9: 87,
11: 1}
Clusters of chat 18 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_212707719479-1552688726@g.us.chat.log ): {2: 128, 14: 2, 8: 1, 12: 1}
Clusters of chat 19 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_22896523148@c.us.chat.log ): {}
Clusters of chat 20 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_22965141143@c.us.chat.log ): {}
Clusters of chat 21 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_233544217390-1490962626@g.us.chat.log ): {2: 779, 12: 15, 10: 177, 14: 15, 11: 4, 9: 9, 1: 10, 13: 2,
4: 1}
Clusters of chat 22 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2347033038665-1563313466@g.us.chat.log ): {2: 678, 12: 5, 14: 12, 10: 2, 11: 4, 13: 6, 4: 2}
Clusters of chat 23 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2347034328548-1491921470@g.us.chat.log ): {2: 161, 14: 5, 12: 1}
Clusters of chat 24 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2347080462391-1519061348@g.us.chat.log ): {2: 140, 12: 6, 14: 1}
Clusters of chat 25 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348024790559-1559600306@g.us.chat.log ): {2: 186, 12: 1, 14: 1, 11: 1}
Clusters of chat 26 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348032322513-1559282930@g.us.chat.log ): {2: 512}
Clusters of chat 27 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348033488384-1525616235@g.us.chat.log ): {2: 330, 9: 4, 1: 4, 12: 5, 14: 3}
Clusters of chat 28 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348054135475-1563140992@g.us.chat.log ): {2: 172, 12: 1, 4: 1, 11: 1}
Clusters of chat 29 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348062137437-1561882666@g.us.chat.log ): {2: 139, 11: 2, 4: 1}
Clusters of chat 30 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348063025256-1559096470@g.us.chat.log ): {2: 9, 12: 1}
Clusters of chat 31 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348065669534-1541854404@g.us.chat.log ): {2: 652, 9: 40, 1: 29, 12: 3}
Clusters of chat 32 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348109890004-1563317908@g.us.chat.log ): {2: 127, 14: 1, 11: 1, 4: 2, 8: 2}
Clusters of chat 33 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348110107300-1543757473@g.us.chat.log ): {2: 126}
Clusters of chat 34 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348115506965-1549270749@g.us.chat.log ): {2: 56, 14: 1}
Clusters of chat 35 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348145686926-1564015604@g.us.chat.log ): {2: 13}
Clusters of chat 36 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348147250000-1553249631@g.us.chat.log ): {2: 41, 9: 59, 1: 34, 12: 1}
Clusters of chat 37 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348149460003-1547065630@g.us.chat.log ): {2: 149, 13: 1, 12: 4, 14: 5}
```

```

Clusters of chat 38 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348155594783-1563117947@g.us.chat.log ): {14: 2, 12: 3, 2: 1738, 7: 20, 11: 4}
Clusters of chat 39 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348160741717-1468963101@g.us.chat.log ): {2: 115}
Clusters of chat 40 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348189583114-1517124754@g.us.chat.log ): {2: 91, 11: 1}
Clusters of chat 41 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2349079721832@c.us.chat.log ): {}
Clusters of chat 42 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2349086550164-1556835426@g.us.chat.log ): {2: 334, 14: 2, 11: 2, 12: 1}
Clusters of chat 43 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2349098643869-1562995313@g.us.chat.log ): {2: 2712, 12: 89, 14: 49, 10: 18, 13: 35, 4: 21, 11: 34,
6: 6}
Clusters of chat 44 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_23754804184-1517662089@g.us.chat.log ): {2: 701, 14: 31, 12: 47, 3: 27, 1: 28, 9: 27, 11: 7, 13: 9, 4:
2, 6: 3, 10: 2, 8: 1}
Clusters of chat 45 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_237651690064-1552403045@g.us.chat.log ): {2: 385, 14: 7, 12: 27, 11: 2, 10: 2}
Clusters of chat 46 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_254706773827-1534684758@g.us.chat.log ): {2: 279, 8: 1, 13: 1, 14: 2, 10: 2, 9: 4, 1: 6, 12: 7}
Clusters of chat 47 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_254707936002-1541854957@g.us.chat.log ): {2: 51, 12: 1, 1: 6, 9: 6}
Clusters of chat 48 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_254757146467-1563438096@g.us.chat.log ): {2: 183, 10: 1, 14: 2}
Clusters of chat 49 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_254798513186-1558264650@g.us.chat.log ): {2: 457, 12: 22, 14: 5, 11: 1, 10: 2, 13: 2}
Clusters of chat 50 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_255620150767-1527782404@g.us.chat.log ): {2: 35}
Clusters of chat 51 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_255717713926-1562782840@g.us.chat.log ): {2: 552, 11: 9, 5: 60, 12: 6, 13: 1, 10: 1}
Clusters of chat 52 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_260977284225-1550949072@g.us.chat.log ): {12: 5, 2: 91, 9: 32, 1: 20, 14: 1}
Clusters of chat 53 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_263772749518-1460358881@g.us.chat.log ): {2: 1440, 4: 8, 11: 15, 13: 4}
Clusters of chat 54 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_263778888125-1488105200@g.us.chat.log ): {2: 1297, 7: 20}
Clusters of chat 55 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_27621634355-1529677872@g.us.chat.log ): {2: 77, 14: 2, 12: 3}
Clusters of chat 56 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_27644635125-1509175015@g.us.chat.log ): {2: 14}
Clusters of chat 57 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_31621943195-1557752118@g.us.chat.log ): {2: 12949, 11: 28, 0: 5, 4: 8, 13: 11, 10: 8, 5: 1, 8: 3, 12:
1}
Clusters of chat 58 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_353862228246-1543459639@g.us.chat.log ): {2: 113, 11: 1}
Clusters of chat 59 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_353862228246-1543580117@g.us.chat.log ): {}
Clusters of chat 60 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_353862228246-1544539545@g.us.chat.log ): {2: 137, 13: 5, 11: 1}

```



```

Clusters of chat 61 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_353862228246-1544601100@g.us.chat.log ): {2: 120}
Clusters of chat 62 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_447448867845-1532456432@g.us.chat.log ): {2: 390, 13: 3, 12: 2, 11: 11, 10: 2, 4: 4, 8: 2}
Clusters of chat 63 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_447451246621-1564038749@g.us.chat.log ): {10: 1, 2: 152, 12: 11, 14: 1}
Clusters of chat 64 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_447708329340-1560552735@g.us.chat.log ): {2: 2541, 7: 21, 10: 2, 11: 5}
Clusters of chat 65 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_447887810701-1493490356@g.us.chat.log ): {2: 10}
Clusters of chat 66 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_4917629460203-1551197253@g.us.chat.log ): {2: 517, 5: 1, 12: 83, 14: 20, 13: 5, 10: 4, 6: 1}
Clusters of chat 67 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_50762850233-1508010876@g.us.chat.log ): {2: 140, 10: 1, 8: 1, 14: 1}
Clusters of chat 68 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_5492914734432-1550429334@g.us.chat.log ): {2: 301, 8: 43, 14: 2, 1: 113, 9: 113}
Clusters of chat 69 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_5512996817977-1513419479@g.us.chat.log ): {8: 46, 2: 328, 12: 5, 13: 1, 14: 3, 11: 1}
Clusters of chat 70 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_5521996180760-1540595805@g.us.chat.log ): {2: 90, 12: 6}
Clusters of chat 71 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_553497819767-1480261843@g.us.chat.log ): {2: 309, 14: 3, 12: 5, 1: 108, 9: 110, 5: 22}
Clusters of chat 72 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_556295153706-1483193085@g.us.chat.log ): {2: 314, 12: 11, 14: 4, 11: 30, 13: 2, 10: 2}
Clusters of chat 73 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_601126316671-1507101918@g.us.chat.log ): {12: 57, 2: 840, 10: 8, 14: 44, 6: 6, 13: 13, 11: 3, 4: 2, 8:
1}
Clusters of chat 74 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_6738959828-1563623700@g.us.chat.log ): {2: 12097, 11: 160, 12: 217, 10: 66, 13: 159, 14: 105, 4:
124, 8: 7, 6: 5, 5: 2}
Clusters of chat 75 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_85268553604@c.us.chat.log ): {}
Clusters of chat 76 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_8613136156865-1500422552@g.us.chat.log ): {2: 164}
Clusters of chat 77 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_8615099927791-1521182088@g.us.chat.log ): {2: 2}
Clusters of chat 78 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_8801744419079-1509171830@g.us.chat.log ): {2: 911, 13: 15, 11: 12, 12: 56, 14: 49, 9: 67, 1: 43, 8:
1, 6: 5, 4: 3, 10: 5}
Clusters of chat 79 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_8801812755493-1505678107@g.us.chat.log ): {2: 187, 1: 14, 9: 15, 12: 6, 11: 1, 14: 1, 8: 2, 4: 1}
Clusters of chat 80 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_916207231114-1544876477@g.us.chat.log ): {12: 9, 2: 231, 14: 4, 13: 1}
Clusters of chat 81 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_917011284590-1486113958@g.us.chat.log ): {2: 80}
Clusters of chat 82 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_917016354308-1488093290@g.us.chat.log ): {2: 440, 12: 3, 14: 2, 1: 93, 9: 70, 11: 1}
Clusters of chat 83 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_917058803589-1528353496@g.us.chat.log ): {2: 354}

```

```

Clusters of chat 84 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_917066389716@c.us.chat.log ): {2: 3}
Clusters of chat 85 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_917096666815-1563845218@g.us.chat.log ): {2: 21, 6: 3}
Clusters of chat 86 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_917340270078-1563980314@g.us.chat.log ): {2: 1}
Clusters of chat 87 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_917504961079-1519030093@g.us.chat.log ): {2: 451, 12: 15, 1: 61, 9: 36, 11: 2, 8: 5, 10: 1, 14: 2}
Clusters of chat 88 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_917504961079-1519136340@g.us.chat.log ): {2: 64, 9: 63, 1: 36}
Clusters of chat 89 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_917504961079-1519185392@g.us.chat.log ): {2: 306, 14: 4, 12: 3, 11: 5, 1: 4, 9: 4}
Clusters of chat 90 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_917504961079-1519185435@g.us.chat.log ): {2: 322, 12: 7, 1: 8, 9: 7, 11: 2, 14: 1}
Clusters of chat 91 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_917896159081-1506248485@g.us.chat.log ): {2: 122, 12: 6, 14: 6, 10: 3, 8: 3, 4: 1, 11: 1}
Clusters of chat 92 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_917900186560-1550952626@g.us.chat.log ): {2: 566, 12: 46, 14: 14, 10: 5, 6: 1, 13: 7}
Clusters of chat 93 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_917985455537-1488972204@g.us.chat.log ): {2: 402, 12: 18, 10: 2, 14: 4, 8: 1, 13: 4, 5: 1}
Clusters of chat 94 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_918011717118-1513346963@g.us.chat.log ): {12: 20, 2: 858, 14: 11, 10: 2, 6: 1, 1: 53, 9: 50, 11: 6}
Clusters of chat 95 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_918074820605@c.us.chat.log ): {}
Clusters of chat 96 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_918112269746-1515737968@g.us.chat.log ): {}
Clusters of chat 97 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_918154995870-1558890931@g.us.chat.log ): {2: 38}
Clusters of chat 98 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_918178289603-1561741836@g.us.chat.log ): {2: 416, 12: 2, 9: 75, 1: 70}
Clusters of chat 99 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_918197180324-1564038241@g.us.chat.log ): {2: 55}
Clusters of chat 100 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_918197180324-1564039245@g.us.chat.log ): {2: 65, 12: 2}
Clusters of chat 101 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_918507097725-1557250497@g.us.chat.log ): {2: 299, 11: 1, 10: 7, 14: 3, 12: 2}
Clusters of chat 102 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_918795192508-1483164938@g.us.chat.log ): {2: 597, 12: 14, 14: 7, 13: 1, 10: 1, 8: 2, 1: 15, 9: 11, 11:
1, 6: 2}
Clusters of chat 103 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_918795744594-1541928672@g.us.chat.log ): {2: 414, 12: 15, 8: 1, 1: 42, 9: 39, 14: 5, 10: 2, 5: 8}
Clusters of chat 104 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_918826310112-1553089301@g.us.chat.log ): {2: 403, 14: 7, 12: 19, 10: 5, 11: 1, 1: 2, 9: 3, 4: 2}
Clusters of chat 105 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_918850732033-1560277858@g.us.chat.log ): {2: 522, 12: 21, 11: 2, 14: 12, 8: 1, 9: 6, 1: 6, 10: 2}
Clusters of chat 106 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_918850823474-1551458263@g.us.chat.log ): {2: 99, 11: 1, 9: 13, 1: 13, 12: 1}
Clusters of chat 107 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_918866126299-1502878641@g.us.chat.log ): {14: 16, 2: 729, 12: 15, 8: 78, 1: 69, 9: 47, 13: 2}

```

```

Clusters of chat 108 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_918874547144-1539766758@g.us.chat.log ): {2: 602, 5: 10, 12: 8}
Clusters of chat 109 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_918903675398-1525952582@g.us.chat.log ): {2: 511, 12: 27, 14: 12, 10: 1, 5: 4, 11: 5, 13: 2, 4: 1}
Clusters of chat 110 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_918956783671-1552825100@g.us.chat.log ): {2: 831, 9: 111, 1: 98, 12: 13, 11: 28, 14: 5, 13: 2, 6: 1,
8: 32}
Clusters of chat 111 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_919102711964-1562017267@g.us.chat.log ): {2: 156, 12: 1}
Clusters of chat 112 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_919137566247-1548228361@g.us.chat.log ): {2: 450}
Clusters of chat 113 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_919321629786-1491165067@g.us.chat.log ): {12: 19, 2: 521, 14: 2, 11: 2}
Clusters of chat 114 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_919504171173-1521911266@g.us.chat.log ): {2: 869, 12: 21, 9: 3, 1: 2, 14: 6, 13: 3, 10: 1}
Clusters of chat 115 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_919567635886-1544363903@g.us.chat.log ): {2: 144, 1: 18, 9: 15, 14: 5, 12: 15}
Clusters of chat 116 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_919623418138-1490536344@g.us.chat.log ): {2: 283, 12: 11, 14: 4, 10: 1, 5: 1}
Clusters of chat 117 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_919672741666-1560305831@g.us.chat.log ): {2: 1324}
Clusters of chat 118 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_919737834397-1560511400@g.us.chat.log ): {2: 119, 12: 3, 11: 5}
Clusters of chat 119 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_919828556027-1547698927@g.us.chat.log ): {2: 75}
Clusters of chat 120 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_919860598106-1522141649@g.us.chat.log ): {2: 287, 12: 6, 11: 1, 14: 1, 4: 1, 1: 18, 9: 17}
Clusters of chat 121 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_923056474175@c.us.chat.log ): {}
Clusters of chat 122 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_923056733587-1508906176@g.us.chat.log ): {1: 36, 9: 63, 2: 585, 14: 20, 12: 31, 11: 6, 13: 1, 6: 1}
Clusters of chat 123 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_923106438145@c.us.chat.log ): {}
Clusters of chat 124 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_923129261430-1532194872@g.us.chat.log ): {2: 129, 12: 3, 8: 1, 11: 3}
Clusters of chat 125 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_923147117253-1520597965@g.us.chat.log ): {12: 22, 2: 624, 14: 12, 13: 4, 10: 2}
Clusters of chat 126 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_923147117253-1528994513@g.us.chat.log ): {2: 421, 7: 1}
Clusters of chat 127 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_923169964268-1494772094@g.us.chat.log ): {2: 203, 1: 4, 9: 4, 12: 1, 14: 1, 10: 1}
Clusters of chat 128 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_923214324132-1552761743@g.us.chat.log ): {2: 653, 9: 12, 1: 6, 14: 1}
Clusters of chat 129 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_923455024491-1564300348@g.us.chat.log ): {2: 367, 12: 8, 4: 3, 13: 12, 10: 2, 14: 2, 6: 1}
Clusters of chat 130 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_94750947291@c.us.chat.log ): {}
Clusters of chat 131 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_971528904770-1563885486@g.us.chat.log ): {2: 200, 4: 1, 11: 3, 13: 3, 8: 1, 10: 2, 14: 1}

```

```

Clusters of chat 132 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_97338381710-1487691608@g.us.chat.log ): {2: 41, 10: 2, 11: 1}
Clusters of chat 133 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_97338381710-1497431182@g.us.chat.log ): {2: 154, 14: 1, 13: 1, 10: 1, 4: 1, 11: 1}
Clusters of chat 134 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_97338381710-1500435936@g.us.chat.log ): {2: 30}
Clusters of chat 135 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_97338381710-1502880819@g.us.chat.log ): {11: 2, 2: 123, 4: 4}
Clusters of chat 136 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_97338381710-1504447471@g.us.chat.log ): {2: 21}
Clusters of chat 137 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_97338381710-1507218698@g.us.chat.log ): {2: 15}
Clusters of chat 138 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_97338381710-1511318890@g.us.chat.log ): {2: 87, 11: 1, 10: 1, 4: 1}
Clusters of chat 139 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_97338381710-1520734066@g.us.chat.log ): {2: 140, 13: 3, 11: 5, 12: 3, 10: 1, 14: 2, 8: 1}
Clusters of chat 140 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_994557753110-1563735846@g.us.chat.log ): {2: 3190, 14: 2, 11: 16, 12: 13, 5: 9, 10: 8, 13: 2}
Clusters of chat 141 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safecat_12053466425-1552762183@g.us.chat.log ): {12: 42, 14: 33, 2: 446, 10: 8, 9: 29, 6: 19, 13: 4, 11:
2, 8: 1}
Clusters of chat 142 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safecat_12064721374-1552734205@g.us.chat.log ): {2: 123, 10: 1, 12: 4, 11: 1}
Clusters of chat 143 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safecat_12533139925-1533654308@g.us.chat.log ): {2: 29, 12: 1}
Clusters of chat 144 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safecat_13159493958-1518354992@g.us.chat.log ): {2: 1116, 12: 23, 14: 23, 13: 1, 5: 1, 8: 1, 10: 11, 6: 1,
11: 1}
Clusters of chat 145 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safecat_13159493958-1518355216@g.us.chat.log ): {2: 376, 8: 1, 11: 2, 10: 7, 14: 3, 6: 46, 9: 41, 12: 2}
Clusters of chat 146 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safecat_13159933205-1542629155@g.us.chat.log ): {2: 580, 12: 6, 10: 8, 14: 11, 6: 1, 11: 1}
Clusters of chat 147 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safecat_13317012408-1506449192@g.us.chat.log ): {6: 56, 9: 45, 2: 145, 14: 2, 12: 3}
Clusters of chat 148 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safecat_13317012408-1506462608@g.us.chat.log ): {2: 421, 6: 13, 10: 13, 9: 11, 12: 11, 14: 12, 11: 3, 13:
1, 4: 1}
Clusters of chat 149 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safecat_14087211456-1539971560@g.us.chat.log ): {2: 294, 14: 5, 12: 20, 10: 1, 6: 1, 13: 2}
Clusters of chat 150 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safecat_14803950754-1529862422@g.us.chat.log ): {2: 1}
Clusters of chat 151 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safecat_17122506796-1545616164@g.us.chat.log ): {2: 90, 14: 4, 12: 6, 10: 1}
Clusters of chat 152 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safecat_17753823154-1545637451@g.us.chat.log ): {2: 406, 14: 7, 11: 2, 8: 1, 5: 4, 6: 85, 9: 92}
Clusters of chat 153 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safecat_17863435548-1534713807@g.us.chat.log ): {}
Clusters of chat 154 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safecat_19793987619@c.us.chat.log ): {}

```

```

Clusters of chat 155 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_201018051555-1493575339@g.us.chat.log ): {2: 32, 12: 1, 13: 3}
Clusters of chat 156 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_212639422141-1477567255@g.us.chat.log ): {2: 382, 10: 9, 11: 2, 12: 24, 14: 10, 8: 2, 9: 78, 6:
84}
Clusters of chat 157 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_212657458806-1520152237@g.us.chat.log ): {2: 731, 12: 28, 8: 9, 10: 10, 4: 2, 14: 33, 6: 129, 9:
92, 11: 3, 13: 4}
Clusters of chat 158 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_212707719479-1552688726@g.us.chat.log ): {2: 117, 14: 9, 8: 1, 12: 3, 10: 2}
Clusters of chat 159 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_22896523148@c.us.chat.log ): {}
Clusters of chat 160 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_22965141143@c.us.chat.log ): {}
Clusters of chat 161 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_233544217390-1490962626@g.us.chat.log ): {2: 769, 12: 16, 14: 23, 10: 178, 11: 4, 9: 9, 6: 10,
13: 2, 4: 1}
Clusters of chat 162 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2347033038665-1563313466@g.us.chat.log ): {2: 671, 12: 5, 14: 12, 10: 12, 11: 2, 13: 7}
Clusters of chat 163 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2347034328548-1491921470@g.us.chat.log ): {2: 160, 14: 6, 12: 1}
Clusters of chat 164 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2347080462391-1519061348@g.us.chat.log ): {2: 136, 12: 6, 14: 4, 10: 1}
Clusters of chat 165 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2348024790559-1559600306@g.us.chat.log ): {2: 185, 12: 1, 14: 1, 11: 2}
Clusters of chat 166 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2348032322513-1559282930@g.us.chat.log ): {2: 509, 7: 1, 12: 2}
Clusters of chat 167 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2348033488384-1525616235@g.us.chat.log ): {2: 322, 14: 7, 9: 4, 6: 4, 8: 4, 12: 5}
Clusters of chat 168 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2348054135475-1563140992@g.us.chat.log ): {2: 169, 8: 4, 12: 1, 4: 1}
Clusters of chat 169 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2348062137437-1561882666@g.us.chat.log ): {2: 140, 11: 1, 4: 1}
Clusters of chat 170 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2348063025256-1559096470@g.us.chat.log ): {2: 9, 12: 1}
Clusters of chat 171 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2348065669534-1541854404@g.us.chat.log ): {2: 635, 14: 16, 9: 40, 6: 29, 12: 4}
Clusters of chat 172 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2348109890004-1563317908@g.us.chat.log ): {2: 127, 11: 2, 4: 2, 8: 2}
Clusters of chat 173 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2348110107300-1543757473@g.us.chat.log ): {2: 126}
Clusters of chat 174 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2348115506965-1549270749@g.us.chat.log ): {2: 55, 10: 1, 14: 1}
Clusters of chat 175 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2348145686926-1564015604@g.us.chat.log ): {2: 13}
Clusters of chat 176 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2348147250000-1553249631@g.us.chat.log ): {2: 40, 14: 1, 9: 59, 6: 34, 12: 1}
Clusters of chat 177 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2348149460003-1547065630@g.us.chat.log ): {2: 144, 13: 1, 12: 4, 14: 9, 10: 1}

```

```
Clusters of chat 178 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\safechat_2348155594783-1563117947@g.us.chat.log ): {14: 2, 12: 3, 2: 1742, 7: 20}
Clusters of chat 179 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\safechat_2348160741717-1468963101@g.us.chat.log ): {2: 115}
Clusters of chat 180 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\safechat_2348189583114-1517124754@g.us.chat.log ): {2: 90, 11: 1, 14: 1}
Clusters of chat 181 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\safechat_2349079721832@c.us.chat.log ): {}
Clusters of chat 182 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\safechat_2349086550164-1556835426@g.us.chat.log ): {14: 5, 2: 329, 7: 1, 12: 1, 11: 3}
Clusters of chat 183 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\safechat_2349098643869-1562995313@g.us.chat.log ): {2: 2709, 12: 90, 14: 57, 10: 20, 13: 36, 4: 20, 11: 26, 6: 6}
Clusters of chat 184 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\safechat_23754804184-1517662089@g.us.chat.log ): {2: 674, 14: 35, 12: 51, 10: 22, 6: 31, 9: 27, 11: 6, 13: 9, 8: 7, 3: 21, 4: 2}
Clusters of chat 185 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\safechat_237651690064-1552403045@g.us.chat.log ): {2: 378, 14: 14, 12: 27, 11: 1, 10: 3}
Clusters of chat 186 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\safechat_254706773827-1534684758@g.us.chat.log ): {2: 268, 10: 5, 14: 9, 8: 1, 13: 1, 12: 8, 9: 4, 6: 6}
Clusters of chat 187 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\safechat_254707936002-1541854957@g.us.chat.log ): {2: 51, 12: 1, 6: 6, 9: 6}
Clusters of chat 188 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\safechat_254757146467-1563438096@g.us.chat.log ): {2: 181, 10: 3, 14: 2}
Clusters of chat 189 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\safechat_254798513186-1558264650@g.us.chat.log ): {2: 428, 10: 4, 12: 41, 14: 13, 11: 1, 13: 2}
Clusters of chat 190 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\safechat_255620150767-1527782404@g.us.chat.log ): {2: 34, 10: 1}
Clusters of chat 191 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\safechat_255717713926-1562782840@g.us.chat.log ): {2: 510, 10: 18, 11: 8, 8: 26, 5: 60, 12: 6, 13: 1}
Clusters of chat 192 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\safechat_260977284225-1550949072@g.us.chat.log ): {12: 5, 14: 5, 2: 87, 9: 32, 6: 20}
Clusters of chat 193 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\safechat_263772749518-1460358881@g.us.chat.log ): {2: 1436, 4: 6, 11: 20, 13: 4, 8: 1}
Clusters of chat 194 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\safechat_263778888125-1488105200@g.us.chat.log ): {2: 1286, 7: 21, 12: 10}
Clusters of chat 195 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\safechat_27621634355-1529677872@g.us.chat.log ): {2: 73, 14: 5, 12: 3, 10: 1}
Clusters of chat 196 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\safechat_27644635125-1509175015@g.us.chat.log ): {2: 14}
Clusters of chat 197 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\safechat_31621943195-1557752118@g.us.chat.log ): {2: 12936, 10: 10, 11: 36, 0: 5, 4: 10, 13: 11, 5: 1, 14: 1, 8: 3, 12: 1}
Clusters of chat 198 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\safechat_353862228246-1543459639@g.us.chat.log ): {2: 114}
Clusters of chat 199 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\safechat_353862228246-1543580117@g.us.chat.log ): {}
Clusters of chat 200 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\safechat_353862228246-1544539545@g.us.chat.log ): {2: 136, 13: 4, 11: 1, 14: 1, 12: 1}
```

```

Clusters of chat 201 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_353862228246-1544601100@g.us.chat.log ): {2: 119, 8: 1}
Clusters of chat 202 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_447448867845-1532456432@g.us.chat.log ): {2: 395, 13: 3, 12: 2, 11: 8, 10: 2, 4: 3, 8: 1}
Clusters of chat 203 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_447451246621-1564038749@g.us.chat.log ): {10: 1, 2: 152, 12: 11, 14: 1}
Clusters of chat 204 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_447708329340-1560552735@g.us.chat.log ): {2: 2543, 10: 2, 14: 1, 11: 2, 7: 21}
Clusters of chat 205 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_447887810701-1493490356@g.us.chat.log ): {2: 7, 10: 3}
Clusters of chat 206 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_4917629460203-1551197253@g.us.chat.log ): {2: 494, 14: 39, 5: 1, 12: 83, 8: 1, 10: 7, 13: 5, 6: 1}
Clusters of chat 207 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_50762850233-1508010876@g.us.chat.log ): {2: 137, 10: 1, 8: 1, 14: 4}
Clusters of chat 208 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_5492914734432-1550429334@g.us.chat.log ): {14: 7, 8: 43, 2: 296, 6: 113, 9: 113}
Clusters of chat 209 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_5512996817977-1513419479@g.us.chat.log ): {8: 96, 2: 274, 14: 8, 12: 5, 13: 1}
Clusters of chat 210 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_5521996180760-1540595805@g.us.chat.log ): {14: 2, 12: 7, 2: 87}
Clusters of chat 211 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_553497819767-1480261843@g.us.chat.log ): {2: 295, 8: 5, 14: 9, 10: 2, 12: 5, 6: 108, 9: 110, 5:
23}
Clusters of chat 212 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_556295153706-1483193085@g.us.chat.log ): {2: 305, 12: 11, 8: 1, 10: 9, 14: 4, 13: 3, 11: 30}
Clusters of chat 213 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_601126316671-1507101918@g.us.chat.log ): {12: 57, 2: 817, 8: 6, 10: 19, 14: 52, 6: 6, 13: 13, 11:
2, 4: 2}
Clusters of chat 214 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_6738959828-1563623700@g.us.chat.log ): {2: 12124, 11: 122, 12: 215, 10: 72, 13: 165, 14: 107,
4: 118, 8: 11, 6: 5, 5: 3}
Clusters of chat 215 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_85268553604@c.us.chat.log ): {}
Clusters of chat 216 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_8613136156865-1500422552@g.us.chat.log ): {2: 159, 14: 4, 10: 1}
Clusters of chat 217 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_8615099927791-1521182088@g.us.chat.log ): {2: 2}
Clusters of chat 218 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_8801744419079-1509171830@g.us.chat.log ): {8: 17, 2: 869, 13: 15, 11: 9, 12: 60, 10: 22, 14: 57,
9: 67, 6: 48, 4: 3}
Clusters of chat 219 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_8801812755493-1505678107@g.us.chat.log ): {2: 184, 6: 14, 9: 15, 12: 6, 14: 3, 11: 1, 8: 3, 4: 1}
Clusters of chat 220 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_916207231114-1544876477@g.us.chat.log ): {12: 9, 2: 227, 14: 5, 4: 1, 13: 3}
Clusters of chat 221 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_917011284590-1486113958@g.us.chat.log ): {2: 72, 10: 5, 8: 2, 14: 1}
Clusters of chat 222 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_917016354308-1488093290@g.us.chat.log ): {2: 415, 12: 3, 14: 7, 10: 11, 6: 93, 9: 70, 13: 2, 7: 1,
11: 1}

```

```

Clusters of chat 223 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_917058803589-1528353496@g.us.chat.log ): {2: 354}
Clusters of chat 224 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_917066389716@c.us.chat.log ): {2: 3}
Clusters of chat 225 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_917096666815-1563845218@g.us.chat.log ): {2: 24}
Clusters of chat 226 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_917340270078-1563980314@g.us.chat.log ): {2: 1}
Clusters of chat 227 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_917504961079-1519030093@g.us.chat.log ): {14: 20, 2: 428, 12: 16, 6: 61, 9: 36, 8: 5, 10: 6, 11:
1}
Clusters of chat 228 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_917504961079-1519136340@g.us.chat.log ): {2: 64, 9: 63, 6: 36}
Clusters of chat 229 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_917504961079-1519185392@g.us.chat.log ): {2: 302, 14: 4, 12: 6, 11: 5, 6: 4, 9: 4, 10: 1}
Clusters of chat 230 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_917504961079-1519185435@g.us.chat.log ): {2: 321, 12: 7, 6: 8, 9: 7, 11: 2, 10: 1, 14: 1}
Clusters of chat 231 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_917896159081-1506248485@g.us.chat.log ): {2: 121, 14: 8, 12: 6, 10: 3, 8: 3, 11: 1}
Clusters of chat 232 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_917900186560-1550952626@g.us.chat.log ): {2: 542, 12: 46, 14: 34, 10: 5, 8: 4, 6: 1, 13: 7}
Clusters of chat 233 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_917985455537-1488972204@g.us.chat.log ): {8: 2, 2: 384, 12: 18, 10: 7, 14: 14, 7: 2, 13: 4, 5: 1}
Clusters of chat 234 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918011717118-1513346963@g.us.chat.log ): {12: 21, 2: 799, 14: 55, 10: 5, 6: 54, 8: 10, 9: 50, 11:
7}
Clusters of chat 235 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918074820605@c.us.chat.log ): {}
Clusters of chat 236 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918112269746-1515737968@g.us.chat.log ): {}
Clusters of chat 237 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918154995870-1558890931@g.us.chat.log ): {2: 38}
Clusters of chat 238 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918178289603-1561741836@g.us.chat.log ): {2: 409, 10: 2, 12: 2, 8: 4, 14: 1, 9: 75, 6: 70}
Clusters of chat 239 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918197180324-1564038241@g.us.chat.log ): {2: 54, 8: 1}
Clusters of chat 240 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918197180324-1564039245@g.us.chat.log ): {2: 65, 12: 2}
Clusters of chat 241 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918507097725-1557250497@g.us.chat.log ): {2: 300, 10: 7, 14: 3, 12: 2}
Clusters of chat 242 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918795192508-1483164938@g.us.chat.log ): {2: 563, 12: 15, 8: 3, 14: 15, 10: 18, 13: 2, 6: 17, 9:
11, 11: 1}
Clusters of chat 243 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918795744594-1541928672@g.us.chat.log ): {2: 389, 12: 15, 8: 18, 14: 8, 6: 42, 9: 39, 10: 7, 5: 8}
Clusters of chat 244 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918826310112-1553089301@g.us.chat.log ): {2: 387, 14: 22, 12: 19, 10: 8, 11: 1, 6: 2, 9: 3}
Clusters of chat 245 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918850732033-1560277858@g.us.chat.log ): {2: 504, 12: 24, 11: 2, 14: 16, 5: 1, 8: 1, 10: 12, 9: 6,

```



```

6: 6}
Clusters of chat 246 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918850823474-1551458263@g.us.chat.log ): {2: 96, 11: 1, 14: 1, 10: 1, 9: 13, 6: 13, 8: 1, 12: 1}
Clusters of chat 247 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918866126299-1502878641@g.us.chat.log ): {14: 26, 2: 715, 12: 16, 8: 79, 6: 69, 9: 47, 13: 2, 10:
2}
Clusters of chat 248 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918874547144-1539766758@g.us.chat.log ): {2: 571, 8: 22, 5: 10, 10: 8, 12: 8, 14: 1}
Clusters of chat 249 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918903675398-1525952582@g.us.chat.log ): {2: 502, 12: 28, 14: 15, 10: 5, 5: 4, 13: 2, 4: 1, 11: 2,
8: 4}
Clusters of chat 250 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918956783671-1552825100@g.us.chat.log ): {2: 770, 9: 111, 6: 99, 12: 15, 14: 9, 10: 52, 13: 5,
11: 27, 8: 33}
Clusters of chat 251 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_919102711964-1562017267@g.us.chat.log ): {2: 156, 12: 1}
Clusters of chat 252 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_919137566247-1548228361@g.us.chat.log ): {2: 450}
Clusters of chat 253 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_919321629786-1491165067@g.us.chat.log ): {12: 20, 2: 504, 10: 4, 14: 12, 4: 1, 11: 2, 8: 1}
Clusters of chat 254 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_919504171173-1521911266@g.us.chat.log ): {2: 843, 10: 17, 12: 22, 9: 3, 6: 2, 14: 14, 13: 3, 8: 1}
Clusters of chat 255 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_919567635886-1544363903@g.us.chat.log ): {2: 142, 14: 7, 6: 18, 9: 15, 12: 15}
Clusters of chat 256 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_919623418138-1490536344@g.us.chat.log ): {2: 274, 12: 11, 14: 9, 10: 3, 5: 1, 13: 2}
Clusters of chat 257 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_919672741666-1560305831@g.us.chat.log ): {2: 1324}
Clusters of chat 258 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_919737834397-1560511400@g.us.chat.log ): {2: 116, 12: 4, 8: 3, 10: 1, 11: 3}
Clusters of chat 259 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_919828556027-1547698927@g.us.chat.log ): {2: 74, 14: 1}
Clusters of chat 260 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_919860598106-1522141649@g.us.chat.log ): {2: 278, 12: 6, 11: 1, 14: 8, 4: 1, 6: 18, 9: 17, 10: 2}
Clusters of chat 261 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_923056474175@c.us.chat.log ): {}
Clusters of chat 262 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_923056733587-1508906176@g.us.chat.log ): {6: 37, 9: 58, 2: 539, 14: 72, 12: 31, 11: 5, 13: 1}
Clusters of chat 263 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_923106438145@c.us.chat.log ): {}
Clusters of chat 264 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_923129261430-1532194872@g.us.chat.log ): {2: 129, 12: 3, 8: 2, 11: 2}
Clusters of chat 265 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_923147117253-1520597965@g.us.chat.log ): {12: 23, 2: 620, 14: 13, 13: 4, 10: 4}
Clusters of chat 266 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_923147117253-1528994513@g.us.chat.log ): {2: 421, 7: 1}
Clusters of chat 267 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_923169964268-1494772094@g.us.chat.log ): {2: 199, 14: 4, 6: 4, 9: 4, 12: 1, 10: 2}
Clusters of chat 268 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\

```

```

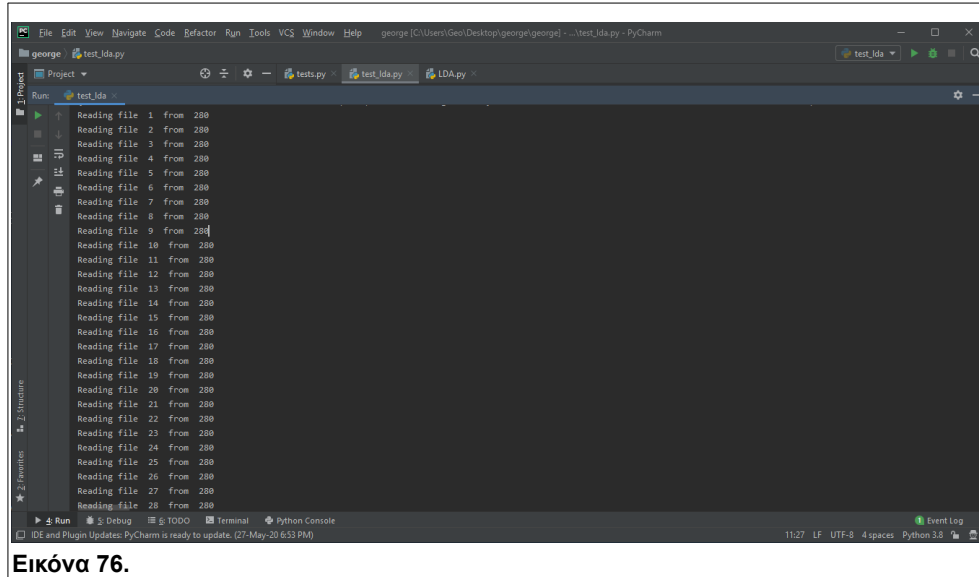
safechat_923214324132-1552761743@g.us.chat.log ): {2: 650, 9: 12, 6: 6, 14: 3, 12: 1}
Clusters of chat 269 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_923455024491-1564300348@g.us.chat.log ): {2: 364, 12: 8, 4: 4, 11: 1, 13: 14, 10: 1, 14: 2, 6: 1}
Clusters of chat 270 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_94750947291@c.us.chat.log ): {}
Clusters of chat 271 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_971528904770-1563885486@g.us.chat.log ): {2: 202, 4: 1, 11: 2, 13: 4, 10: 2}
Clusters of chat 272 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_97338381710-1487691608@g.us.chat.log ): {2: 40, 10: 2, 8: 1, 14: 1}
Clusters of chat 273 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_97338381710-1497431182@g.us.chat.log ): {2: 155, 14: 1, 13: 1, 10: 1, 11: 1}
Clusters of chat 274 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_97338381710-1500435936@g.us.chat.log ): {2: 29, 14: 1}
Clusters of chat 275 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_97338381710-1502880819@g.us.chat.log ): {2: 122, 11: 1, 13: 1, 10: 1, 12: 4}
Clusters of chat 276 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_97338381710-1504447471@g.us.chat.log ): {2: 21}
Clusters of chat 277 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_97338381710-1507218698@g.us.chat.log ): {2: 15}
Clusters of chat 278 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_97338381710-1511318890@g.us.chat.log ): {2: 88, 11: 1, 10: 1}
Clusters of chat 279 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_97338381710-1520734066@g.us.chat.log ): {2: 140, 13: 4, 11: 4, 12: 3, 10: 1, 14: 2, 8: 1}
Clusters of chat 280 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_994557753110-1563735846@g.us.chat.log ): {2: 3107, 14: 19, 8: 23, 11: 22, 12: 18, 10: 40, 5: 9,
13: 2}

```

Πίνακας 3. Τελικά αποτελέσματα WhatsApp με K-means.

6.3.2 Ανάλυση δεδομένων με LDA (Topic Extraction)

Αφού αναλύσαμε τα δεδομένα που συλλέξαμε με την βοήθεια του αλγορίθμου K-means και πήραμε τα παραπάνω αποτελέσματα του clustering, στο PyCharm τρέχουμε το script test_lda.py για να αναλύσουμε τα δεδομένα μας αυτή τη φορά με την βοήθεια του LDA κάνοντας Topic Extraction. Τρέχοντας το script test_lda.py παίρνουμε αρχικά το παρακάτω ουτρυτ, το οποίο μας λέει ότι διαβάζει τα αρχεία όπου είναι αποθηκευμένα τα μηνύματα (Εικόνα 76.).



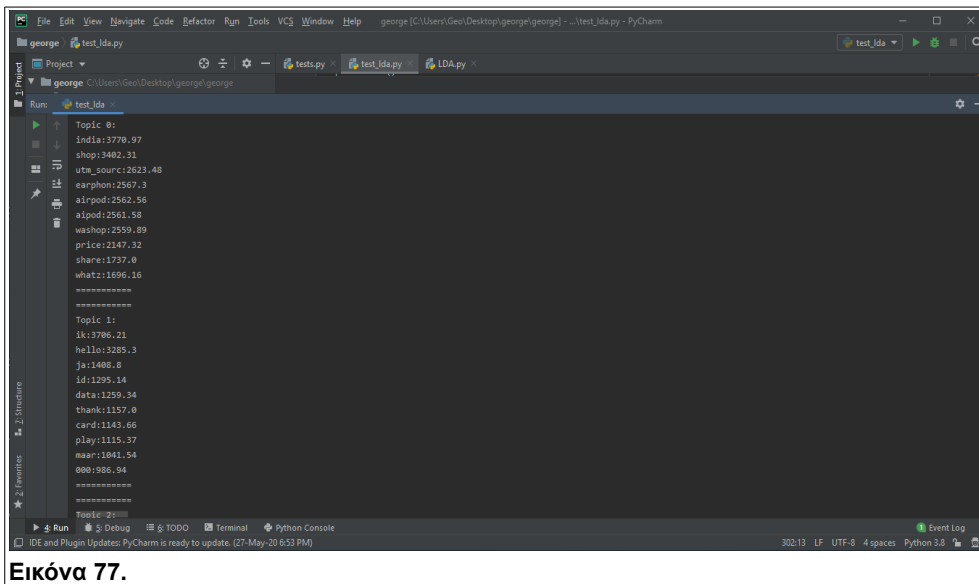
```

Run: test_lda
Reading file 1 from 280
Reading file 2 from 280
Reading file 3 from 280
Reading file 4 from 280
Reading file 5 from 280
Reading file 6 from 280
Reading file 7 from 280
Reading file 8 from 280
Reading file 9 from 280
Reading file 10 from 280
Reading file 11 from 280
Reading file 12 from 280
Reading file 13 from 280
Reading file 14 from 280
Reading file 15 from 280
Reading file 16 from 280
Reading file 17 from 280
Reading file 18 from 280
Reading file 19 from 280
Reading file 20 from 280
Reading file 21 from 280
Reading file 22 from 280
Reading file 23 from 280
Reading file 24 from 280
Reading file 25 from 280
Reading file 26 from 280
Reading file 27 from 280
Reading file 28 from 280

```

Εικόνα 76.

Στην συνέχεια του ουτρυτ βλέπουμε τα top terms για το κάθε topic (Εικόνα 77.).



```

Run: test_lda
Topic 0:
india:3770.97
shop:3482.31
utm_sourc:2623.48
earphon:2567.3
airpod:2562.56
airpod:2561.58
washop:2359.09
price:2147.32
share:1737.0
whatz:1696.16
=====
Topic 1:
ik:3786.21
hello:3285.3
ja:1488.8
id:1295.14
data:1259.34
thank:1157.0
card:1143.66
play:1115.37
maar:1041.54
000:986.94
=====
Topic 2:
=====

```

Εικόνα 77.

Και τέλος όπως και στην περίπτωση του Telegram προηγουμένως, το τελευταίο στοιχείο που μας δίνει το output είναι για το κάθε group chat το ποσοστό του κάθε topic όπου κατατάσσεται το group αυτό (Εικόνα 78.).

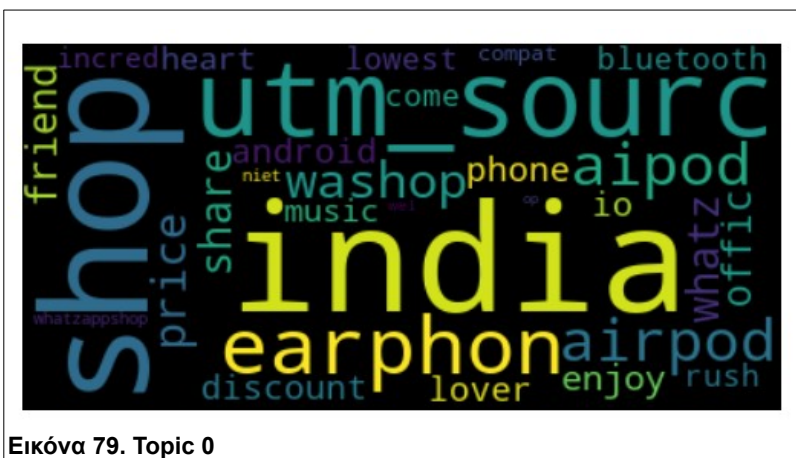
```

C:\Users\Geo\Desktop\george\george - ...\test_lda.py - PyCharm
Project: george
C:\Users\Geo\Desktop\george\george
Run: test_lda.py
Chat: 1 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_12053466425-1552762183@g.us.chat.log ):
Topics: {'topic_0': 0.0611465403702923, 'topic_1': 0.08163461123352113, 'topic_2': 0.04620933022124594, 'topic_3': 0.04440588784384422, 'topic_4': 0.0688802154169757, 'topic_5': 0.10178932571853419, 'topic_6': 0.09270922105279876, 'topic_7': 0.05054664483225688, 'topic_8': 0.1159904922881222, 'topic_9': 0.0869864898202743, 'topic_10': 0.038404435848569896, 'topic_11': 0.08061006896193679, 'topic_12': 0.1306867363916312}
Chat: 2 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_12064721374-1552734205@g.us.chat.log ):
Topics: {'topic_0': 0.03261578542496183, 'topic_1': 0.05474165914561662, 'topic_2': 0.04067744602928795, 'topic_3': 0.07285426281713209, 'topic_4': 0.0663766386582008, 'topic_5': 0.10178932571853419, 'topic_6': 0.09270922105279876, 'topic_7': 0.05054664483225688, 'topic_8': 0.1159904922881222, 'topic_9': 0.0869864898202743, 'topic_10': 0.038404435848569896, 'topic_11': 0.08061006896193679, 'topic_12': 0.1306867363916312}
Chat: 3 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_12533139925-1533654308@g.us.chat.log ):
Topics: {'topic_0': 0.048934662389024296, 'topic_1': 0.03861720454345875, 'topic_2': 0.038617204099041194, 'topic_3': 0.08640696016933598, 'topic_4': 0.07671243265450994, 'topic_5': 0.10178932571853419, 'topic_6': 0.09270922105279876, 'topic_7': 0.05054664483225688, 'topic_8': 0.1159904922881222, 'topic_9': 0.0869864898202743, 'topic_10': 0.038404435848569896, 'topic_11': 0.08061006896193679, 'topic_12': 0.1306867363916312}
Chat: 4 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_13159493958-1518354992@g.us.chat.log ):
Topics: {'topic_0': 0.04931615087702648, 'topic_1': 0.05757350582793708, 'topic_2': 0.09476918514143805, 'topic_3': 0.0655959256002862, 'topic_4': 0.06499673761273486, 'topic_5': 0.10178932571853419, 'topic_6': 0.09270922105279876, 'topic_7': 0.05054664483225688, 'topic_8': 0.1159904922881222, 'topic_9': 0.0869864898202743, 'topic_10': 0.038404435848569896, 'topic_11': 0.08061006896193679, 'topic_12': 0.1306867363916312}
Chat: 5 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_13159493958-1518355216@g.us.chat.log ):
Topics: {'topic_0': 0.13446745862019313, 'topic_1': 0.04716535764262206, 'topic_2': 0.05488521211938397, 'topic_3': 0.0561682482074918, 'topic_4': 0.04591461642363578, 'topic_5': 0.10178932571853419, 'topic_6': 0.09270922105279876, 'topic_7': 0.05054664483225688, 'topic_8': 0.1159904922881222, 'topic_9': 0.0869864898202743, 'topic_10': 0.038404435848569896, 'topic_11': 0.08061006896193679, 'topic_12': 0.1306867363916312}
Chat: 6 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_13159933205-1542629155@g.us.chat.log ):
Topics: {'topic_0': 0.04830123585741569, 'topic_1': 0.05356602449012666, 'topic_2': 0.08420757294666062, 'topic_3': 0.06548344660268428, 'topic_4': 0.05410346303789492, 'topic_5': 0.10178932571853419, 'topic_6': 0.09270922105279876, 'topic_7': 0.05054664483225688, 'topic_8': 0.1159904922881222, 'topic_9': 0.0869864898202743, 'topic_10': 0.038404435848569896, 'topic_11': 0.08061006896193679, 'topic_12': 0.1306867363916312}
Chat: 7 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_13317012408-1506449192@g.us.chat.log ):
Topics: {'topic_0': 0.2306017612823128, 'topic_1': 0.042133691610199676, 'topic_2': 0.08055241627687695, 'topic_3': 0.01098921566748787, 'topic_4': 0.01686032747918099, 'topic_5': 0.10178932571853419, 'topic_6': 0.09270922105279876, 'topic_7': 0.05054664483225688, 'topic_8': 0.1159904922881222, 'topic_9': 0.0869864898202743, 'topic_10': 0.038404435848569896, 'topic_11': 0.08061006896193679, 'topic_12': 0.1306867363916312}
Chat: 8 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_13317012408-1506462608@g.us.chat.log ):
Topics: {'topic_0': 0.04947172267976854, 'topic_1': 0.041504939576476135, 'topic_2': 0.05978621543828286, 'topic_3': 0.091941881827365, 'topic_4': 0.0384812016630209, 'topic_5': 0.10178932571853419, 'topic_6': 0.09270922105279876, 'topic_7': 0.05054664483225688, 'topic_8': 0.1159904922881222, 'topic_9': 0.0869864898202743, 'topic_10': 0.038404435848569896, 'topic_11': 0.08061006896193679, 'topic_12': 0.1306867363916312}
Chat: 9 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_14087211456-1539971560@g.us.chat.log ):
Topics: {'topic_0': 0.03605268806107613, 'topic_1': 0.051792855998344434, 'topic_2': 0.05583187381295325, 'topic_3': 0.054288751138713474, 'topic_4': 0.0696648666036194, 'topic_5': 0.10178932571853419, 'topic_6': 0.09270922105279876, 'topic_7': 0.05054664483225688, 'topic_8': 0.1159904922881222, 'topic_9': 0.0869864898202743, 'topic_10': 0.038404435848569896, 'topic_11': 0.08061006896193679, 'topic_12': 0.1306867363916312}
Chat: 10 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_14803950754-1529862422@g.us.chat.log ):
Topics: {'topic_0': 0.07692307692307694, 'topic_1': 0.07692307692307694, 'topic_2': 0.07692307692307694, 'topic_3': 0.07692307692307694, 'topic_4': 0.07692307692307694, 'topic_5': 0.10178932571853419, 'topic_6': 0.09270922105279876, 'topic_7': 0.05054664483225688, 'topic_8': 0.1159904922881222, 'topic_9': 0.0869864898202743, 'topic_10': 0.038404435848569896, 'topic_11': 0.08061006896193679, 'topic_12': 0.1306867363916312}
Chat: 11 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_17122506796-1545616164@g.us.chat.log ):
Topics: {'topic_0': 0.03250939894254178, 'topic_1': 0.07427935114553944, 'topic_2': 0.0803352724541523, 'topic_3': 0.0439397884481609, 'topic_4': 0.06625945108744292, 'topic_5': 0.10178932571853419, 'topic_6': 0.09270922105279876, 'topic_7': 0.05054664483225688, 'topic_8': 0.1159904922881222, 'topic_9': 0.0869864898202743, 'topic_10': 0.038404435848569896, 'topic_11': 0.08061006896193679, 'topic_12': 0.1306867363916312}
Chat: 12 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_17753823154-1545637451@g.us.chat.log ):
Topics: {'topic_0': 0.1621444683166072, 'topic_1': 0.05731919891975417, 'topic_2': 0.04502483322869061, 'topic_3': 0.01854920187165356, 'topic_4': 0.01594491028459936, 'topic_5': 0.10178932571853419, 'topic_6': 0.09270922105279876, 'topic_7': 0.05054664483225688, 'topic_8': 0.1159904922881222, 'topic_9': 0.0869864898202743, 'topic_10': 0.038404435848569896, 'topic_11': 0.08061006896193679, 'topic_12': 0.1306867363916312}
Chat: 13 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_17863435548-1534713807@g.us.chat.log ):
Topics: {}

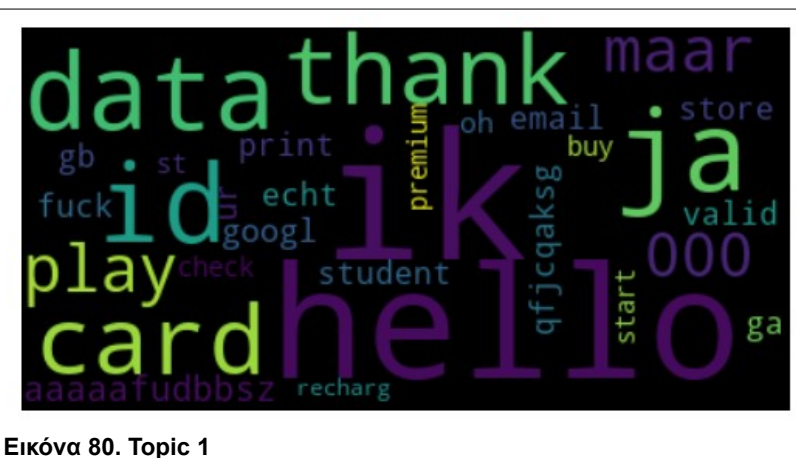
```

Εικόνα 78.

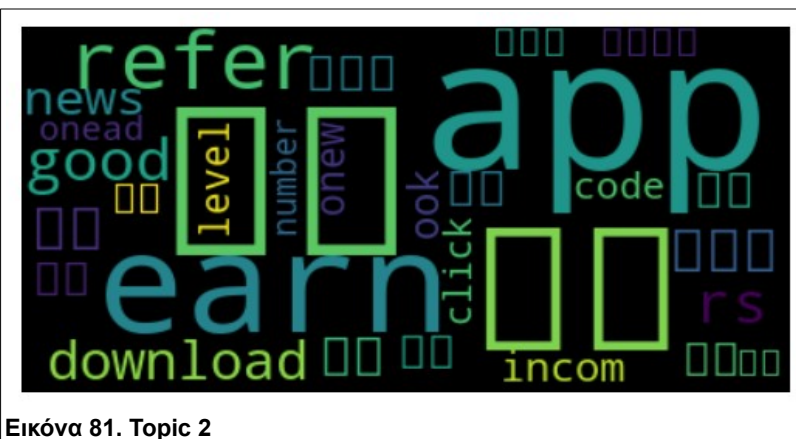
Για παράδειγμα το output “Chat: 1 / 280 (C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_12053466425-1552762183@g.us.chat.log): Topics: {'topic_0': 0.0611465403702923, 'topic_1': 0.08163461123352113, 'topic_2': 0.04620933022124594, 'topic_3': 0.04440588784384422, 'topic_4': 0.0688802154169757, 'topic_5': 0.10178932571853419, 'topic_6': 0.09270922105279876, 'topic_7': 0.05054664483225688, 'topic_8': 0.1159904922881222, 'topic_9': 0.0869864898202743, 'topic_10': 0.038404435848569896, 'topic_11': 0.08061006896193679, 'topic_12': 0.1306867363916312}” μας λέει ότι στο συγκεκριμένο chat με όνομα “chat_12053466425-1552762183@g.us.chat.log” το θέμα συζήτησης των μελών του αντιστοιχεί σε ποσοστό 6.11 % στο Topic_0, 8.16 % στο Topic_1, 4.62 % στο Topic_2, 4.44 % στο Topic_3, 6.88 % στο Topic_4, 10.17 % στο Topic_5, 9.27 % στο Topic_6, 5.05 % στο Topic_7, 11.59 % στο Topic_8, 8.69 % στο Topic_9, 3.84 % στο Topic_10, 8.06% στο Topic_11, και 13.06 στο Topic_12. Τα συνολικά αποτελέσματα τα βλέπουμε στον Πίνακα 4. που παρατίθεται, ενώ στις εικόνες 79 – 91 βλέπουμε τα wordclouds τα οποία είναι αποθηκευμένα στον φάκελο images/LDA. Στις εικόνες αυτές βλέπουμε τα top terms για το κάθε Topic και από τα top terms μπορούμε να βγάλουμε ένα συμπέρασμα για το τι είδους μηνύματα περιέχει το κάθε chat.



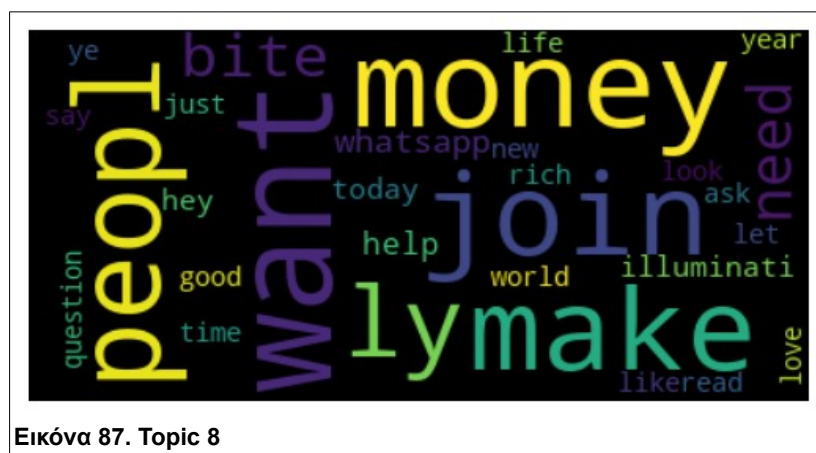
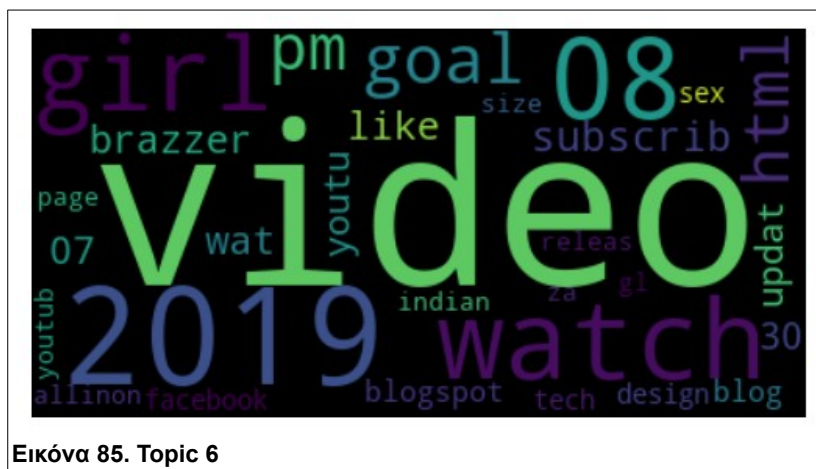
Εικόνα 79. Topic 0

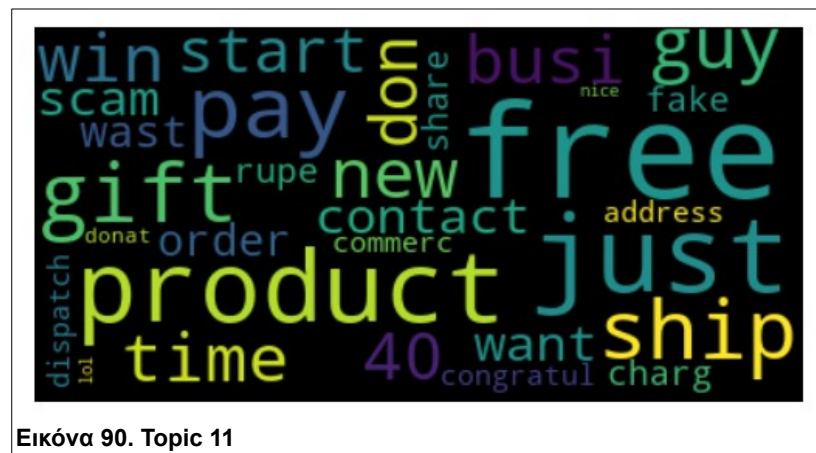
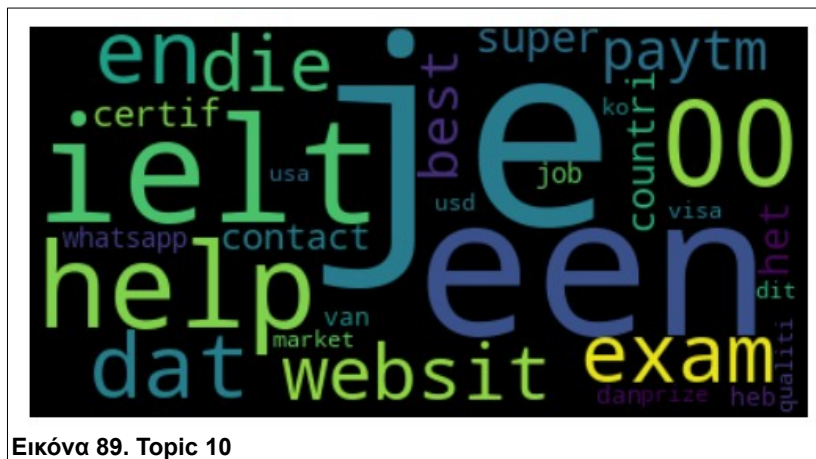
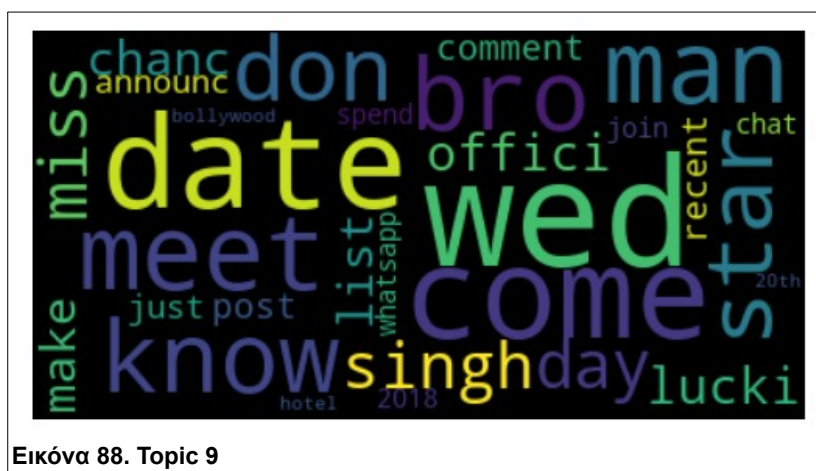


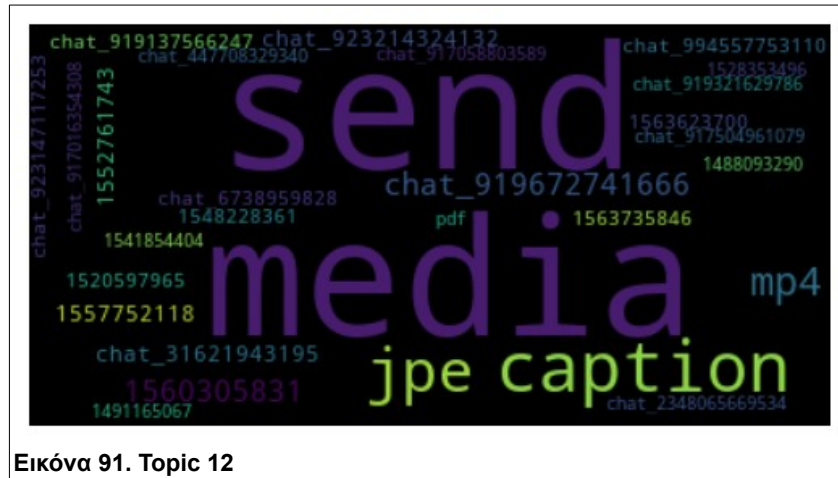
Εικόνα 80. Topic 1



Εικόνα 81. Topic 2







Εικόνα 91. Topic 12

Ακολουθεί ο Πίνακας 4.

```
Chat: 1 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_12053466425-1552762183@g.us.chat.log ):
Topics: {'topic_0': 0.0611465403702923, 'topic_1': 0.08163461123352113, 'topic_2': 0.04620933022124594, 'topic_3': 0.04440588784384422, 'topic_4': 0.0688802154169757, 'topic_5': 0.10178932571853419, 'topic_6': 0.09270922105279876, 'topic_7': 0.05054664483225688, 'topic_8': 0.1159904922881222, 'topic_9': 0.0869864898202743, 'topic_10': 0.038404435848569896, 'topic_11': 0.08061006896193679, 'topic_12': 0.1306867363916312}
Chat: 2 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_12064721374-1552734205@g.us.chat.log ):
Topics: {'topic_0': 0.03261578542496183, 'topic_1': 0.05474165914561662, 'topic_2': 0.04067744602928795, 'topic_3': 0.07285426281713209, 'topic_4': 0.06637663865582008, 'topic_5': 0.07381510395096295, 'topic_6': 0.06357561423032056, 'topic_7': 0.034097066326744176, 'topic_8': 0.07680813849175235, 'topic_9': 0.05484146223326473, 'topic_10': 0.048012353278475875, 'topic_11': 0.06762873993914723, 'topic_12': 0.31395572947651323}
Chat: 3 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_12533139925-1533654308@g.us.chat.log ):
Topics: {'topic_0': 0.048934662389024296, 'topic_1': 0.03861720454345875, 'topic_2': 0.038617204099041194, 'topic_3': 0.08640696016933598, 'topic_4': 0.07671243265450994, 'topic_5': 0.04033674944749687, 'topic_6': 0.08918272447526271, 'topic_7': 0.08119304207835425, 'topic_8': 0.10309000708851974, 'topic_9': 0.0386172246202937, 'topic_10': 0.03861720786448576, 'topic_11': 0.04161295245333022, 'topic_12': 0.2780616281168866}
Chat: 4 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_13159493958-1518354992@g.us.chat.log ):
Topics: {'topic_0': 0.04931615087702648, 'topic_1': 0.05757350582793708, 'topic_2': 0.09476810514143805, 'topic_3': 0.0655959256002862, 'topic_4': 0.06499673761273486, 'topic_5': 0.10759179930934251, 'topic_6': 0.09905463069759328, 'topic_7': 0.0850050797854946, 'topic_8': 0.11728154442691997, 'topic_9': 0.04504902337306712, 'topic_10': 0.063048188565264, 'topic_11': 0.05618448539197276, 'topic_12': 0.09453482339093518}
```

```

Chat: 5 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_13159493958-1518355216@g.us.chat.log ):
Topics: {'topic_0': 0.13446745862019313, 'topic_1': 0.04716535764262206, 'topic_2': 0.05488521211938397, 'topic_3': 0.0561682482074918, 'topic_4': 0.04591461642363578, 'topic_5': 0.05660928056257232, 'topic_6': 0.07434986021847752, 'topic_7': 0.06069271364434366, 'topic_8': 0.08049665144878052, 'topic_9': 0.11861174210575424, 'topic_10': 0.04768420403105249, 'topic_11': 0.11969008374718607, 'topic_12': 0.10326457122850846}
Chat: 6 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_13159933205-1542629155@g.us.chat.log ):
Topics: {'topic_0': 0.04830123585741569, 'topic_1': 0.05356602449012666, 'topic_2': 0.08420757294666062, 'topic_3': 0.06548344660268428, 'topic_4': 0.05410346303709492, 'topic_5': 0.09022983033107959, 'topic_6': 0.11118104905904831, 'topic_7': 0.06317330924519449, 'topic_8': 0.14999918278940957, 'topic_9': 0.043756545431962474, 'topic_10': 0.05561120731040748, 'topic_11': 0.055249575731870185, 'topic_12': 0.12513755716704966}
Chat: 7 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_13317012408-1506449192@g.us.chat.log ):
Topics: {'topic_0': 0.2306017612823128, 'topic_1': 0.042133691610199676, 'topic_2': 0.08055241627687695, 'topic_3': 0.010989215566748787, 'topic_4': 0.01686032747918099, 'topic_5': 0.1975162158381854, 'topic_6': 0.021365671889553434, 'topic_7': 0.05971164260524972, 'topic_8': 0.05693675426867049, 'topic_9': 0.20354713105906988, 'topic_10': 0.007164658718087104, 'topic_11': 0.05423582922080583, 'topic_12': 0.018384684185059195}
Chat: 8 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_13317012408-1506462608@g.us.chat.log ):
Topics: {'topic_0': 0.04947172267976854, 'topic_1': 0.041504939576476135, 'topic_2': 0.05970621543828286, 'topic_3': 0.091941881827365, 'topic_4': 0.03848212016630299, 'topic_5': 0.10835759697317024, 'topic_6': 0.0717217653949985, 'topic_7': 0.08028503542197535, 'topic_8': 0.16145599540095862, 'topic_9': 0.08302752268586835, 'topic_10': 0.04881352355572696, 'topic_11': 0.037143070690839615, 'topic_12': 0.1280886101882672}
Chat: 9 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_14087211456-1539971560@g.us.chat.log ):
Topics: {'topic_0': 0.03605268806107613, 'topic_1': 0.051792855998344434, 'topic_2': 0.05583187381295325, 'topic_3': 0.054288751138713474, 'topic_4': 0.06966486666036194, 'topic_5': 0.10332912280904936, 'topic_6': 0.06798848952624874, 'topic_7': 0.05308515225927498, 'topic_8': 0.08231677012339285, 'topic_9': 0.045027675958659835, 'topic_10': 0.04739867982987731, 'topic_11': 0.03974246480245685, 'topic_12': 0.2934806090195909}
Chat: 10 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_14803950754-1529862422@g.us.chat.log ):
Topics: {'topic_0': 0.07692307692307694, 'topic_1': 0.07692307692307694, 'topic_2': 0.07692307692307694, 'topic_3': 0.07692307692307694, 'topic_4': 0.07692307692307694, 'topic_5': 0.07692307692307694, 'topic_6': 0.07692307692307694, 'topic_7': 0.07692307692307694, 'topic_8': 0.07692307692307694, 'topic_9': 0.07692307692307694, 'topic_10': 0.07692307692307694, 'topic_11': 0.07692307692307694, 'topic_12': 0.07692307692307694}
Chat: 11 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_17122506796-1545616164@g.us.chat.log ):
Topics: {'topic_0': 0.03250939894254178, 'topic_1': 0.07427935114553944, 'topic_2': 0.0803352724541523, 'topic_3': 0.0439397884481609, 'topic_4': 0.06625945108744292, 'topic_5': 0.11067407674900197, 'topic_6': 0.14252923384348037, 'topic_7': 0.050551845384871424, 'topic_8': 0.103724304458676, 'topic_9': 0.038904895239005685, 'topic_10': 0.046082763309880405, 'topic_11': 0.08237612327510697, 'topic_12': 0.12783349566213972}

```

```

Chat: 12 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_17753823154-1545637451@g.us.chat.log ):
Topics: {'topic_0': 0.1621444683166072, 'topic_1': 0.05731919891975417, 'topic_2': 0.04502483322869061, 'topic_3': 0.01854920187165356, 'topic_4': 0.01594491028459936, 'topic_5': 0.037882766233524975, 'topic_6': 0.045132253011874034, 'topic_7': 0.041192649730444536, 'topic_8': 0.07547756131662363, 'topic_9': 0.15056851915270822, 'topic_10': 0.04005760799168528, 'topic_11': 0.15905430717894187, 'topic_12': 0.15165172276289376}
Chat: 13 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_17863435548-1534713807@g.us.chat.log ):
Topics: {}
Chat: 14 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_19793987619@c.us.chat.log ):
Topics: {}
Chat: 15 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_201018051555-1493575339@g.us.chat.log ):
Topics: {'topic_0': 0.031507949060809386, 'topic_1': 0.08849015230983832, 'topic_2': 0.031507931377784834, 'topic_3': 0.05453765450589143, 'topic_4': 0.06614797865558276, 'topic_5': 0.08920721675147775, 'topic_6': 0.04494770758858909, 'topic_7': 0.0633385368370353, 'topic_8': 0.2592693391886901, 'topic_9': 0.03465263511710767, 'topic_10': 0.0903374822235481, 'topic_11': 0.03865608573500465, 'topic_12': 0.10739933064864066}
Chat: 16 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_212639422141-1477567255@g.us.chat.log ):
Topics: {'topic_0': 0.1609834943848627, 'topic_1': 0.031038759109668686, 'topic_2': 0.08379901397169756, 'topic_3': 0.04916089950373276, 'topic_4': 0.043107488387623044, 'topic_5': 0.04931402010991389, 'topic_6': 0.044284130806240524, 'topic_7': 0.04154516348384124, 'topic_8': 0.061496219315291326, 'topic_9': 0.14153274094229734, 'topic_10': 0.030352963731859372, 'topic_11': 0.14922211924065912, 'topic_12': 0.11416298701231452}
Chat: 17 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_212657458806-1520152237@g.us.chat.log ):
Topics: {'topic_0': 0.15588087574303308, 'topic_1': 0.03196633897190194, 'topic_2': 0.04364677346720403, 'topic_3': 0.060976867100973144, 'topic_4': 0.04735975413432531, 'topic_5': 0.0791237208874368, 'topic_6': 0.032376860080156444, 'topic_7': 0.11059563026553264, 'topic_8': 0.1030809701401095, 'topic_9': 0.11063356306936681, 'topic_10': 0.061301400045566716, 'topic_11': 0.09372746268042273, 'topic_12': 0.0693297834139723}
Chat: 18 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_212707719479-1552688726@g.us.chat.log ):
Topics: {'topic_0': 0.018107619968866968, 'topic_1': 0.05006382276620659, 'topic_2': 0.0712363037491867, 'topic_3': 0.05316520144052413, 'topic_4': 0.02421916602333822, 'topic_5': 0.192586633412296, 'topic_6': 0.08683804492144266, 'topic_7': 0.1585540176326344, 'topic_8': 0.10406425699674118, 'topic_9': 0.01992292781129059, 'topic_10': 0.02934606229742073, 'topic_11': 0.12158613438730066, 'topic_12': 0.070309808592751}
Chat: 19 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_22896523148@c.us.chat.log ):
Topics: {}
Chat: 20 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_22965141143@c.us.chat.log ):
Topics: {}
Chat: 21 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_233544217390-1490962626@g.us.chat.log ):

```

```

Topics:      {'topic_0': 0.038580203854091784, 'topic_1': 0.053012034780832405, 'topic_2':
0.05585868794762794, 'topic_3': 0.13652224686674577, 'topic_4': 0.052415364387957, 'topic_5':
0.09171030951835359, 'topic_6': 0.07736821834620437, 'topic_7': 0.05232012018119288, 'topic_8':
0.17605850182314287, 'topic_9': 0.03876575492018618, 'topic_10': 0.04641425719164907, 'topic_11':
0.05391861744510841, 'topic_12': 0.1270556827369121}
Chat:       22 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2347033038665-1563313466@g.us.chat.log ):
Topics:      {'topic_0': 0.034959925374121965, 'topic_1': 0.05721423317946934, 'topic_2':
0.04548469275460575, 'topic_3': 0.06212923143930142, 'topic_4': 0.1163211553708416, 'topic_5':
0.1097683861036716, 'topic_6': 0.0560037222922487, 'topic_7': 0.05140866122972445, 'topic_8':
0.155437163871433, 'topic_9': 0.05786555296261611, 'topic_10': 0.05212944085166394, 'topic_11':
0.05861159380243992, 'topic_12': 0.14266624076786696}
Chat:       23 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2347034328548-1491921470@g.us.chat.log ):
Topics:      {'topic_0': 0.03881364054078689, 'topic_1': 0.05337672398088203, 'topic_2':
0.1544589673389534, 'topic_3': 0.0896777150048108, 'topic_4': 0.049954986273842795, 'topic_5':
0.10213032213104835, 'topic_6': 0.06136761344866988, 'topic_7': 0.11902427157592962, 'topic_8':
0.1194860145253893, 'topic_9': 0.029740556253250932, 'topic_10': 0.04767539732649985, 'topic_11':
0.058848541224504676, 'topic_12': 0.07544525037543079}
Chat:       24 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2347080462391-1519061348@g.us.chat.log ):
Topics:      {'topic_0': 0.01854724033231941, 'topic_1': 0.14450512120621906, 'topic_2':
0.048580716880489296, 'topic_3': 0.038907506122597274, 'topic_4': 0.044032739500149055, 'topic_5':
0.1271828935350032, 'topic_6': 0.07665092567189938, 'topic_7': 0.09695294758069538, 'topic_8':
0.1291860850110405, 'topic_9': 0.026455415426675215, 'topic_10': 0.03433479181110754, 'topic_11':
0.10155008688171464, 'topic_12': 0.11311353004008987}
Chat:       25 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348024790559-1559600306@g.us.chat.log ):
Topics:      {'topic_0': 0.03137113200733387, 'topic_1': 0.03074483840861454, 'topic_2':
0.024809728417357114, 'topic_3': 0.04141348454980667, 'topic_4': 0.03124680908190127, 'topic_5':
0.16283194744279525, 'topic_6': 0.02560664848084461, 'topic_7': 0.06961981894945313, 'topic_8':
0.1896256906300142, 'topic_9': 0.035595262691062966, 'topic_10': 0.030682435880309993, 'topic_11':
0.0576728077043789, 'topic_12': 0.2687793957561275}
Chat:       26 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348032322513-1559282930@g.us.chat.log ):
Topics:      {'topic_0': 0.018527171507765557, 'topic_1': 0.01617877246435391, 'topic_2':
0.01516996575149287, 'topic_3': 0.03574117096706664, 'topic_4': 0.3525169053623966, 'topic_5':
0.21319740068124218, 'topic_6': 0.05410866782957998, 'topic_7': 0.02350730835609191, 'topic_8':
0.09773409789559469, 'topic_9': 0.018497160756859914, 'topic_10': 0.020278746325832382, 'topic_11':
0.018175206650556473, 'topic_12': 0.11636742545116902}
Chat:       27 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348033488384-1525616235@g.us.chat.log ):
Topics:      {'topic_0': 0.03458014166219672, 'topic_1': 0.10575042427512181, 'topic_2':
0.051193834801644765, 'topic_3': 0.028239359476641843, 'topic_4': 0.036752641916410374, 'topic_5':
0.14511535760123556, 'topic_6': 0.06312392797451546, 'topic_7': 0.1242731709712608, 'topic_8':
0.15346036037896882, 'topic_9': 0.0376082825040672, 'topic_10': 0.03599788700249476, 'topic_11':
0.07928875386453511, 'topic_12': 0.10461585757090741}
Chat:       28 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348054135475-1563140992@g.us.chat.log ):

```

```

Topics:      {'topic_0': 0.023601671642543114, 'topic_1': 0.02218866730342327, 'topic_2':
0.022939653674784413, 'topic_3': 0.0447148507096198, 'topic_4': 0.04176119960948828, 'topic_5':
0.13710924948095696, 'topic_6': 0.026898941890012407, 'topic_7': 0.07301397078263856, 'topic_8':
0.11420492954487603, 'topic_9': 0.027203634078730304, 'topic_10': 0.04988255014896917, 'topic_11':
0.06183966800404066, 'topic_12': 0.3546410131299174}
Chat:       29 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348062137437-1561882666@g.us.chat.log ):
Topics:      {'topic_0': 0.03323617241377447, 'topic_1': 0.069532418864943, 'topic_2': 0.03693583110324656,
'topic_3': 0.06520564190706521, 'topic_4': 0.07142422640120229, 'topic_5': 0.17782091967126346,
'topic_6': 0.07071797275377366, 'topic_7': 0.056102231161375506, 'topic_8': 0.12879924827422493,
'topic_9': 0.06405893886510373, 'topic_10': 0.07295983824744051, 'topic_11': 0.054533660837400987,
'topic_12': 0.09867289949918549}
Chat:       30 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348063025256-1559096470@g.us.chat.log ):
Topics:      {'topic_0': 0.04006410256410751, 'topic_1': 0.040064109060313956, 'topic_2':
0.052564140803878924, 'topic_3': 0.0400641076208158, 'topic_4': 0.09006411074756482, 'topic_5':
0.07383504187751433, 'topic_6': 0.12339738923903668, 'topic_7': 0.06019101899507554, 'topic_8':
0.09006412881685152, 'topic_9': 0.07339744598130249, 'topic_10': 0.0400641086902081, 'topic_11':
0.16116620145862665, 'topic_12': 0.11506409414470373}
Chat:       31 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348065669534-1541854404@g.us.chat.log ):
Topics:      {'topic_0': 0.05497057543210692, 'topic_1': 0.04754599124903521, 'topic_2':
0.045086668082417174, 'topic_3': 0.028027998892401238, 'topic_4': 0.019988909288186866, 'topic_5':
0.07204698308890553, 'topic_6': 0.0261813988801574, 'topic_7': 0.1726002874384916, 'topic_8':
0.10087269311006375, 'topic_9': 0.06910382272409721, 'topic_10': 0.03847108112009956, 'topic_11':
0.056316756976419455, 'topic_12': 0.26878683371762}
Chat:       32 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348109890004-1563317908@g.us.chat.log ):
Topics:      {'topic_0': 0.026371581579765654, 'topic_1': 0.044735970147164635, 'topic_2':
0.04508087851469046, 'topic_3': 0.0798787108349484, 'topic_4': 0.04547448649101736, 'topic_5':
0.05280412869348742, 'topic_6': 0.06632484390689983, 'topic_7': 0.07081525539433098, 'topic_8':
0.3999310706352806, 'topic_9': 0.043986457255007404, 'topic_10': 0.044002773377439755, 'topic_11':
0.04798838940395059, 'topic_12': 0.03260545376601712}
Chat:       33 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348110107300-1543757473@g.us.chat.log ):
Topics:      {'topic_0': 0.009995736186215258, 'topic_1': 0.009995751563658534, 'topic_2':
0.009995736186214547, 'topic_3': 0.009995744851713975, 'topic_4': 0.02079820417368764, 'topic_5':
0.009995743899577924, 'topic_6': 0.00999573618621608, 'topic_7': 0.012979409632141788, 'topic_8':
0.06441672030262915, 'topic_9': 0.009995748007190967, 'topic_10': 0.01840542646862452, 'topic_11':
0.023774380559944238, 'topic_12': 0.7896556619821861}
Chat:       34 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348115506965-1549270749@g.us.chat.log ):
Topics:      {'topic_0': 0.012598281650582861, 'topic_1': 0.04593968042049407, 'topic_2':
0.057460500157434785, 'topic_3': 0.021266771547078806, 'topic_4': 0.019658441773675, 'topic_5':
0.12460846333630875, 'topic_6': 0.09460311839595424, 'topic_7': 0.1626694268528653, 'topic_8':
0.17249697228409203, 'topic_9': 0.01769439463507324, 'topic_10': 0.03292209098970624, 'topic_11':
0.16770960514752745, 'topic_12': 0.07037225280920718}
Chat:       35 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348145686926-1564015604@g.us.chat.log ):

```

```

Topics:      {'topic_0': 0.05207100591716018, 'topic_1': 0.05207100962276922, 'topic_2':
0.05207102448206914, 'topic_3': 0.0520710059171602, 'topic_4': 0.052071009487123246, 'topic_5':
0.09937686161249162, 'topic_6': 0.2782834947790705, 'topic_7': 0.07261886243060887, 'topic_8':
0.0810816918672118, 'topic_9': 0.05207100591716, 'topic_10': 0.0520710065508861, 'topic_11':
0.05207101549912929, 'topic_12': 0.052071005917160025}
Chat:       36 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348147250000-1553249631@g.us.chat.log ):
Topics:      {'topic_0': 0.2537595957906262, 'topic_1': 0.022183729550979362, 'topic_2':
0.01848273377492522, 'topic_3': 0.010746652834698027, 'topic_4': 0.020114795960300604, 'topic_5':
0.020431194952012652, 'topic_6': 0.025268224150866183, 'topic_7': 0.03153130411772384, 'topic_8':
0.032432559810345624, 'topic_9': 0.43829838938599436, 'topic_10': 0.021128079076895418, 'topic_11':
0.06461614327615894, 'topic_12': 0.04100659731847412}
Chat:       37 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348149460003-1547065630@g.us.chat.log ):
Topics:      {'topic_0': 0.02465271737754033, 'topic_1': 0.1311908208088211, 'topic_2':
0.05819878235371077, 'topic_3': 0.04067454360466205, 'topic_4': 0.04243849491448733, 'topic_5':
0.12261431080769176, 'topic_6': 0.05403959394993984, 'topic_7': 0.10119539151184183, 'topic_8':
0.12997515938544, 'topic_9': 0.029507522809072345, 'topic_10': 0.058886445368684155, 'topic_11':
0.1239417203015887, 'topic_12': 0.08268449680651943}
Chat:       38 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348155594783-1563117947@g.us.chat.log ):
Topics:      {'topic_0': 0.011897150232610012, 'topic_1': 0.012083491498844526, 'topic_2':
0.011436752686321064, 'topic_3': 0.04002877780696807, 'topic_4': 0.38895766975354884, 'topic_5':
0.2689208033251972, 'topic_6': 0.06401326377846314, 'topic_7': 0.012076288071469078, 'topic_8':
0.04447593775864708, 'topic_9': 0.010636011138845422, 'topic_10': 0.018661717774117712, 'topic_11':
0.01619183496882285, 'topic_12': 0.10062030120614476}
Chat:       39 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348160741717-1468963101@g.us.chat.log ):
Topics:      {'topic_0': 0.05130430825213938, 'topic_1': 0.030476767913094626, 'topic_2':
0.02249696737829716, 'topic_3': 0.04537500605660112, 'topic_4': 0.02732788856781291, 'topic_5':
0.02101007152001549, 'topic_6': 0.04093445894495863, 'topic_7': 0.02714298002753351, 'topic_8':
0.05659788421411476, 'topic_9': 0.017804107516369, 'topic_10': 0.03436078189002831, 'topic_11':
0.1221519077544012, 'topic_12': 0.5030168699646342}
Chat:       40 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2348189583114-1517124754@g.us.chat.log ):
Topics:      {'topic_0': 0.01564860440812232, 'topic_1': 0.07780357358619938, 'topic_2':
0.07185640525859444, 'topic_3': 0.03695082330998278, 'topic_4': 0.022516771327610787, 'topic_5':
0.15294680837238103, 'topic_6': 0.08631816267012221, 'topic_7': 0.1839211916169751, 'topic_8':
0.16375606523316658, 'topic_9': 0.01833499460126836, 'topic_10': 0.017346451171574254, 'topic_11':
0.057188509482359776, 'topic_12': 0.09541163896164291}
Chat:       41 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2349079721832@c.us.chat.log ):
Topics:      {}
Chat:       42 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2349086550164-1556835426@g.us.chat.log ):
Topics:      {'topic_0': 0.029012529308199673, 'topic_1': 0.07107687012577851, 'topic_2':
0.04462791084283287, 'topic_3': 0.06818377922013891, 'topic_4': 0.04887701707150803, 'topic_5':
0.09373181219300511, 'topic_6': 0.12237120575189722, 'topic_7': 0.044104915186007745, 'topic_8':
0.2875953799137172, 'topic_9': 0.03864059796789394, 'topic_10': 0.06634689720204687, 'topic_11':

```

```

0.05104269014208091, 'topic_12': 0.034388395074893024}
Chat: 43 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_2349098643869-1562995313@g.us.chat.log ):
Topics: {'topic_0': 0.05578854150230599, 'topic_1': 0.07908355602943824, 'topic_2':
0.061876581048362134, 'topic_3': 0.08305755411626907, 'topic_4': 0.07604920757593167, 'topic_5':
0.07845091036632476, 'topic_6': 0.07164336724506616, 'topic_7': 0.06042532113857635, 'topic_8':
0.1355556127811933, 'topic_9': 0.06348885842073945, 'topic_10': 0.06576281172764303, 'topic_11':
0.08213350009297576, 'topic_12': 0.08668417795516604}
Chat: 44 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_23754804184-
1517662089@g.us.chat.log ):
Topics: {'topic_0': 0.07392018052269535, 'topic_1': 0.07358381359206163, 'topic_2':
0.06434840201552633, 'topic_3': 0.060839235870327885, 'topic_4': 0.09614494597651164, 'topic_5':
0.06926543218341322, 'topic_6': 0.053941331495328926, 'topic_7': 0.04718280844225933, 'topic_8':
0.13574321417515756, 'topic_9': 0.07431012246616024, 'topic_10': 0.08423319790896505, 'topic_11':
0.07907670614414987, 'topic_12': 0.0874106092074494}
Chat: 45 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_237651690064-
1552403045@g.us.chat.log ):
Topics: {'topic_0': 0.035694639228507734, 'topic_1': 0.06945514737295401, 'topic_2':
0.08197617065280317, 'topic_3': 0.0611663554966269, 'topic_4': 0.06533548729332887, 'topic_5':
0.09151004770501062, 'topic_6': 0.15244290601432417, 'topic_7': 0.06943249405684158, 'topic_8':
0.10498588926555913, 'topic_9': 0.030411118976200857, 'topic_10': 0.05723525471027701, 'topic_11':
0.05572023156167563, 'topic_12': 0.12463425766589188}
Chat: 46 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_254706773827-
1534684758@g.us.chat.log ):
Topics: {'topic_0': 0.05599534100564576, 'topic_1': 0.04175403690131536, 'topic_2':
0.10046334818011184, 'topic_3': 0.04781588150128043, 'topic_4': 0.049085946260059715, 'topic_5':
0.1426266513942349, 'topic_6': 0.06373504551133215, 'topic_7': 0.12917981565761216, 'topic_8':
0.12018182371498091, 'topic_9': 0.042641317766379964, 'topic_10': 0.06010322231859663, 'topic_11':
0.07471452322039376, 'topic_12': 0.07170304656805646}
Chat: 47 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_254707936002-
1541854957@g.us.chat.log ):
Topics: {'topic_0': 0.11611285227093245, 'topic_1': 0.048136732667430494, 'topic_2':
0.10357353662044222, 'topic_3': 0.02238148813198801, 'topic_4': 0.023928446335415066, 'topic_5':
0.07394803002320434, 'topic_6': 0.016600510398498372, 'topic_7': 0.040728600631502715, 'topic_8':
0.12823154646244392, 'topic_9': 0.10791285967131642, 'topic_10': 0.01761163442165522, 'topic_11':
0.11367996972154276, 'topic_12': 0.18715379264362805}
Chat: 48 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_254757146467-
1563438096@g.us.chat.log ):
Topics: {'topic_0': 0.03673082000889865, 'topic_1': 0.05744914181472304, 'topic_2':
0.13403037447791355, 'topic_3': 0.04418059616454294, 'topic_4': 0.13141117321810866, 'topic_5':
0.09729633007939781, 'topic_6': 0.09739461100768701, 'topic_7': 0.046852876032794435, 'topic_8':
0.07075244645856395, 'topic_9': 0.035834788466456144, 'topic_10': 0.053507518027977576, 'topic_11':
0.05276579938695915, 'topic_12': 0.14179352485597665}
Chat: 49 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_254798513186-
1558264650@g.us.chat.log ):
Topics: {'topic_0': 0.035397427718261924, 'topic_1': 0.06564137254929246, 'topic_2':
0.08294588878541223, 'topic_3': 0.05937340702791121, 'topic_4': 0.06672641247367454, 'topic_5':
0.08590552307477475, 'topic_6': 0.08970116979724033, 'topic_7': 0.057761726140820036, 'topic_8':
0.06788734040609555, 'topic_9': 0.035950218247320954, 'topic_10': 0.046855766688458984, 'topic_11':

```

```

0.04987682287750056, 'topic_12': 0.25597692421323703}
Chat: 50 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_255620150767-1527782404@g.us.chat.log ):
Topics:    {'topic_0': 0.01046850243940353, 'topic_1': 0.11455329939935784, 'topic_2': 0.09594246451004097, 'topic_3': 0.01820682209742761, 'topic_4': 0.017791901232450568, 'topic_5': 0.19053753590818776, 'topic_6': 0.1838911211571953, 'topic_7': 0.03870789382912835, 'topic_8': 0.16346869429959018, 'topic_9': 0.019449964021736082, 'topic_10': 0.02196301290807626, 'topic_11': 0.06341039286932326, 'topic_12': 0.061608395328082544}
Chat: 51 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_255717713926-1562782840@g.us.chat.log ):
Topics:    {'topic_0': 0.02220877858455366, 'topic_1': 0.043008493400195955, 'topic_2': 0.12162245644806388, 'topic_3': 0.07013710410630038, 'topic_4': 0.032295400348237246, 'topic_5': 0.073188672648521, 'topic_6': 0.053000482254595326, 'topic_7': 0.13416527324601835, 'topic_8': 0.07146600367397192, 'topic_9': 0.030764500628766917, 'topic_10': 0.09645747161961657, 'topic_11': 0.03916675913067242, 'topic_12': 0.21251860391048796}
Chat: 52 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_260977284225-1550949072@g.us.chat.log ):
Topics:    {'topic_0': 0.14263025600331647, 'topic_1': 0.041705328655866486, 'topic_2': 0.04988956353775219, 'topic_3': 0.051538356393872824, 'topic_4': 0.03236313746342835, 'topic_5': 0.09055531067025485, 'topic_6': 0.03511139532803189, 'topic_7': 0.0372449744169423, 'topic_8': 0.08812538274741691, 'topic_9': 0.2268548475817923, 'topic_10': 0.01601574576359121, 'topic_11': 0.0618215867912688, 'topic_12': 0.12614411464646555}
Chat: 53 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_263772749518-1460358881@g.us.chat.log ):
Topics:    {'topic_0': 0.050980138478271736, 'topic_1': 0.07282307637951632, 'topic_2': 0.06225657467509284, 'topic_3': 0.06636830200825457, 'topic_4': 0.09051057224643318, 'topic_5': 0.09008287964799996, 'topic_6': 0.06220717530897234, 'topic_7': 0.055417534094043104, 'topic_8': 0.1663344478410289, 'topic_9': 0.06727383081938627, 'topic_10': 0.06918370952077323, 'topic_11': 0.07705023415503122, 'topic_12': 0.06951152482520609}
Chat: 54 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_26377888125-1488105200@g.us.chat.log ):
Topics:    {'topic_0': 0.016613144823221487, 'topic_1': 0.015343518843272463, 'topic_2': 0.011936479072947865, 'topic_3': 0.049168510238875286, 'topic_4': 0.41614528967728104, 'topic_5': 0.26770627995445784, 'topic_6': 0.05832058266996739, 'topic_7': 0.014587536003042117, 'topic_8': 0.041673526037778824, 'topic_9': 0.013498479298642507, 'topic_10': 0.024277744185995265, 'topic_11': 0.019767244882214284, 'topic_12': 0.05096166431230575}
Chat: 55 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_27621634355-1529677872@g.us.chat.log ):
Topics:    {'topic_0': 0.014979399779379695, 'topic_1': 0.05871951156017297, 'topic_2': 0.04266904921993977, 'topic_3': 0.028811578261309954, 'topic_4': 0.03885766428132647, 'topic_5': 0.15808754059408064, 'topic_6': 0.10257996790306555, 'topic_7': 0.2579197698844149, 'topic_8': 0.0921700450270312, 'topic_9': 0.019265339516478165, 'topic_10': 0.02641770655806053, 'topic_11': 0.12520257805968552, 'topic_12': 0.03431984935505439}
Chat: 56 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_27644635125-1509175015@g.us.chat.log ):
Topics:    {'topic_0': 0.0743535268835903, 'topic_1': 0.043600246031521746, 'topic_2': 0.03079415348906492, 'topic_3': 0.056559844277969326, 'topic_4': 0.03423717419475837, 'topic_5': 0.042910695542431924, 'topic_6': 0.042662720184537005, 'topic_7': 0.13720730646551624, 'topic_8': 0.07040155357988634, 'topic_9': 0.022674059239441817, 'topic_10': 0.14633289670650354, 'topic_11':

```



```
0.06195021675793242, 'topic_12': 0.23631560664684612}
Chat: 57 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_31621943195-1557752118@g.us.chat.log ):
Topics: {'topic_0': 0.10984994990095333, 'topic_1': 0.1263897821247624, 'topic_2': 0.052149284194233994, 'topic_3': 0.048478349445180584, 'topic_4': 0.06234002289742552, 'topic_5': 0.055701550115792614, 'topic_6': 0.052526957652601104, 'topic_7': 0.04621261281860672, 'topic_8': 0.053503744939330654, 'topic_9': 0.054984383737482975, 'topic_10': 0.20391241911237445, 'topic_11': 0.05010783260844777, 'topic_12': 0.0838431104528119}
Chat: 58 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_353862228246-1543459639@g.us.chat.log ):
Topics: {'topic_0': 0.025110023605183655, 'topic_1': 0.05322401443635983, 'topic_2': 0.04902451526348928, 'topic_3': 0.06566182338630902, 'topic_4': 0.032526503964369145, 'topic_5': 0.03242184132811916, 'topic_6': 0.035755947678779966, 'topic_7': 0.05215243008256181, 'topic_8': 0.3407416456580344, 'topic_9': 0.048158044942385535, 'topic_10': 0.055059887587970355, 'topic_11': 0.04040332448530092, 'topic_12': 0.16975999758113697}
Chat: 59 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_353862228246-1543580117@g.us.chat.log ):
Topics: {}
Chat: 60 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_353862228246-1544539545@g.us.chat.log ):
Topics: {'topic_0': 0.0415602497949629, 'topic_1': 0.04560870111930049, 'topic_2': 0.04859611339608074, 'topic_3': 0.05226037455153953, 'topic_4': 0.05488418647118438, 'topic_5': 0.043637911628842785, 'topic_6': 0.053426585615957325, 'topic_7': 0.04878164437547343, 'topic_8': 0.2899382582674798, 'topic_9': 0.04304374510834957, 'topic_10': 0.04363162561248402, 'topic_11': 0.053909930288755216, 'topic_12': 0.1807206737695898}
Chat: 61 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_353862228246-1544601100@g.us.chat.log ):
Topics: {'topic_0': 0.039109135309812276, 'topic_1': 0.05728415328727868, 'topic_2': 0.04838449501574619, 'topic_3': 0.06625059931471748, 'topic_4': 0.05511239178360527, 'topic_5': 0.0549339120282812, 'topic_6': 0.0569320779455953, 'topic_7': 0.0630812441924038, 'topic_8': 0.23487275752806414, 'topic_9': 0.06420610296455742, 'topic_10': 0.05924721287234241, 'topic_11': 0.04680698997067446, 'topic_12': 0.1537789277869213}
Chat: 62 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_447448867845-1532456432@g.us.chat.log ):
Topics: {'topic_0': 0.04828088764156236, 'topic_1': 0.06332046331352427, 'topic_2': 0.06368580199729264, 'topic_3': 0.08268584513653932, 'topic_4': 0.061721411932689596, 'topic_5': 0.06582609794491479, 'topic_6': 0.07163299945886341, 'topic_7': 0.05520521866447979, 'topic_8': 0.21038617404274135, 'topic_9': 0.06305774647387467, 'topic_10': 0.057986858996325365, 'topic_11': 0.08083621583853264, 'topic_12': 0.07537427855866001}
Chat: 63 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_447451246621-1564038749@g.us.chat.log ):
Topics: {'topic_0': 0.03669193770225504, 'topic_1': 0.04551031125946616, 'topic_2': 0.07459433051877284, 'topic_3': 0.05723037828999337, 'topic_4': 0.07670853606361572, 'topic_5': 0.0527175133249415, 'topic_6': 0.05701765218982655, 'topic_7': 0.044088769935171905, 'topic_8': 0.0633352777014518, 'topic_9': 0.043859488811996104, 'topic_10': 0.041992794131958014, 'topic_11': 0.04687462005300612, 'topic_12': 0.35937839001754507}
Chat: 64 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_447708329340-1560552735@g.us.chat.log ):
Topics: {'topic_0': 0.008872387850778521, 'topic_1': 0.01112358130577388, 'topic_2':
```

```

0.008823376962804564, 'topic_3': 0.03944756272651029, 'topic_4': 0.40064421559564645, 'topic_5':
0.2775353142686443, 'topic_6': 0.06409652679068104, 'topic_7': 0.01076677775255304, 'topic_8':
0.04690743548705173, 'topic_9': 0.009951954344256416, 'topic_10': 0.01694058184018739, 'topic_11':
0.016241963216113584, 'topic_12': 0.0886483218362958}
Chat: 65 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_447887810701-
1493490356@g.us.chat.log ):
Topics: {'topic_0': 0.015837109957898977, 'topic_1': 0.015837104072400097, 'topic_2':
0.015837123682737153, 'topic_3': 0.04438982353010409, 'topic_4': 0.029536617565459538, 'topic_5':
0.08463256220190798, 'topic_6': 0.08857611782622603, 'topic_7': 0.13688919057578314, 'topic_8':
0.2731546549753744, 'topic_9': 0.015837143670076242, 'topic_10': 0.06070437881068719, 'topic_11':
0.07305528863677045, 'topic_12': 0.1457128844945747}
Chat: 66 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_4917629460203-1551197253@g.us.chat.log ):
Topics: {'topic_0': 0.04013682630826814, 'topic_1': 0.06156684539512604, 'topic_2':
0.07100993067672513, 'topic_3': 0.07507055684664937, 'topic_4': 0.10538518185875485, 'topic_5':
0.09465687251601936, 'topic_6': 0.10389577406263768, 'topic_7': 0.05376629650820366, 'topic_8':
0.06334618129003623, 'topic_9': 0.04149334191576264, 'topic_10': 0.055673776733518456, 'topic_11':
0.07241596917605891, 'topic_12': 0.16158244671224387}
Chat: 67 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_50762850233-
1508010876@g.us.chat.log ):
Topics: {'topic_0': 0.01628489115841241, 'topic_1': 0.08655230986743392, 'topic_2':
0.16560759179684575, 'topic_3': 0.05431222704813277, 'topic_4': 0.027718557693395422, 'topic_5':
0.2592573172150899, 'topic_6': 0.04183161456676292, 'topic_7': 0.07075276547866718, 'topic_8':
0.14255672822270443, 'topic_9': 0.02124126806342934, 'topic_10': 0.025246059690114484, 'topic_11':
0.022628389316668367, 'topic_12': 0.06601027988234331}
Chat: 68 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_5492914734432-1550429334@g.us.chat.log ):
Topics: {'topic_0': 0.2108204023683212, 'topic_1': 0.028200990786989275, 'topic_2':
0.026714567330964212, 'topic_3': 0.013245672626003964, 'topic_4': 0.011958541559107451, 'topic_5':
0.02975236509265391, 'topic_6': 0.018937658941317128, 'topic_7': 0.07818906522002814, 'topic_8':
0.08677925141096697, 'topic_9': 0.20234324222099634, 'topic_10': 0.00969627956035179, 'topic_11':
0.17860592399842926, 'topic_12': 0.1047560388838715}
Chat: 69 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_5512996817977-1513419479@g.us.chat.log ):
Topics: {'topic_0': 0.013735169851996904, 'topic_1': 0.04407071084601868, 'topic_2':
0.020457899164833676, 'topic_3': 0.032834307352176635, 'topic_4': 0.022813823679093594, 'topic_5':
0.08835930332585984, 'topic_6': 0.038031166713268956, 'topic_7': 0.39220881938315855, 'topic_8':
0.18143916591234535, 'topic_9': 0.01719672367336426, 'topic_10': 0.022512329640311104, 'topic_11':
0.03133698709672066, 'topic_12': 0.09500359336085186}
Chat: 70 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_5521996180760-1540595805@g.us.chat.log ):
Topics: {'topic_0': 0.041822528859517434, 'topic_1': 0.04468701904838523, 'topic_2':
0.060339465577572304, 'topic_3': 0.055777663079016666, 'topic_4': 0.07879102579764263, 'topic_5':
0.08584821658830448, 'topic_6': 0.12928164857765087, 'topic_7': 0.08983985334310539, 'topic_8':
0.18962122000042425, 'topic_9': 0.04216552591827167, 'topic_10': 0.04962601853553922, 'topic_11':
0.05265728182359658, 'topic_12': 0.07954253285097317}
Chat: 71 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_553497819767-
1480261843@g.us.chat.log ):
Topics: {'topic_0': 0.20366632062071321, 'topic_1': 0.025834760912811866, 'topic_2':

```

```

0.05537000865506003, 'topic_3': 0.027816178401256404, 'topic_4': 0.01994003099437771, 'topic_5':
0.0460820787528688, 'topic_6': 0.032572130065211265, 'topic_7': 0.03537210606755682, 'topic_8':
0.039187329349117186, 'topic_9': 0.2119385312965443, 'topic_10': 0.03377912561649993, 'topic_11':
0.19761206351913943, 'topic_12': 0.07082933574884508}
Chat: 72 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_556295153706-
1483193085@g.us.chat.log ):
Topics: {'topic_0': 0.03068172826440021, 'topic_1': 0.045124488985196186, 'topic_2':
0.10128453716944351, 'topic_3': 0.037199557512234756, 'topic_4': 0.05396310838113852, 'topic_5':
0.09240577544491413, 'topic_6': 0.11088693056339693, 'topic_7': 0.06494683915069256, 'topic_8':
0.08052971306697708, 'topic_9': 0.038364888599410694, 'topic_10': 0.03884146385359315, 'topic_11':
0.03838621092262563, 'topic_12': 0.26738475808597767}
Chat: 73 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_601126316671-
1507101918@g.us.chat.log ):
Topics: {'topic_0': 0.04475753258471449, 'topic_1': 0.08572396862499797, 'topic_2':
0.05762417301316585, 'topic_3': 0.06957612941389386, 'topic_4': 0.09957232772337642, 'topic_5':
0.06720853973895895, 'topic_6': 0.05281343013717888, 'topic_7': 0.048481252430437764, 'topic_8':
0.1370093348952891, 'topic_9': 0.053422916813687694, 'topic_10': 0.10617207301824631, 'topic_11':
0.07392926061438905, 'topic_12': 0.10370906099166938}
Chat: 74 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_6738959828-
1563623700@g.us.chat.log ):
Topics: {'topic_0': 0.05551445968901617, 'topic_1': 0.0740741029732392, 'topic_2':
0.06350100839281045, 'topic_3': 0.07988565147188346, 'topic_4': 0.07333507601546041, 'topic_5':
0.0792434241491256, 'topic_6': 0.06767512705670946, 'topic_7': 0.05835904043751191, 'topic_8':
0.1445699754712164, 'topic_9': 0.06797200655137538, 'topic_10': 0.06217105151784088, 'topic_11':
0.08478316894077315, 'topic_12': 0.08891590733309501}
Chat: 75 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_85268553604@c.us.chat.log ):
Topics: {}
Chat: 76 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_8613136156865-1500422552@g.us.chat.log ):
Topics: {'topic_0': 0.00906001641515945, 'topic_1': 0.1052868257396274, 'topic_2':
0.027979671341126118, 'topic_3': 0.040884024464358154, 'topic_4': 0.014297862566175679, 'topic_5':
0.1290529642521614, 'topic_6': 0.06989948053242541, 'topic_7': 0.2431064697046703, 'topic_8':
0.13204991746807732, 'topic_9': 0.014813289293463096, 'topic_10': 0.023164551349289485, 'topic_11':
0.16243278424200752, 'topic_12': 0.027972142631458488}
Chat: 77 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_8615099927791-1521182088@g.us.chat.log ):
Topics: {'topic_0': 0.009615384615384767, 'topic_1': 0.009615384615384795, 'topic_2':
0.009615397723604752, 'topic_3': 0.009615386450836521, 'topic_4': 0.009615384615384746, 'topic_5':
0.8846151939619467, 'topic_6': 0.009615392002183303, 'topic_7': 0.009615401931077393, 'topic_8':
0.00961541968460169, 'topic_9': 0.009615490899455573, 'topic_10': 0.009615394269370199, 'topic_11':
0.009615384615384717, 'topic_12': 0.009615384615384704}
Chat: 78 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_8801744419079-1509171830@g.us.chat.log ):
Topics: {'topic_0': 0.07732512579574914, 'topic_1': 0.07649390460803658, 'topic_2':
0.05557800603132947, 'topic_3': 0.05518606066344168, 'topic_4': 0.08911670380718666, 'topic_5':
0.07528839680899828, 'topic_6': 0.04661690385527257, 'topic_7': 0.041958989665661556, 'topic_8':
0.14434076583173314, 'topic_9': 0.11904723718213236, 'topic_10': 0.07831190516492821, 'topic_11':
0.07987561961863686, 'topic_12': 0.060860380966897845}

```

```

Chat:      79 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_8801812755493-1505678107@g.us.chat.log ):
Topics:    {'topic_0': 0.07237157067501913, 'topic_1': 0.07733096528750485, 'topic_2':
0.08935359508237078, 'topic_3': 0.018603109640464053, 'topic_4': 0.04051463517600982, 'topic_5':
0.32028009757961556, 'topic_6': 0.035956396688819184, 'topic_7': 0.07005892340020613, 'topic_8':
0.11726187804465703, 'topic_9': 0.07626558333840677, 'topic_10': 0.014885874557280345, 'topic_11':
0.03719806068330256, 'topic_12': 0.029919309846343622}
Chat: 80 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_916207231114-
1544876477@g.us.chat.log ):
Topics:    {'topic_0': 0.04161162211286591, 'topic_1': 0.06174617959192148, 'topic_2':
0.1201790983247784, 'topic_3': 0.060218379960302014, 'topic_4': 0.0601145045383109, 'topic_5':
0.12308445646053003, 'topic_6': 0.10069326642106105, 'topic_7': 0.09450757278224388, 'topic_8':
0.10704951424933662, 'topic_9': 0.040256024122626924, 'topic_10': 0.06505459719006858, 'topic_11':
0.05547322474690161, 'topic_12': 0.07001155949905176}
Chat: 81 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_917011284590-
1486113958@g.us.chat.log ):
Topics:    {'topic_0': 0.015206247227519279, 'topic_1': 0.04566647734507272, 'topic_2':
0.0676293974265145, 'topic_3': 0.054830051289055236, 'topic_4': 0.09122451187185392, 'topic_5':
0.12048659331316462, 'topic_6': 0.055303741083768634, 'topic_7': 0.056997377132350216, 'topic_8':
0.2166678312323834, 'topic_9': 0.020319948951587365, 'topic_10': 0.20107981316548088, 'topic_11':
0.02627823833626175, 'topic_12': 0.02830977162498763}
Chat: 82 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_917016354308-
1488093290@g.us.chat.log ):
Topics:    {'topic_0': 0.16436549496966218, 'topic_1': 0.05999811577958141, 'topic_2':
0.06293971441108275, 'topic_3': 0.02694020902073378, 'topic_4': 0.017969584485972568, 'topic_5':
0.05950263248597405, 'topic_6': 0.04851088931338672, 'topic_7': 0.05361997931077017, 'topic_8':
0.06863804172921538, 'topic_9': 0.11862570654647611, 'topic_10': 0.036821082781754506, 'topic_11':
0.11291905753412225, 'topic_12': 0.16914949163126916}
Chat: 83 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_917058803589-
1528353496@g.us.chat.log ):
Topics:    {'topic_0': 0.019987221652591734, 'topic_1': 0.018021008894970745, 'topic_2':
0.015700471843265833, 'topic_3': 0.017783316525083387, 'topic_4': 0.03004990767330737, 'topic_5':
0.013035882916138105, 'topic_6': 0.03115073161670551, 'topic_7': 0.014598719907222162, 'topic_8':
0.01833242964174504, 'topic_9': 0.011740823318386063, 'topic_10': 0.03244467854560151, 'topic_11':
0.046140277696529494, 'topic_12': 0.7310145297684552}
Chat:      84 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_917066389716@c.us.chat.log ):
Topics:    {'topic_0': 0.004116272688088078, 'topic_1': 0.02384014869225735, 'topic_2':
0.04554947617992353, 'topic_3': 0.01826883443977195, 'topic_4': 0.03744961062924122, 'topic_5':
0.004116290798570925, 'topic_6': 0.021951520836543708, 'topic_7': 0.013121234400504977, 'topic_8':
0.5577445879211792, 'topic_9': 0.004116284888008994, 'topic_10': 0.02488225255180999, 'topic_11':
0.04578294136821523, 'topic_12': 0.19906054460588476}
Chat: 85 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\chat_917096666815-
1563845218@g.us.chat.log ):
Topics:    {'topic_0': 0.12879658447254758, 'topic_1': 0.06403094351247708, 'topic_2':
0.25835554769408237, 'topic_3': 0.0234358728678352, 'topic_4': 0.02343586718586804, 'topic_5':
0.027940403157207427, 'topic_6': 0.023435867185868448, 'topic_7': 0.023435871329962468, 'topic_8':
0.06510255238378253, 'topic_9': 0.023435877238141552, 'topic_10': 0.0343018370288671, 'topic_11':
0.08085691754490457, 'topic_12': 0.22343585839845578}

```

```

Chat: 86 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_917340270078-1563980314@g.us.chat.log ):
Topics:    {'topic_0': 0.005128205128208588, 'topic_1': 0.14540643011212162, 'topic_2': 0.3262656881516597, 'topic_3': 0.07179487121151668, 'topic_4': 0.0051282291772124025, 'topic_5': 0.005128311094309438, 'topic_6': 0.005128230117283588, 'topic_7': 0.005128278489507945, 'topic_8': 0.005128323485841604, 'topic_9': 0.0771380823985944, 'topic_10': 0.1506570957693238, 'topic_11': 0.1928400497362132, 'topic_12': 0.005128205128207179}
Chat: 87 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_917504961079-1519030093@g.us.chat.log ):
Topics:    {'topic_0': 0.1316672481952389, 'topic_1': 0.04097380027755819, 'topic_2': 0.042723002134124395, 'topic_3': 0.05060259893118191, 'topic_4': 0.043150653261606116, 'topic_5': 0.10635989339493257, 'topic_6': 0.05123798563730802, 'topic_7': 0.09252431969652354, 'topic_8': 0.1212237252470041, 'topic_9': 0.08885723825398187, 'topic_10': 0.03891478618819332, 'topic_11': 0.06993747941860091, 'topic_12': 0.1218272693637488}
Chat: 88 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_917504961079-1519136340@g.us.chat.log ):
Topics:    {'topic_0': 0.2247605459406087, 'topic_1': 0.014176997466120271, 'topic_2': 0.13379463723908902, 'topic_3': 0.014320676453772989, 'topic_4': 0.011661679122035468, 'topic_5': 0.0177868574046, 'topic_6': 0.017608304260214, 'topic_7': 0.03362240845018892, 'topic_8': 0.02508754420649231, 'topic_9': 0.3750615641728215, 'topic_10': 0.014005622139061409, 'topic_11': 0.05606968445158336, 'topic_12': 0.06204347869341193}
Chat: 89 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_917504961079-1519185392@g.us.chat.log ):
Topics:    {'topic_0': 0.050624478163874234, 'topic_1': 0.056573544630224085, 'topic_2': 0.15502665141531477, 'topic_3': 0.08609391813291078, 'topic_4': 0.044353681840272184, 'topic_5': 0.07388725943228534, 'topic_6': 0.0854467603763131, 'topic_7': 0.07923576599621547, 'topic_8': 0.08939786529442091, 'topic_9': 0.048691107794978566, 'topic_10': 0.05291110573958182, 'topic_11': 0.08860487289038482, 'topic_12': 0.0891529882932231}
Chat: 90 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_917504961079-1519185435@g.us.chat.log ):
Topics:    {'topic_0': 0.051518382145897185, 'topic_1': 0.044725945494336936, 'topic_2': 0.16282071009201857, 'topic_3': 0.08345470922374668, 'topic_4': 0.048824416484455666, 'topic_5': 0.07603911162857697, 'topic_6': 0.0652595628605888, 'topic_7': 0.06024485371553271, 'topic_8': 0.11023452812592997, 'topic_9': 0.05159989924886423, 'topic_10': 0.07358862467998015, 'topic_11': 0.07381041376733058, 'topic_12': 0.09787884253274151}
Chat: 91 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_917896159081-1506248485@g.us.chat.log ):
Topics:    {'topic_0': 0.02312974955627954, 'topic_1': 0.1301302573654485, 'topic_2': 0.02468157894512994, 'topic_3': 0.06125553419046639, 'topic_4': 0.04483519991942042, 'topic_5': 0.0936196390763585, 'topic_6': 0.1159025084418818, 'topic_7': 0.12037366813147829, 'topic_8': 0.15129795993797046, 'topic_9': 0.03555444098630349, 'topic_10': 0.038243656526311884, 'topic_11': 0.05489578526146349, 'topic_12': 0.10608002166148713}
Chat: 92 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_917900186560-1550952626@g.us.chat.log ):
Topics:    {'topic_0': 0.036893630640542824, 'topic_1': 0.06218846412957144, 'topic_2': 0.06412931391839183, 'topic_3': 0.09429770405028462, 'topic_4': 0.08117233673260651, 'topic_5': 0.0821459273331371, 'topic_6': 0.06143018862843549, 'topic_7': 0.09066705645073658, 'topic_8': 0.10246014840120861, 'topic_9': 0.03781726961672482, 'topic_10': 0.08029906338455788, 'topic_11': 0.0686116473582646, 'topic_12': 0.13788724935554122}

```

```
Chat: 93 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_917985455537-1488972204@g.us.chat.log ):
Topics: {'topic_0': 0.03598372698571713, 'topic_1': 0.048352485792405614, 'topic_2': 0.12392505789123626, 'topic_3': 0.08073124190771448, 'topic_4': 0.05363950816621982, 'topic_5': 0.08678556980221035, 'topic_6': 0.08295809202903894, 'topic_7': 0.0630899676194227, 'topic_8': 0.0738070430018706, 'topic_9': 0.03642890034376196, 'topic_10': 0.05183734833481169, 'topic_11': 0.04481867364068803, 'topic_12': 0.21764238448490314}
Chat: 94 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_918011717118-1513346963@g.us.chat.log ):
Topics: {'topic_0': 0.07508999143550248, 'topic_1': 0.04224075090166977, 'topic_2': 0.05541179944248831, 'topic_3': 0.047842177799637606, 'topic_4': 0.05262792249819091, 'topic_5': 0.15115031174692112, 'topic_6': 0.048097147670792685, 'topic_7': 0.07373272070413166, 'topic_8': 0.13136371351028822, 'topic_9': 0.0712383728621516, 'topic_10': 0.07865024480857259, 'topic_11': 0.08732214690499983, 'topic_12': 0.08523269971465554}
Chat: 95 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_918074820605@c.us.chat.log ):
Topics: {}
Chat: 96 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_918112269746-1515737968@g.us.chat.log ):
Topics: {}
Chat: 97 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_918154995870-1558890931@g.us.chat.log ):
Topics: {'topic_0': 0.06475496589496199, 'topic_1': 0.07375542876987455, 'topic_2': 0.07781551841795284, 'topic_3': 0.07289795841081353, 'topic_4': 0.06677925748326023, 'topic_5': 0.08257525393248831, 'topic_6': 0.06756950362102616, 'topic_7': 0.08486195380480399, 'topic_8': 0.08142383401983959, 'topic_9': 0.08114519460410857, 'topic_10': 0.0677481379660024, 'topic_11': 0.07181276596154024, 'topic_12': 0.10686022711332818}
Chat: 98 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_918178289603-1561741836@g.us.chat.log ):
Topics: {'topic_0': 0.14052072939950128, 'topic_1': 0.0288057142219706, 'topic_2': 0.11564167346253733, 'topic_3': 0.034260047260053515, 'topic_4': 0.030239958226599434, 'topic_5': 0.05154977848715648, 'topic_6': 0.06647902951657951, 'topic_7': 0.04450246404974303, 'topic_8': 0.0500109156995907, 'topic_9': 0.15836674036733825, 'topic_10': 0.03508614893772048, 'topic_11': 0.16151103003919562, 'topic_12': 0.08302577033201497}
Chat: 99 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_918197180324-1564038241@g.us.chat.log ):
Topics: {'topic_0': 0.01530729339277379, 'topic_1': 0.017694665324908466, 'topic_2': 0.029249422040178004, 'topic_3': 0.014259263767079486, 'topic_4': 0.06741290607234825, 'topic_5': 0.06353171601973617, 'topic_6': 0.02886559543587746, 'topic_7': 0.037305481429304975, 'topic_8': 0.02715985184754175, 'topic_9': 0.01604304785957811, 'topic_10': 0.028200960919206684, 'topic_11': 0.02406923548534784, 'topic_12': 0.6309005604061191}
Chat: 100 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_918197180324-1564039245@g.us.chat.log ):
Topics: {'topic_0': 0.02336356953881544, 'topic_1': 0.016533109251261547, 'topic_2': 0.030730997193929608, 'topic_3': 0.016533112389374743, 'topic_4': 0.0746305417120335, 'topic_5': 0.030861250986574752, 'topic_6': 0.06434426268283198, 'topic_7': 0.022814114523326122, 'topic_8': 0.027100328754741835, 'topic_9': 0.017997409012528057, 'topic_10': 0.01653310365050908, 'topic_11': 0.031170783249182234, 'topic_12': 0.6273874170548916}
Chat: 101 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/
```

```

chat_918507097725-1557250497@g.us.chat.log ):
Topics:    {'topic_0': 0.03160031795342985, 'topic_1': 0.035347269167245475, 'topic_2':
0.27091371589221885, 'topic_3': 0.05850215488315499, 'topic_4': 0.042416509901601764, 'topic_5':
0.05459413532569933, 'topic_6': 0.047457266986474775, 'topic_7': 0.04094518362427749, 'topic_8':
0.07693861059453466, 'topic_9': 0.03795716532280529, 'topic_10': 0.058551986348178243, 'topic_11':
0.04179259398960013, 'topic_12': 0.20298309001078021}
Chat:      102 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_918795192508-1483164938@g.us.chat.log ):
Topics:    {'topic_0': 0.0496152240589887, 'topic_1': 0.07529407365125416, 'topic_2':
0.12626191447551716, 'topic_3': 0.04239202806085333, 'topic_4': 0.03937324522846319, 'topic_5':
0.06225995867037008, 'topic_6': 0.05799768286768014, 'topic_7': 0.10218805499328681, 'topic_8':
0.09480887346213646, 'topic_9': 0.04048285735590189, 'topic_10': 0.060007200903817906, 'topic_11':
0.060250394423022526, 'topic_12': 0.1890684918487094}
Chat:      103 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_918795744594-1541928672@g.us.chat.log ):
Topics:    {'topic_0': 0.1004312001042653, 'topic_1': 0.03990082770107865, 'topic_2':
0.06447405596716874, 'topic_3': 0.056647291139980745, 'topic_4': 0.04060363948610821, 'topic_5':
0.05388554117850037, 'topic_6': 0.0809305057670713, 'topic_7': 0.08053775521258467, 'topic_8':
0.057156274925696575, 'topic_9': 0.09383939792405245, 'topic_10': 0.05756471418969919, 'topic_11':
0.09025765598387495, 'topic_12': 0.18377114041992}
Chat:      104 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_918826310112-1553089301@g.us.chat.log ):
Topics:    {'topic_0': 0.04685708719548215, 'topic_1': 0.06706069958790524, 'topic_2':
0.10013213893317978, 'topic_3': 0.06181668880262264, 'topic_4': 0.06905687469023802, 'topic_5':
0.0743111326569528, 'topic_6': 0.09650267131572966, 'topic_7': 0.09726260176938305, 'topic_8':
0.12325085429581352, 'topic_9': 0.048849787859307915, 'topic_10': 0.06699797997893706, 'topic_11':
0.05912451538399018, 'topic_12': 0.088776967530459}
Chat:      105 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_918850732033-1560277858@g.us.chat.log ):
Topics:    {'topic_0': 0.05478117323416577, 'topic_1': 0.06520420453441897, 'topic_2':
0.05267052317463081, 'topic_3': 0.06531152019134488, 'topic_4': 0.0674888754449523, 'topic_5':
0.10844151771407225, 'topic_6': 0.09948097680997253, 'topic_7': 0.07583550940280008, 'topic_8':
0.12625536962239256, 'topic_9': 0.060188298890827815, 'topic_10': 0.07171170816097276, 'topic_11':
0.06861366722668218, 'topic_12': 0.08401665559277062}
Chat:      106 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_918850823474-1551458263@g.us.chat.log ):
Topics:    {'topic_0': 0.1327739724979287, 'topic_1': 0.05267211780656919, 'topic_2':
0.054719869206725745, 'topic_3': 0.03841633646716569, 'topic_4': 0.040176644687210966, 'topic_5':
0.0551169580849258, 'topic_6': 0.09375600120368957, 'topic_7': 0.07877479887063013, 'topic_8':
0.059330653211369024, 'topic_9': 0.13643088605163442, 'topic_10': 0.04656305613369485, 'topic_11':
0.13999800371462065, 'topic_12': 0.07127070206383507}
Chat:      107 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_918866126299-1502878641@g.us.chat.log ):
Topics:    {'topic_0': 0.08993359474896043, 'topic_1': 0.08075821562886316, 'topic_2':
0.04247007023138779, 'topic_3': 0.057904195883737776, 'topic_4': 0.03253476335142275, 'topic_5':
0.09924466403949941, 'topic_6': 0.038377804909625994, 'topic_7': 0.11786701093560001, 'topic_8':
0.1688105352990239, 'topic_9': 0.06557950639507185, 'topic_10': 0.05073405122894644, 'topic_11':
0.05185516098062177, 'topic_12': 0.103930426367241}
Chat:      108 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\

```

```

chat_918874547144-1539766758@g.us.chat.log ):
Topics:      {'topic_0': 0.025422084291708877, 'topic_1': 0.041277013823734804, 'topic_2':
0.13881972234432663, 'topic_3': 0.1041475032688435, 'topic_4': 0.03500196076500955, 'topic_5':
0.06024636438966051, 'topic_6': 0.04980880974167178, 'topic_7': 0.14924536264923466, 'topic_8':
0.05392150850230216, 'topic_9': 0.025764915286740107, 'topic_10': 0.08468383500673356, 'topic_11':
0.03311754133203779, 'topic_12': 0.19854337859799762}
Chat:       109 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_918903675398-1525952582@g.us.chat.log ):
Topics:      {'topic_0': 0.03206886696092281, 'topic_1': 0.060394584029065265, 'topic_2':
0.09326453567352143, 'topic_3': 0.08371116309402914, 'topic_4': 0.0700343251836527, 'topic_5':
0.059757722274750374, 'topic_6': 0.0799161588370732, 'topic_7': 0.0750495326829312, 'topic_8':
0.08110042400167498, 'topic_9': 0.03397124472591481, 'topic_10': 0.05465499668795623, 'topic_11':
0.05242642950814064, 'topic_12': 0.22365001634036716}
Chat:       110 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_918956783671-1552825100@g.us.chat.log ):
Topics:      {'topic_0': 0.11113326187075233, 'topic_1': 0.03787487041516499, 'topic_2':
0.08679469394365155, 'topic_3': 0.04048842020055388, 'topic_4': 0.039295145944196475, 'topic_5':
0.0789820298628768, 'topic_6': 0.06474791393584454, 'topic_7': 0.11818970328919012, 'topic_8':
0.09423954430636981, 'topic_9': 0.1226030355527616, 'topic_10': 0.03567087918399786, 'topic_11':
0.09898946937654522, 'topic_12': 0.0709910321180974}
Chat:       111 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_919102711964-1562017267@g.us.chat.log ):
Topics:      {'topic_0': 0.043487815639039265, 'topic_1': 0.07817389999869194, 'topic_2':
0.15562903530445038, 'topic_3': 0.07977231816943249, 'topic_4': 0.09750072198866887, 'topic_5':
0.09275746822991236, 'topic_6': 0.07091791222601784, 'topic_7': 0.06805293746596487, 'topic_8':
0.07791994258727311, 'topic_9': 0.03548989668315198, 'topic_10': 0.053543911659870706, 'topic_11':
0.060537630322631986, 'topic_12': 0.08621650972489368}
Chat:       112 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_919137566247-1548228361@g.us.chat.log ):
Topics:      {'topic_0': 0.020138467006065888, 'topic_1': 0.013900481878981585, 'topic_2':
0.016114430215227953, 'topic_3': 0.021825244152570724, 'topic_4': 0.01958181183539652, 'topic_5':
0.013791569887756604, 'topic_6': 0.021239909805136424, 'topic_7': 0.019629760935834968, 'topic_8':
0.025788671208086474, 'topic_9': 0.013597769094055249, 'topic_10': 0.026965630405373472, 'topic_11':
0.039298111688630566, 'topic_12': 0.74812814188688}
Chat:       113 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_919321629786-1491165067@g.us.chat.log ):
Topics:      {'topic_0': 0.029806288227046784, 'topic_1': 0.035517903646527835, 'topic_2':
0.09107533354565389, 'topic_3': 0.05478174913428171, 'topic_4': 0.0711126932227357, 'topic_5':
0.06980132377604685, 'topic_6': 0.06015560272840027, 'topic_7': 0.055617054330574664, 'topic_8':
0.05264307800831125, 'topic_9': 0.03281388183099215, 'topic_10': 0.04348187523617083, 'topic_11':
0.04449987279598249, 'topic_12': 0.35869334351727583}
Chat:       114 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_919504171173-1521911266@g.us.chat.log ):
Topics:      {'topic_0': 0.023388486825539144, 'topic_1': 0.03982954772175923, 'topic_2':
0.10403435708350592, 'topic_3': 0.06985401525699252, 'topic_4': 0.04134901737027665, 'topic_5':
0.1716623226916516, 'topic_6': 0.05901244341106565, 'topic_7': 0.08921051953527363, 'topic_8':
0.20297794581987955, 'topic_9': 0.0243070551373931, 'topic_10': 0.07209691375037175, 'topic_11':
0.03745401766603695, 'topic_12': 0.06482335773025732}
Chat:       115 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\

```



```

chat_919567635886-1544363903@g.us.chat.log ):
Topics:    {'topic_0': 0.11325392198518003, 'topic_1': 0.057822059063408204, 'topic_2':
0.09626679217237699, 'topic_3': 0.03579564578404835, 'topic_4': 0.06680488117458047, 'topic_5':
0.07873136507088536, 'topic_6': 0.0534223663029865, 'topic_7': 0.03641345775743078, 'topic_8':
0.07028079946948938, 'topic_9': 0.10130928433014647, 'topic_10': 0.03168744076084862, 'topic_11':
0.08550689417388492, 'topic_12': 0.1727050919547338}
Chat:      116 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_919623418138-1490536344@g.us.chat.log ):
Topics:    {'topic_0': 0.03346230793788969, 'topic_1': 0.060624107216675445, 'topic_2':
0.09626872414512637, 'topic_3': 0.06424540650426444, 'topic_4': 0.0553652089344506, 'topic_5':
0.11162123601329933, 'topic_6': 0.0771908639096527, 'topic_7': 0.09972654167897203, 'topic_8':
0.1157540814839591, 'topic_9': 0.02927794902486958, 'topic_10': 0.054419091236703764, 'topic_11':
0.051429887028635685, 'topic_12': 0.15061459488550188}
Chat:      117 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_919672741666-1560305831@g.us.chat.log ):
Topics:    {'topic_0': 0.012018041765389287, 'topic_1': 0.019709998150889495, 'topic_2':
0.03541657054939467, 'topic_3': 0.01743475836429394, 'topic_4': 0.053303640262662925, 'topic_5':
0.010342773318354977, 'topic_6': 0.03396753722105951, 'topic_7': 0.014984256670515481, 'topic_8':
0.014660909774636218, 'topic_9': 0.010189738751814793, 'topic_10': 0.011102376135947554, 'topic_11':
0.02531463757972995, 'topic_12': 0.7415547614553295}
Chat:      118 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_919737834397-1560511400@g.us.chat.log ):
Topics:    {'topic_0': 0.0388880480911949, 'topic_1': 0.04449675057461553, 'topic_2':
0.07493174208378924, 'topic_3': 0.044887793725282656, 'topic_4': 0.051245154277892696, 'topic_5':
0.0734653742216264, 'topic_6': 0.09075820224288311, 'topic_7': 0.09053250619150882, 'topic_8':
0.0646016563035998, 'topic_9': 0.03871034373317761, 'topic_10': 0.08529803540177268, 'topic_11':
0.04467875296005207, 'topic_12': 0.2575056401926039}
Chat:      119 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_919828556027-1547698927@g.us.chat.log ):
Topics:    {'topic_0': 0.03008490342162241, 'topic_1': 0.041266129047274294, 'topic_2':
0.03207422623848065, 'topic_3': 0.1383962299648093, 'topic_4': 0.04283320827461078, 'topic_5':
0.10407143613869584, 'topic_6': 0.03881821350115531, 'topic_7': 0.06192254287505264, 'topic_8':
0.09536714908955111, 'topic_9': 0.032838547408058874, 'topic_10': 0.03743681997524403, 'topic_11':
0.05820669260238536, 'topic_12': 0.2866839014630593}
Chat:      120 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_919860598106-1522141649@g.us.chat.log ):
Topics:    {'topic_0': 0.08681442528754416, 'topic_1': 0.04249665049984223, 'topic_2':
0.07712756977525694, 'topic_3': 0.05673428378863376, 'topic_4': 0.047014064533340054, 'topic_5':
0.07951865185938876, 'topic_6': 0.09192685246140259, 'topic_7': 0.05177855309301093, 'topic_8':
0.14645103937810092, 'topic_9': 0.08352864460904345, 'topic_10': 0.049405647044299214, 'topic_11':
0.09319691141500912, 'topic_12': 0.09400670625512686}
Chat:      121 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_923056474175@c.us.chat.log ):
Topics:    {}
Chat:      122 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_923056733587-1508906176@g.us.chat.log ):
Topics:    {'topic_0': 0.07667922345375865, 'topic_1': 0.05435547559252089, 'topic_2':
0.06584076744624096, 'topic_3': 0.03760522058593475, 'topic_4': 0.07245653938104059, 'topic_5':
0.06886684634713706, 'topic_6': 0.07679821547062765, 'topic_7': 0.06901119112475328, 'topic_8':

```

```

0.09402313034218895, 'topic_9': 0.10998732711544715, 'topic_10': 0.06760975738601895, 'topic_11':
0.056118527376779304, 'topic_12': 0.15064777837755888}
Chat: 123 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_923106438145@c.us.chat.log ):
Topics: {}
Chat: 124 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_923129261430-1532194872@g.us.chat.log ):
Topics: {'topic_0': 0.04709302340914976, 'topic_1': 0.10886973754775935, 'topic_2':
0.053380061597157295, 'topic_3': 0.0578904601657119, 'topic_4': 0.07011031272169528, 'topic_5':
0.04171030695744363, 'topic_6': 0.06519221005333174, 'topic_7': 0.16914302932040973, 'topic_8':
0.09351284143272955, 'topic_9': 0.042794377271071944, 'topic_10': 0.08293261469020974, 'topic_11':
0.09032294046400709, 'topic_12': 0.07704808436932244}
Chat: 125 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_923147117253-1520597965@g.us.chat.log ):
Topics: {'topic_0': 0.02499564903806895, 'topic_1': 0.04715753078640419, 'topic_2':
0.0344292976740248, 'topic_3': 0.041485370448615166, 'topic_4': 0.043901579506807925, 'topic_5':
0.049138610100419255, 'topic_6': 0.0415327322375914, 'topic_7': 0.034616507722542635, 'topic_8':
0.04730139749769644, 'topic_9': 0.02523225555525226, 'topic_10': 0.03248803326420464, 'topic_11':
0.036616958176836564, 'topic_12': 0.5411040779915357}
Chat: 126 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_923147117253-1528994513@g.us.chat.log ):
Topics: {'topic_0': 0.009909618828072314, 'topic_1': 0.019846061117338897, 'topic_2':
0.006725677214778614, 'topic_3': 0.03415734130698531, 'topic_4': 0.3658166115886853, 'topic_5':
0.28213779761393115, 'topic_6': 0.0921982257396886, 'topic_7': 0.009158264588040931, 'topic_8':
0.045667654222861814, 'topic_9': 0.008065565524566855, 'topic_10': 0.012743075269427485, 'topic_11':
0.012057108744244876, 'topic_12': 0.10151699824137754}
Chat: 127 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_923169964268-1494772094@g.us.chat.log ):
Topics: {'topic_0': 0.053334645459011475, 'topic_1': 0.046366103576143805, 'topic_2':
0.10721700224354737, 'topic_3': 0.0603602416651429, 'topic_4': 0.04236919940574386, 'topic_5':
0.1134578107710153, 'topic_6': 0.09203377448670193, 'topic_7': 0.08174277619600163, 'topic_8':
0.17039960321412495, 'topic_9': 0.05426610987887465, 'topic_10': 0.047173171583756605, 'topic_11':
0.04346980041326738, 'topic_12': 0.08780976110666765}
Chat: 128 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_923214324132-1552761743@g.us.chat.log ):
Topics: {'topic_0': 0.022175691915176814, 'topic_1': 0.04288552648063424, 'topic_2':
0.01624131949628708, 'topic_3': 0.018296479883597894, 'topic_4': 0.01365014402276469, 'topic_5':
0.01980285047083345, 'topic_6': 0.018009147068613736, 'topic_7': 0.014361755255436133, 'topic_8':
0.019210841187022522, 'topic_9': 0.031693632898309544, 'topic_10': 0.017170205558194103, 'topic_11':
0.015827162329976656, 'topic_12': 0.7506752434331534}
Chat: 129 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_923455024491-1564300348@g.us.chat.log ):
Topics: {'topic_0': 0.04280918451432276, 'topic_1': 0.07429621718956202, 'topic_2':
0.06799459229286539, 'topic_3': 0.07085522154532536, 'topic_4': 0.06526182846509174, 'topic_5':
0.07760699504491425, 'topic_6': 0.0747131120842132, 'topic_7': 0.04644001241562877, 'topic_8':
0.14194309757244958, 'topic_9': 0.05059154134773928, 'topic_10': 0.08322101376167487, 'topic_11':
0.05331312312737253, 'topic_12': 0.1509540606388405}
Chat: 130 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_94750947291@c.us.chat.log ):

```

```

Topics: {}
Chat: 131 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
chat_971528904770-1563885486@g.us.chat.log ):
Topics: {'topic_0': 0.0408976482468849, 'topic_1': 0.0543279998712795, 'topic_2': 0.05189927392056664,
'topic_3': 0.052695668478275795, 'topic_4': 0.11533213278671597, 'topic_5': 0.0969182760431297,
'topic_6': 0.06668307686757079, 'topic_7': 0.05623422050577682, 'topic_8': 0.14666174700214024,
'topic_9': 0.051856857896130475, 'topic_10': 0.04563093469266224, 'topic_11': 0.05607725688662358,
'topic_12': 0.16478490680224298}
Chat: 132 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_97338381710-
1487691608@g.us.chat.log ):
Topics: {'topic_0': 0.035026548747295154, 'topic_1': 0.09782768979889599, 'topic_2':
0.03340318667578384, 'topic_3': 0.06579526543968109, 'topic_4': 0.054765793353321504, 'topic_5':
0.056949624312845316, 'topic_6': 0.0566784315873639, 'topic_7': 0.07085769160946359, 'topic_8':
0.24081906426698155, 'topic_9': 0.053587138101831494, 'topic_10': 0.11253236979277438, 'topic_11':
0.05718520127650151, 'topic_12': 0.06457199503726076}
Chat: 133 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_97338381710-
1497431182@g.us.chat.log ):
Topics: {'topic_0': 0.05077061867291657, 'topic_1': 0.06754211643799815, 'topic_2':
0.05496348443426342, 'topic_3': 0.07795555991186758, 'topic_4': 0.05705992844599661, 'topic_5':
0.08409169300948437, 'topic_6': 0.06056779261194949, 'topic_7': 0.07141731609585496, 'topic_8':
0.17020591075411198, 'topic_9': 0.058791626926223646, 'topic_10': 0.06590624097829274, 'topic_11':
0.09353184888683912, 'topic_12': 0.0871958628342004}
Chat: 134 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_97338381710-
1500435936@g.us.chat.log ):
Topics: {'topic_0': 0.022918458259715793, 'topic_1': 0.12050704875533755, 'topic_2':
0.034029543148835724, 'topic_3': 0.08986036764267254, 'topic_4': 0.041153219783417676, 'topic_5':
0.059833148643543846, 'topic_6': 0.04207651818152166, 'topic_7': 0.07222628517653172, 'topic_8':
0.2658879492489572, 'topic_9': 0.03780317097056426, 'topic_10': 0.11784948434677424, 'topic_11':
0.03149627218429171, 'topic_12': 0.06435853365783605}
Chat: 135 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_97338381710-
1502880819@g.us.chat.log ):
Topics: {'topic_0': 0.0504802363852455, 'topic_1': 0.108537368626918, 'topic_2': 0.04199409555647123,
'topic_3': 0.07519377841674928, 'topic_4': 0.06869240514790062, 'topic_5': 0.0709868502591784,
'topic_6': 0.06210362009407054, 'topic_7': 0.05626537814111267, 'topic_8': 0.18656942739124047,
'topic_9': 0.05042528994263827, 'topic_10': 0.08370514958353212, 'topic_11': 0.058449495275406195,
'topic_12': 0.08659690517953611}
Chat: 136 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_97338381710-
1504447471@g.us.chat.log ):
Topics: {'topic_0': 0.02604348752688824, 'topic_1': 0.12052522822716111, 'topic_2':
0.026043521054256565, 'topic_3': 0.09220531495601866, 'topic_4': 0.04705412286312658, 'topic_5':
0.04304900291427784, 'topic_6': 0.05330963780941195, 'topic_7': 0.09313322230181251, 'topic_8':
0.2194890302786533, 'topic_9': 0.03143435293156529, 'topic_10': 0.18295730412705372, 'topic_11':
0.03871229028550557, 'topic_12': 0.02604348472426869}
Chat: 137 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_97338381710-
1507218698@g.us.chat.log ):
Topics: {'topic_0': 0.04487179487179811, 'topic_1': 0.05320519259467866, 'topic_2':
0.0782050433773545, 'topic_3': 0.053205169218060955, 'topic_4': 0.04487179487179765, 'topic_5':
0.044871990684427085, 'topic_6': 0.05911814783643823, 'topic_7': 0.07732156829535834, 'topic_8':
0.3309826886783405, 'topic_9': 0.044871810102318936, 'topic_10': 0.05650898830990822, 'topic_11':

```

```

0.0670940162877223, 'topic_12': 0.044871794871796725}
Chat: 138 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_97338381710-1511318890@g.us.chat.log ):
Topics: {'topic_0': 0.04353410375976718, 'topic_1': 0.06467228218935198, 'topic_2': 0.06020075915598805, 'topic_3': 0.08571738694509157, 'topic_4': 0.06528749122927974, 'topic_5': 0.054656042224001146, 'topic_6': 0.04749582532617589, 'topic_7': 0.08789245522416028, 'topic_8': 0.2007027820070515, 'topic_9': 0.05752322278062008, 'topic_10': 0.081616641652103, 'topic_11': 0.06434846235424566, 'topic_12': 0.08635254515216384}
Chat: 139 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_97338381710-1520734066@g.us.chat.log ):
Topics: {'topic_0': 0.059349192583770166, 'topic_1': 0.07545676634947507, 'topic_2': 0.05612337242827384, 'topic_3': 0.07438448278992715, 'topic_4': 0.07825367747087289, 'topic_5': 0.07429308736432885, 'topic_6': 0.07520251588829403, 'topic_7': 0.06413760547925242, 'topic_8': 0.15179880892351955, 'topic_9': 0.07058485456407941, 'topic_10': 0.08263142298938002, 'topic_11': 0.07283735837055061, 'topic_12': 0.06494685479827467}
Chat: 140 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/chat_994557753110-1563735846@g.us.chat.log ):
Topics: {'topic_0': 0.03857602610950535, 'topic_1': 0.0635085244923565, 'topic_2': 0.1354701878987223, 'topic_3': 0.0591702907383873, 'topic_4': 0.05821420615866853, 'topic_5': 0.07331276722760713, 'topic_6': 0.13610461334789845, 'topic_7': 0.07759063701003364, 'topic_8': 0.07174019414188094, 'topic_9': 0.034303718108555516, 'topic_10': 0.08051221824245966, 'topic_11': 0.04419900814686251, 'topic_12': 0.1272976083770602}
Chat: 141 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/safechat_12053466425-1552762183@g.us.chat.log ):
Topics: {'topic_0': 0.06166613800817827, 'topic_1': 0.09502185679233256, 'topic_2': 0.05490438392170375, 'topic_3': 0.059652800220307986, 'topic_4': 0.07607823998902044, 'topic_5': 0.08596123165422052, 'topic_6': 0.08323256938615957, 'topic_7': 0.05374400499650559, 'topic_8': 0.07578420768929098, 'topic_9': 0.09866434933849817, 'topic_10': 0.04504959189017306, 'topic_11': 0.07147899470202068, 'topic_12': 0.13876163141159176}
Chat: 142 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/safechat_12064721374-1552734205@g.us.chat.log ):
Topics: {'topic_0': 0.03695090709576843, 'topic_1': 0.04534883808351444, 'topic_2': 0.040826878783841586, 'topic_3': 0.07299741169071473, 'topic_4': 0.06149871548001097, 'topic_5': 0.07950726649687308, 'topic_6': 0.06240309972490613, 'topic_7': 0.04082688154385174, 'topic_8': 0.070115504989458721, 'topic_9': 0.04819234229419052, 'topic_10': 0.045348836573637, 'topic_11': 0.07682234057434846, 'topic_12': 0.31912143176375535}
Chat: 143 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/safechat_12533139925-1533654308@g.us.chat.log ):
Topics: {'topic_0': 0.04547008547008632, 'topic_1': 0.045470125681205266, 'topic_2': 0.04547008946366241, 'topic_3': 0.04547009347502317, 'topic_4': 0.09547007437071481, 'topic_5': 0.04547008880261589, 'topic_6': 0.06246842865046474, 'topic_7': 0.0843589270467506, 'topic_8': 0.08991455436613968, 'topic_9': 0.04547010156123135, 'topic_10': 0.056249507330315786, 'topic_11': 0.05380341198103347, 'topic_12': 0.2849145118007565}
Chat: 144 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample/safechat_13159493958-1518354992@g.us.chat.log ):
Topics: {'topic_0': 0.05382933560028872, 'topic_1': 0.06971656532189115, 'topic_2': 0.0571810447201035, 'topic_3': 0.06491572835909329, 'topic_4': 0.07833414122164188, 'topic_5': 0.08990394474680152, 'topic_6': 0.10539852412583177, 'topic_7': 0.07662941897303964, 'topic_8': 0.11039365521172871, 'topic_9': 0.056213794020284806, 'topic_10': 0.0662336830151582, 'topic_11':

```

```

0.06790075525660878, 'topic_12': 0.10334940942754073}
Chat: 145 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_13159493958-1518355216@g.us.chat.log ):
Topics: {'topic_0': 0.11658637299320518, 'topic_1': 0.051933526435297576, 'topic_2':
0.04655633975943668, 'topic_3': 0.050223787709580164, 'topic_4': 0.04345547383288667, 'topic_5':
0.06054431527398959, 'topic_6': 0.11945371000443716, 'topic_7': 0.051538633386689, 'topic_8':
0.06842275115041643, 'topic_9': 0.11872497048246362, 'topic_10': 0.042180719970384174, 'topic_11':
0.12607931831097777, 'topic_12': 0.10430008069023664}
Chat: 146 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_13159933205-1542629155@g.us.chat.log ):
Topics: {'topic_0': 0.05384861904685516, 'topic_1': 0.06531780704186532, 'topic_2':
0.06310177251257917, 'topic_3': 0.06547276756302935, 'topic_4': 0.07400138620545332, 'topic_5':
0.08058038207775756, 'topic_6': 0.11916986072599896, 'topic_7': 0.0690455376086652, 'topic_8':
0.09511126476541656, 'topic_9': 0.056142051618643, 'topic_10': 0.06314040376546423, 'topic_11':
0.06255557096603998, 'topic_12': 0.1325125761022372}
Chat: 147 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_13317012408-1506449192@g.us.chat.log ):
Topics: {'topic_0': 0.19900904609737563, 'topic_1': 0.03627053687837631, 'topic_2':
0.07921928767385987, 'topic_3': 0.03317949816301774, 'topic_4': 0.04147962900295438, 'topic_5':
0.0906249157152055, 'topic_6': 0.050955679850165386, 'topic_7': 0.06338560290330623, 'topic_8':
0.031187480139643314, 'topic_9': 0.22685599588435038, 'topic_10': 0.031187465987259657, 'topic_11':
0.07797582603180343, 'topic_12': 0.03866903567268412}
Chat: 148 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_13317012408-1506462608@g.us.chat.log ):
Topics: {'topic_0': 0.05900666184275537, 'topic_1': 0.07782371975333932, 'topic_2':
0.05845078749329335, 'topic_3': 0.06633805197480393, 'topic_4': 0.0547899854474941, 'topic_5':
0.1037280047778144, 'topic_6': 0.055416244508812114, 'topic_7': 0.05212443145399018, 'topic_8':
0.12430709669611757, 'topic_9': 0.09425498058232781, 'topic_10': 0.059694911431174016, 'topic_11':
0.05032740782999212, 'topic_12': 0.1437377162080866}
Chat: 149 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_14087211456-1539971560@g.us.chat.log ):
Topics: {'topic_0': 0.04283957162317565, 'topic_1': 0.0545109410530316, 'topic_2':
0.05286720540452155, 'topic_3': 0.058732235907930384, 'topic_4': 0.07620729873856526, 'topic_5':
0.06608216422315842, 'topic_6': 0.06676097680132306, 'topic_7': 0.0630849935058022, 'topic_8':
0.07294110825426639, 'topic_9': 0.051271200181926366, 'topic_10': 0.04977470918760814, 'topic_11':
0.04825751996368077, 'topic_12': 0.2966700751550096}
Chat: 150 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_14803950754-1529862422@g.us.chat.log ):
Topics: {'topic_0': 0.07692307692307694, 'topic_1': 0.07692307692307694, 'topic_2':
0.07692307692307694, 'topic_3': 0.07692307692307694, 'topic_4': 0.07692307692307694, 'topic_5':
0.07692307692307694, 'topic_6': 0.07692307692307694, 'topic_7': 0.07692307692307694, 'topic_8':
0.07692307692307694, 'topic_9': 0.07692307692307694, 'topic_10': 0.07692307692307694, 'topic_11':
0.07692307692307694, 'topic_12': 0.07692307692307694}
Chat: 151 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_17122506796-1545616164@g.us.chat.log ):
Topics: {'topic_0': 0.0429423711601952, 'topic_1': 0.08988851221738133, 'topic_2':
0.08692884123512071, 'topic_3': 0.06405588793112908, 'topic_4': 0.0726453479896227, 'topic_5':
0.05944405897218271, 'topic_6': 0.09987309355868952, 'topic_7': 0.05614369347818533, 'topic_8':
0.10812388554905031, 'topic_9': 0.05944403319036257, 'topic_10': 0.04624272228733597, 'topic_11':

```

```

0.0775958140319337, 'topic_12': 0.13667173839881058}
Chat: 152 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_17753823154-1545637451@g.us.chat.log ):
Topics: {'topic_0': 0.13730629523675592, 'topic_1': 0.06876332463226471, 'topic_2':
0.04479853218341675, 'topic_3': 0.031092658836317816, 'topic_4': 0.021030476597840072, 'topic_5':
0.041773073359875804, 'topic_6': 0.043034753058690936, 'topic_7': 0.03825027881577171, 'topic_8':
0.061252449128233244, 'topic_9': 0.1604034456586588, 'topic_10': 0.045230023037814746, 'topic_11':
0.1451315817367697, 'topic_12': 0.161933107717591}
Chat: 153 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_17863435548-1534713807@g.us.chat.log ):
Topics: {}
Chat: 154 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_19793987619@c.us.chat.log ):
Topics: {}
Chat: 155 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_201018051555-1493575339@g.us.chat.log ):
Topics: {'topic_0': 0.05202988712653969, 'topic_1': 0.16314105393876516, 'topic_2':
0.042770675169857796, 'topic_3': 0.04607234281454778, 'topic_4': 0.0798076923093652, 'topic_5':
0.0674053292467664, 'topic_6': 0.042770655270658764, 'topic_7': 0.06324366404272706, 'topic_8':
0.17450064132254509, 'topic_9': 0.05202991868929468, 'topic_10': 0.045536135695907434, 'topic_11':
0.052029929412685166, 'topic_12': 0.11866207496033983}
Chat: 156 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_212639422141-1477567255@g.us.chat.log ):
Topics: {'topic_0': 0.13966071256270815, 'topic_1': 0.03906119228345908, 'topic_2':
0.050151056135961464, 'topic_3': 0.03256843491061774, 'topic_4': 0.05012813628980231, 'topic_5':
0.05019768967180477, 'topic_6': 0.05340924170900314, 'topic_7': 0.04957316288352796, 'topic_8':
0.06460188440978779, 'topic_9': 0.15766668341433826, 'topic_10': 0.03357163719709214, 'topic_11':
0.15658906907333023, 'topic_12': 0.12282109945856887}
Chat: 157 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_212657458806-1520152237@g.us.chat.log ):
Topics: {'topic_0': 0.13023386885192292, 'topic_1': 0.05690618978064087, 'topic_2':
0.04233348054506402, 'topic_3': 0.06239066608061331, 'topic_4': 0.06703397324747695, 'topic_5':
0.06631467321159953, 'topic_6': 0.04880636268797009, 'topic_7': 0.06917720661058278, 'topic_8':
0.08952977876032855, 'topic_9': 0.11926910651658662, 'topic_10': 0.060585401668202365, 'topic_11':
0.10614823631145007, 'topic_12': 0.08127105572756486}
Chat: 158 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_212707719479-1552688726@g.us.chat.log ):
Topics: {'topic_0': 0.037404344511896226, 'topic_1': 0.09105126562953776, 'topic_2':
0.0696957337743142, 'topic_3': 0.11130210655349523, 'topic_4': 0.04687404136391219, 'topic_5':
0.08787066541782396, 'topic_6': 0.06786051248445235, 'topic_7': 0.16094653270553863, 'topic_8':
0.10484435505574588, 'topic_9': 0.04507302242329308, 'topic_10': 0.04001381080208728, 'topic_11':
0.05066194053934303, 'topic_12': 0.08640166873855991}
Chat: 159 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_22896523148@c.us.chat.log ):
Topics: {}
Chat: 160 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_22965141143@c.us.chat.log ):
Topics: {}
Chat: 161 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\

```

```

safechat_233544217390-1490962626@g.us.chat.log ):
Topics:      {'topic_0': 0.04557494307870227, 'topic_1': 0.06478308190144132, 'topic_2':
0.06353231016899027, 'topic_3': 0.05793106440569049, 'topic_4': 0.06333642740449452, 'topic_5':
0.0807804537878576, 'topic_6': 0.06561309657923921, 'topic_7': 0.0497212295959723, 'topic_8':
0.1872574730865805, 'topic_9': 0.049318573101375315, 'topic_10': 0.05069836891955563, 'topic_11':
0.06408247920554205, 'topic_12': 0.15737049876456288}
Chat:       162 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2347033038665-1563313466@g.us.chat.log ):
Topics:      {'topic_0': 0.041648500953279974, 'topic_1': 0.06421549656877647, 'topic_2':
0.056975238508053375, 'topic_3': 0.06999063010791381, 'topic_4': 0.09939453605151603, 'topic_5':
0.09515611239449696, 'topic_6': 0.058201847193686554, 'topic_7': 0.05148878019250955, 'topic_8':
0.12181492948707238, 'topic_9': 0.0645172990615661, 'topic_10': 0.05873011108304163, 'topic_11':
0.06744779052197827, 'topic_12': 0.15041872787611493}
Chat:       163 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2347034328548-1491921470@g.us.chat.log ):
Topics:      {'topic_0': 0.04700982650733093, 'topic_1': 0.0729579545678963, 'topic_2':
0.12331936306497218, 'topic_3': 0.05025335554782265, 'topic_4': 0.060981889257451544, 'topic_5':
0.12375628336240149, 'topic_6': 0.0674567720959004, 'topic_7': 0.09112159878803695, 'topic_8':
0.07545293410211498, 'topic_9': 0.04551284610854835, 'topic_10': 0.08169046304137141, 'topic_11':
0.055492861117373594, 'topic_12': 0.10499385243877851}
Chat:       164 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2347080462391-1519061348@g.us.chat.log ):
Topics:      {'topic_0': 0.031191002575795553, 'topic_1': 0.17453514600864442, 'topic_2':
0.056167563402398174, 'topic_3': 0.05939102749008419, 'topic_4': 0.05428706341371601, 'topic_5':
0.10490190158028509, 'topic_6': 0.05489397759107446, 'topic_7': 0.11500998157517288, 'topic_8':
0.1057100810953722, 'topic_9': 0.04077671693658169, 'topic_10': 0.040365323277686556, 'topic_11':
0.04150421021079841, 'topic_12': 0.12126600484239024}
Chat:       165 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2348024790559-1559600306@g.us.chat.log ):
Topics:      {'topic_0': 0.04417813815117106, 'topic_1': 0.03980305718443077, 'topic_2':
0.033495404378766476, 'topic_3': 0.04525663350567197, 'topic_4': 0.038214375919590544, 'topic_5':
0.1152226886807821, 'topic_6': 0.03393633211701597, 'topic_7': 0.0655934745491842, 'topic_8':
0.14252048510272364, 'topic_9': 0.047989165833495086, 'topic_10': 0.03641418511531938, 'topic_11':
0.07761863823303088, 'topic_12': 0.27975784104152185}
Chat:       166 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2348032322513-1559282930@g.us.chat.log ):
Topics:      {'topic_0': 0.030609298303862123, 'topic_1': 0.03109758522751515, 'topic_2':
0.03956112444242689, 'topic_3': 0.046448131774186247, 'topic_4': 0.34625311072727694, 'topic_5':
0.11490994578942425, 'topic_6': 0.052031842332721075, 'topic_7': 0.03419002824001439, 'topic_8':
0.05056029990404505, 'topic_9': 0.04687780322349836, 'topic_10': 0.03831507973412253, 'topic_11':
0.03646867159411042, 'topic_12': 0.13267707870679799}
Chat:       167 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2348033488384-1525616235@g.us.chat.log ):
Topics:      {'topic_0': 0.04284448814256268, 'topic_1': 0.10544198708215592, 'topic_2':
0.06559795342841733, 'topic_3': 0.04952414450697886, 'topic_4': 0.04526024402627326, 'topic_5':
0.10354499595615133, 'topic_6': 0.05509744745382767, 'topic_7': 0.10189318588397571, 'topic_8':
0.12885835126977124, 'topic_9': 0.05394860685113268, 'topic_10': 0.05680718464321768, 'topic_11':
0.08078922740210764, 'topic_12': 0.11039218335342833}
Chat:       168 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\

```

```

safechat_2348054135475-1563140992@g.us.chat.log ):
Topics:      {'topic_0': 0.024148613291473532, 'topic_1': 0.034815327576645926, 'topic_2':
0.027196244190158616, 'topic_3': 0.03923512778853056, 'topic_4': 0.04802451288560743, 'topic_5':
0.11337580102392075, 'topic_6': 0.03405337448011582, 'topic_7': 0.05902980457148096, 'topic_8':
0.11880050145162045, 'topic_9': 0.031767673807717285, 'topic_10': 0.04652957663338628, 'topic_11':
0.0619241559773206, 'topic_12': 0.3610992863220221}
Chat:       169 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2348062137437-1561882666@g.us.chat.log ):
Topics:      {'topic_0': 0.047679675101846855, 'topic_1': 0.07842809573281169, 'topic_2':
0.0476796924553191, 'topic_3': 0.04709283482571947, 'topic_4': 0.04767967150975789, 'topic_5':
0.15759496590402727, 'topic_6': 0.05237451642201372, 'topic_7': 0.06411158280423021, 'topic_8':
0.12218407538266132, 'topic_9': 0.0792013304319477, 'topic_10': 0.09345433067956674, 'topic_11':
0.05310247469663055, 'topic_12': 0.10941675405346711}
Chat:       170 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2348063025256-1559096470@g.us.chat.log ):
Topics:      {'topic_0': 0.05064102564103037, 'topic_1': 0.050641050782390576, 'topic_2':
0.050641025641029125, 'topic_3': 0.05064105346369404, 'topic_4': 0.10064102564102888, 'topic_5':
0.050641110607365945, 'topic_6': 0.05064102564103177, 'topic_7': 0.05064102594745632, 'topic_8':
0.13397424895210602, 'topic_9': 0.05064103375400292, 'topic_10': 0.05064105058865531, 'topic_11':
0.18397430611858218, 'topic_12': 0.125641017221626668}
Chat:       171 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2348065669534-1541854404@g.us.chat.log ):
Topics:      {'topic_0': 0.05807829130097945, 'topic_1': 0.07078435066592269, 'topic_2':
0.04620402363895782, 'topic_3': 0.04024593959558749, 'topic_4': 0.03620904490669204, 'topic_5':
0.07128823094650091, 'topic_6': 0.032897186571403794, 'topic_7': 0.11192515617518509, 'topic_8':
0.06760679614734448, 'topic_9': 0.07895609275553193, 'topic_10': 0.03579520529433784, 'topic_11':
0.06798231442736954, 'topic_12': 0.28202736757418967}
Chat:       172 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2348109890004-1563317908@g.us.chat.log ):
Topics:      {'topic_0': 0.04514194908455369, 'topic_1': 0.06268583760133634, 'topic_2':
0.0661210409680303, 'topic_3': 0.07892474937533876, 'topic_4': 0.0714577177099937, 'topic_5':
0.07722725310690698, 'topic_6': 0.07521711557368171, 'topic_7': 0.05954324853019755, 'topic_8':
0.23271407428101515, 'topic_9': 0.05599563664780491, 'topic_10': 0.06644518316757686, 'topic_11':
0.05767330892307096, 'topic_12': 0.05085288503049301}
Chat:       173 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2348110107300-1543757473@g.us.chat.log ):
Topics:      {'topic_0': 0.009995736186215258, 'topic_1': 0.009995751563658534, 'topic_2':
0.009995736186214547, 'topic_3': 0.009995744851713975, 'topic_4': 0.02079820417368764, 'topic_5':
0.009995743899577924, 'topic_6': 0.00999573618621608, 'topic_7': 0.012979409632141788, 'topic_8':
0.06441672030262915, 'topic_9': 0.009995748007190967, 'topic_10': 0.01840542646862452, 'topic_11':
0.023774380559944238, 'topic_12': 0.7896556619821861}
Chat:       174 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2348115506965-1549270749@g.us.chat.log ):
Topics:      {'topic_0': 0.0298085173260959, 'topic_1': 0.09296204165597283, 'topic_2':
0.08330444140071318, 'topic_3': 0.08384564021237048, 'topic_4': 0.029233738636579597, 'topic_5':
0.045063952550664636, 'topic_6': 0.07079749420343753, 'topic_7': 0.1515619300470797, 'topic_8':
0.15627303117025643, 'topic_9': 0.06139749586699896, 'topic_10': 0.03569814021607313, 'topic_11':
0.0726685174355489, 'topic_12': 0.08738505927820868}
Chat:       175 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\

```



```

safechat_2348145686926-1564015604@g.us.chat.log ):
Topics:    {'topic_0': 0.057692307692308445, 'topic_1': 0.057692307692308584, 'topic_2':
0.05769235482729573, 'topic_3': 0.05769230769230846, 'topic_4': 0.057692307692308334, 'topic_5':
0.2120618508812987, 'topic_6': 0.15332266297250857, 'topic_7': 0.05769236208812235, 'topic_8':
0.05769230769230816, 'topic_9': 0.05769230769230811, 'topic_10': 0.057692307692308424, 'topic_11':
0.057692307692308195, 'topic_12': 0.057692307692308126}
Chat:      176 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2348147250000-1553249631@g.us.chat.log ):
Topics:    {'topic_0': 0.20315594935982204, 'topic_1': 0.027230022927734056, 'topic_2':
0.024143585473714237, 'topic_3': 0.014267067730589151, 'topic_4': 0.01797076774923692, 'topic_5':
0.04636579327096373, 'topic_6': 0.01673619976202053, 'topic_7': 0.020439907783834633, 'topic_8':
0.04266211223798661, 'topic_9': 0.4459028681003166, 'topic_10': 0.021674470735220226, 'topic_11':
0.07132175109871708, 'topic_12': 0.04812950376984399}
Chat:      177 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2348149460003-1547065630@g.us.chat.log ):
Topics:    {'topic_0': 0.034893448758608975, 'topic_1': 0.13782694441924495, 'topic_2':
0.06149819874972786, 'topic_3': 0.08312459400061654, 'topic_4': 0.048952050956738165, 'topic_5':
0.10150932134097647, 'topic_6': 0.05094498143171982, 'topic_7': 0.09654605335086217, 'topic_8':
0.09996921753824424, 'topic_9': 0.04125884052178421, 'topic_10': 0.0700659296017407, 'topic_11':
0.07899998458334447, 'topic_12': 0.0944104347463911}
Chat:      178 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2348155594783-1563117947@g.us.chat.log ):
Topics:    {'topic_0': 0.028762663482530388, 'topic_1': 0.03229974035788021, 'topic_2':
0.04127954605585715, 'topic_3': 0.048995866828664913, 'topic_4': 0.4261729843841112, 'topic_5':
0.07094825152091445, 'topic_6': 0.06449841468016111, 'topic_7': 0.031835845757454184, 'topic_8':
0.03291617757680862, 'topic_9': 0.032096069537341024, 'topic_10': 0.03680445935238287, 'topic_11':
0.03239404730970061, 'topic_12': 0.12099593315619818}
Chat:      179 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2348160741717-1468963101@g.us.chat.log ):
Topics:    {'topic_0': 0.055986582453001775, 'topic_1': 0.020105417933302557, 'topic_2':
0.02355605321999167, 'topic_3': 0.04208626581060714, 'topic_4': 0.028386966110979066, 'topic_5':
0.022069150673165488, 'topic_6': 0.04199353365851924, 'topic_7': 0.026028162111957652, 'topic_8':
0.05821786253351732, 'topic_9': 0.018863189454456218, 'topic_10': 0.0354198677317228, 'topic_11':
0.1232109925024506, 'topic_12': 0.5040759558063287}
Chat:      180 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2348189583114-1517124754@g.us.chat.log ):
Topics:    {'topic_0': 0.04057208266516334, 'topic_1': 0.11439029929947457, 'topic_2':
0.08173541246295356, 'topic_3': 0.033180320227470375, 'topic_4': 0.06168886179549427, 'topic_5':
0.058226904465827234, 'topic_6': 0.07381616827502341, 'topic_7': 0.1727649026297284, 'topic_8':
0.1196739403224022, 'topic_9': 0.03760783914111775, 'topic_10': 0.03036130534211928, 'topic_11':
0.07565120700451285, 'topic_12': 0.10033075636871264}
Chat:      181 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2349079721832@c.us.chat.log ):
Topics:    {}
Chat:      182 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2349086550164-1556835426@g.us.chat.log ):
Topics:    {'topic_0': 0.04915124003940741, 'topic_1': 0.08159964491510575, 'topic_2':
0.0530843802388157, 'topic_3': 0.07958387658766967, 'topic_4': 0.05857408125543721, 'topic_5':
0.09032198318072643, 'topic_6': 0.11116748562648726, 'topic_7': 0.05455932192880767, 'topic_8':

```

```

0.1876250184993112, 'topic_9': 0.05306362881759717, 'topic_10': 0.06635871891561708, 'topic_11':
0.06279547879791374, 'topic_12': 0.052115141197103754}
Chat: 183 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_2349098643869-1562995313@g.us.chat.log ):
Topics: {'topic_0': 0.05836773976847094, 'topic_1': 0.07936124636479444, 'topic_2':
0.06242638680426185, 'topic_3': 0.08150062640834611, 'topic_4': 0.0783289040110727, 'topic_5':
0.07677723068210307, 'topic_6': 0.07351750479277155, 'topic_7': 0.06172195190846472, 'topic_8':
0.1225115340058728, 'topic_9': 0.0652807106866656, 'topic_10': 0.06637196197691532, 'topic_11':
0.08474150749582537, 'topic_12': 0.08909269509442554}
Chat: 184 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_23754804184-1517662089@g.us.chat.log ):
Topics: {'topic_0': 0.07024443749366353, 'topic_1': 0.07325761994694986, 'topic_2':
0.06017656489405535, 'topic_3': 0.06289979161596113, 'topic_4': 0.08575425258628845, 'topic_5':
0.12632387736248452, 'topic_6': 0.05715591764018646, 'topic_7': 0.0468922763017512, 'topic_8':
0.11093701944486062, 'topic_9': 0.0773973200583466, 'topic_10': 0.05776997432641949, 'topic_11':
0.08156933115084684, 'topic_12': 0.08962161717819259}
Chat: 185 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_237651690064-1552403045@g.us.chat.log ):
Topics: {'topic_0': 0.04541354440876477, 'topic_1': 0.08609311649661441, 'topic_2':
0.07024251814892957, 'topic_3': 0.055528436987591676, 'topic_4': 0.07571579989948635, 'topic_5':
0.08204930542666894, 'topic_6': 0.08258218035796411, 'topic_7': 0.06025330696895078, 'topic_8':
0.09941930893638433, 'topic_9': 0.04307178507437839, 'topic_10': 0.0911400179554361, 'topic_11':
0.07407420137848565, 'topic_12': 0.1344164779603468}
Chat: 186 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_254706773827-1534684758@g.us.chat.log ):
Topics: {'topic_0': 0.06110729981110713, 'topic_1': 0.07314665718837066, 'topic_2':
0.06486903253914447, 'topic_3': 0.057102659003040304, 'topic_4': 0.06871970716919842, 'topic_5':
0.10853645465771522, 'topic_6': 0.07025308783546043, 'topic_7': 0.0912294402784805, 'topic_8':
0.10338706520919073, 'topic_9': 0.06018533927262865, 'topic_10': 0.0643243697372553, 'topic_11':
0.08865659752036117, 'topic_12': 0.08848228977804605}
Chat: 187 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_254707936002-1541854957@g.us.chat.log ):
Topics: {'topic_0': 0.09148780926012964, 'topic_1': 0.1085761536752691, 'topic_2': 0.1181493321576024,
'topic_3': 0.021175348694301124, 'topic_4': 0.02898780912389403, 'topic_5': 0.07846694563226106,
'topic_6': 0.021175321182224253, 'topic_7': 0.03680041797966082, 'topic_8': 0.06414398804058517,
'topic_9': 0.11297226446160194, 'topic_10': 0.021175312464778502, 'topic_11': 0.10636875530744913,
'topic_12': 0.19052054202024282}
Chat: 188 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_254757146467-1563438096@g.us.chat.log ):
Topics: {'topic_0': 0.045572557669333345, 'topic_1': 0.060447109442927874, 'topic_2':
0.0487983676284719, 'topic_3': 0.053637089170972825, 'topic_4': 0.2055186998831128, 'topic_5':
0.07516017662453632, 'topic_6': 0.07887598674011445, 'topic_7': 0.04754536134758767, 'topic_8':
0.07342222505810936, 'topic_9': 0.04467656180033383, 'topic_10': 0.0575797119192102, 'topic_11':
0.06044710004177958, 'topic_12': 0.14831905267351012}
Chat: 189 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_254798513186-1558264650@g.us.chat.log ):
Topics: {'topic_0': 0.04030387035830863, 'topic_1': 0.06835812174811515, 'topic_2':
0.05551718043122136, 'topic_3': 0.060393615381621434, 'topic_4': 0.08465916218582457, 'topic_5':
0.08251076187189062, 'topic_6': 0.0755805758007093, 'topic_7': 0.043214053577970306, 'topic_8':

```

```

0.061837249946634844, 'topic_9': 0.04219155805438677, 'topic_10': 0.05725629848365844, 'topic_11':
0.0655975027681079, 'topic_12': 0.2625800493915506}
Chat: 190 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_255620150767-1527782404@g.us.chat.log ):
Topics: {'topic_0': 0.03139194139194476, 'topic_1': 0.14038408804802832, 'topic_2':
0.15192358726553143, 'topic_3': 0.04567763706475314, 'topic_4': 0.03139197909321604, 'topic_5':
0.045677680912689034, 'topic_6': 0.09329669810244556, 'topic_7': 0.0790109498789234, 'topic_8':
0.16710612296911492, 'topic_9': 0.05996345851901026, 'topic_10': 0.031391972577973, 'topic_11':
0.045677659030081134, 'topic_12': 0.07710622514628893}
Chat: 191 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_255717713926-1562782840@g.us.chat.log ):
Topics: {'topic_0': 0.03556280969651219, 'topic_1': 0.04771639487491656, 'topic_2':
0.04888248010611947, 'topic_3': 0.05875771617800645, 'topic_4': 0.039086937293008696, 'topic_5':
0.06987456079667913, 'topic_6': 0.08450869137430352, 'topic_7': 0.14096755274927625, 'topic_8':
0.0664137211408701, 'topic_9': 0.039211609297472715, 'topic_10': 0.0771383432460915, 'topic_11':
0.06541428048923732, 'topic_12': 0.2264649027575092}
Chat: 192 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_260977284225-1550949072@g.us.chat.log ):
Topics: {'topic_0': 0.12041941361858832, 'topic_1': 0.0664289297010204, 'topic_2':
0.056065351163283365, 'topic_3': 0.040201316501345535, 'topic_4': 0.03932320959391162, 'topic_5':
0.06053635441745232, 'topic_6': 0.04603461686744576, 'topic_7': 0.04451676072989499, 'topic_8':
0.05432021129810809, 'topic_9': 0.23675051431474134, 'topic_10': 0.0197482741049089, 'topic_11':
0.0817489089549965, 'topic_12': 0.13390613873430293}
Chat: 193 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_263772749518-1460358881@g.us.chat.log ):
Topics: {'topic_0': 0.05854717675768037, 'topic_1': 0.0771917628003944, 'topic_2':
0.06596562241782775, 'topic_3': 0.06663148597924362, 'topic_4': 0.08662339285579111, 'topic_5':
0.08007908149616129, 'topic_6': 0.07007981902739926, 'topic_7': 0.06133876196651724, 'topic_8':
0.13299677426047077, 'topic_9': 0.07423989905898982, 'topic_10': 0.07502638921422124, 'topic_11':
0.07334000604394672, 'topic_12': 0.07793982812137144}
Chat: 194 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_263778888125-1488105200@g.us.chat.log ):
Topics: {'topic_0': 0.03125693592664198, 'topic_1': 0.04074820792103795, 'topic_2':
0.04052041436919542, 'topic_3': 0.04861113698299422, 'topic_4': 0.4412628872115454, 'topic_5':
0.06362294955277623, 'topic_6': 0.08104180034129488, 'topic_7': 0.03125694316337346, 'topic_8':
0.0366479863500117, 'topic_9': 0.031920536011112596, 'topic_10': 0.04636702317933276, 'topic_11':
0.036002569765011025, 'topic_12': 0.0707406092256834}
Chat: 195 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_27621634355-1529677872@g.us.chat.log ):
Topics: {'topic_0': 0.03681770092785816, 'topic_1': 0.09754002652786327, 'topic_2':
0.06821420601828064, 'topic_3': 0.08763065739725992, 'topic_4': 0.05674675466528446, 'topic_5':
0.08512648096502629, 'topic_6': 0.062224195136852276, 'topic_7': 0.20867375611006506, 'topic_8':
0.10580843052697232, 'topic_9': 0.05307792662125112, 'topic_10': 0.04088274418586589, 'topic_11':
0.042915267757911185, 'topic_12': 0.05434185315950939}
Chat: 196 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_27644635125-1509175015@g.us.chat.log ):
Topics: {'topic_0': 0.021876571644982686, 'topic_1': 0.02187657530011045, 'topic_2':
0.02187661941213068, 'topic_3': 0.056570849268949015, 'topic_4': 0.034248186123187134, 'topic_5':
0.13709589585907953, 'topic_6': 0.021876614655133055, 'topic_7': 0.11785849821714174, 'topic_8':

```

```

0.09330503872038567, 'topic_9': 0.021876619961186877, 'topic_10': 0.1338115397277814, 'topic_11':
0.0814003733119594, 'topic_12': 0.23632661779797243}
Chat: 197 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_31621943195-1557752118@g.us.chat.log ):
Topics: {'topic_0': 0.102977963342044, 'topic_1': 0.129937942141451, 'topic_2': 0.05388828863170963,
'topic_3': 0.05032986453361192, 'topic_4': 0.062300388913023586, 'topic_5': 0.05813532407109238,
'topic_6': 0.054233030672076285, 'topic_7': 0.04789910760729163, 'topic_8': 0.05488187348311356,
'topic_9': 0.05546784168847062, 'topic_10': 0.19278654821458388, 'topic_11': 0.05117302293995754,
'topic_12': 0.08598880376157884}
Chat: 198 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_353862228246-1543459639@g.us.chat.log ):
Topics: {'topic_0': 0.050170308201717764, 'topic_1': 0.08818199681424019, 'topic_2':
0.05017030576098217, 'topic_3': 0.04724634211516811, 'topic_4': 0.06329196738368774, 'topic_5':
0.04578434274383147, 'topic_6': 0.04703746620919295, 'topic_7': 0.0501703088819624, 'topic_8':
0.1989115125447641, 'topic_9': 0.06490793355115082, 'topic_10': 0.0484994895741701, 'topic_11':
0.05163232154382046, 'topic_12': 0.19399570467531135}
Chat: 199 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_353862228246-1543580117@g.us.chat.log ):
Topics: {}
Chat: 200 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_353862228246-1544539545@g.us.chat.log ):
Topics: {'topic_0': 0.04577089917918993, 'topic_1': 0.07004047649671677, 'topic_2':
0.06142191440615778, 'topic_3': 0.0626706418357846, 'topic_4': 0.07257742721502558, 'topic_5':
0.06191363041552049, 'topic_6': 0.05909091514004653, 'topic_7': 0.04810192512659541, 'topic_8':
0.16536676001310546, 'topic_9': 0.0545003596984771, 'topic_10': 0.045770906795966376, 'topic_11':
0.0547721538263011, 'topic_12': 0.19800198985111245}
Chat: 201 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_353862228246-1544601100@g.us.chat.log ):
Topics: {'topic_0': 0.04935541310541526, 'topic_1': 0.07310543015609607, 'topic_2':
0.059077637644699914, 'topic_3': 0.0667742437673306, 'topic_4': 0.061855412282483466, 'topic_5':
0.06463319501306745, 'topic_6': 0.058522084868535065, 'topic_7': 0.05977208309624677, 'topic_8':
0.12025647793830369, 'topic_9': 0.07831068967055148, 'topic_10': 0.08376363561233625, 'topic_11':
0.05936882275952615, 'topic_12': 0.16520487408540735}
Chat: 202 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_447448867845-1532456432@g.us.chat.log ):
Topics: {'topic_0': 0.050269421250529286, 'topic_1': 0.0627492988527325, 'topic_2':
0.06254799212639817, 'topic_3': 0.08484699138203917, 'topic_4': 0.0665737591484225, 'topic_5':
0.07393536550725426, 'topic_6': 0.0693917975433141, 'topic_7': 0.057113235831040325, 'topic_8':
0.19025842446365748, 'topic_9': 0.06307738824598683, 'topic_10': 0.0606770684706513, 'topic_11':
0.08019000399694685, 'topic_12': 0.07836925318102743}
Chat: 203 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_447451246621-1564038749@g.us.chat.log ):
Topics: {'topic_0': 0.040714859985940974, 'topic_1': 0.04778554840457395, 'topic_2':
0.05889662832133466, 'topic_3': 0.057886528835312846, 'topic_4': 0.07808857808799738, 'topic_5':
0.058391606479032965, 'topic_6': 0.060608616672719766, 'topic_7': 0.0407148520580906, 'topic_8':
0.0612251261348735, 'topic_9': 0.04576536010585162, 'topic_10': 0.040714851482438016, 'topic_11':
0.04677544677544598, 'topic_12': 0.362431996656388}
Chat: 204 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_447708329340-1560552735@g.us.chat.log ):

```

```

Topics:      {'topic_0': 0.028948874316848654, 'topic_1': 0.033123976216870604, 'topic_2':
0.042183597732217255, 'topic_3': 0.04717513705734002, 'topic_4': 0.4266008831911092, 'topic_5':
0.07234754329097977, 'topic_6': 0.06262004155742616, 'topic_7': 0.029651095870053615, 'topic_8':
0.044531160428777784, 'topic_9': 0.03447620358510394, 'topic_10': 0.037959448446082664, 'topic_11':
0.03024117005487384, 'topic_12': 0.11014086825231707}
Chat:       205 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_447887810701-1493490356@g.us.chat.log ):
Topics:      {'topic_0': 0.03166666666666927, 'topic_1': 0.03166666666666977, 'topic_2':
0.0316667112789995, 'topic_3': 0.04355269302874172, 'topic_4': 0.031666666666668904, 'topic_5':
0.07791298116483665, 'topic_6': 0.03166668714931444, 'topic_7': 0.1553694244367623, 'topic_8':
0.18166667031527514, 'topic_9': 0.0316666814710647, 'topic_10': 0.04162239619756769, 'topic_11':
0.13166664524705177, 'topic_12': 0.17820910971037815}
Chat:       206 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_4917629460203-1551197253@g.us.chat.log ):
Topics:      {'topic_0': 0.04403454592196724, 'topic_1': 0.07089691672329447, 'topic_2':
0.057084982016668756, 'topic_3': 0.0686380949085038, 'topic_4': 0.10964395713943993, 'topic_5':
0.08055695757416584, 'topic_6': 0.09862504397088313, 'topic_7': 0.0627435195480187, 'topic_8':
0.06537206877975728, 'topic_9': 0.04628823394807663, 'topic_10': 0.057240361908193337, 'topic_11':
0.07474452792135915, 'topic_12': 0.16413078963967598}
Chat:       207 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_50762850233-1508010876@g.us.chat.log ):
Topics:      {'topic_0': 0.04325969605034556, 'topic_1': 0.06132496158153074, 'topic_2':
0.13183779005255106, 'topic_3': 0.07472506701928298, 'topic_4': 0.07356269493991671, 'topic_5':
0.16330801349238602, 'topic_6': 0.06831796474816536, 'topic_7': 0.06831798538373134, 'topic_8':
0.082863030282269, 'topic_9': 0.05141822962471821, 'topic_10': 0.047309180074749636, 'topic_11':
0.047052318902256245, 'topic_12': 0.08670306784809637}
Chat:       208 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_5492914734432-1550429334@g.us.chat.log ):
Topics:      {'topic_0': 0.1707305871703915, 'topic_1': 0.04073159438594858, 'topic_2':
0.03263317278788722, 'topic_3': 0.01711235780680495, 'topic_4': 0.017488650674190417, 'topic_5':
0.03340924317185376, 'topic_6': 0.022847841428370406, 'topic_7': 0.0659896788695844, 'topic_8':
0.07539734930096219, 'topic_9': 0.20934563945987067, 'topic_10': 0.01469914927954827, 'topic_11':
0.18769949870678712, 'topic_12': 0.11191523695780119}
Chat:       209 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_5512996817977-1513419479@g.us.chat.log ):
Topics:      {'topic_0': 0.035775229947049285, 'topic_1': 0.06572758634003602, 'topic_2':
0.0410826991314519, 'topic_3': 0.04253235525164121, 'topic_4': 0.043863030174868985, 'topic_5':
0.06052903478123242, 'topic_6': 0.04558730637547698, 'topic_7': 0.243312911706381, 'topic_8':
0.1619310498590377, 'topic_9': 0.040023329177408855, 'topic_10': 0.04758256736935154, 'topic_11':
0.056790058220608064, 'topic_12': 0.11526284166545679}
Chat:       210 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_5521996180760-1540595805@g.us.chat.log ):
Topics:      {'topic_0': 0.04640187566712687, 'topic_1': 0.062026876611311406, 'topic_2':
0.05161020208600872, 'topic_3': 0.06115878872257396, 'topic_4': 0.09129273308320372, 'topic_5':
0.06463102172560137, 'topic_6': 0.13928378899488983, 'topic_7': 0.08806852993236437, 'topic_8':
0.13816773317115463, 'topic_9': 0.0559504651030392, 'topic_10': 0.05681852997899717, 'topic_11':
0.06029077056195048, 'topic_12': 0.08429868436177797}
Chat:       211 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_553497819767-1480261843@g.us.chat.log ):

```

```

Topics:      {'topic_0': 0.16716102865030713, 'topic_1': 0.03505450726651779, 'topic_2':
0.0410996518404051, 'topic_3': 0.02986662032663044, 'topic_4': 0.02550932183765635, 'topic_5':
0.05006076084857446, 'topic_6': 0.04102310922491738, 'topic_7': 0.038421605972825915, 'topic_8':
0.0348349843757261, 'topic_9': 0.21848988975802117, 'topic_10': 0.03120374531896401, 'topic_11':
0.21011493676334472, 'topic_12': 0.07715983781611296}
Chat:       212 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_556295153706-1483193085@g.us.chat.log ):
Topics:      {'topic_0': 0.04000803465740856, 'topic_1': 0.04597682063304256, 'topic_2':
0.06836328319446183, 'topic_3': 0.04138544121920178, 'topic_4': 0.058076416262087545, 'topic_5':
0.08726375189743696, 'topic_6': 0.12069736414498872, 'topic_7': 0.05344522550646721, 'topic_8':
0.08107271592967318, 'topic_9': 0.03813292758459947, 'topic_10': 0.04395011866915751, 'topic_11':
0.04435923591901614, 'topic_12': 0.2772686643824594}
Chat:       213 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_601126316671-1507101918@g.us.chat.log ):
Topics:      {'topic_0': 0.04885566656492227, 'topic_1': 0.08804115716929597, 'topic_2':
0.06180492281093961, 'topic_3': 0.07029331957415678, 'topic_4': 0.0896410799229668, 'topic_5':
0.09610510259822279, 'topic_6': 0.055111159357299495, 'topic_7': 0.04683699588268404, 'topic_8':
0.12105700062085314, 'topic_9': 0.0576806640519649, 'topic_10': 0.07478481570301246, 'topic_11':
0.08036795875002303, 'topic_12': 0.10942015699366556}
Chat:       214 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_6738959828-1563623700@g.us.chat.log ):
Topics:      {'topic_0': 0.057381425499828415, 'topic_1': 0.07564711273126208, 'topic_2':
0.06487809708806802, 'topic_3': 0.07896995893573772, 'topic_4': 0.07316957574243493, 'topic_5':
0.0802608895452888, 'topic_6': 0.06806263539537774, 'topic_7': 0.05936442383640362, 'topic_8':
0.1325986031771334, 'topic_9': 0.06993102989094187, 'topic_10': 0.06284656315357375, 'topic_11':
0.08596612993615657, 'topic_12': 0.09092355506785474}
Chat:       215 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_85268553604@c.us.chat.log ):
Topics:      {}
Chat:       216 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_8613136156865-1500422552@g.us.chat.log ):
Topics:      {'topic_0': 0.054322960248914975, 'topic_1': 0.12518715931866595, 'topic_2':
0.060229161784915414, 'topic_3': 0.16627207183812093, 'topic_4': 0.0360303385052782, 'topic_5':
0.040603514517874625, 'topic_6': 0.07617260807387417, 'topic_7': 0.16306277095988658, 'topic_8':
0.09243284543630871, 'topic_9': 0.04619296038755248, 'topic_10': 0.039079215835326725, 'topic_11':
0.049749915922074936, 'topic_12': 0.050664477171206174}
Chat:       217 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_8615099927791-1521182088@g.us.chat.log ):
Topics:      {'topic_0': 0.019230769230769624, 'topic_1': 0.019230769230769697, 'topic_2':
0.019230835755019347, 'topic_3': 0.019230778543997025, 'topic_4': 0.01923076923076957, 'topic_5':
0.7692304313239591, 'topic_6': 0.019230806712748367, 'topic_7': 0.019230820777808414, 'topic_8':
0.019230860753707082, 'topic_9': 0.019230850748143145, 'topic_10': 0.01923076923076961, 'topic_11':
0.01923076923076949, 'topic_12': 0.019230769230769457}
Chat:       218 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_8801744419079-1509171830@g.us.chat.log ):
Topics:      {'topic_0': 0.07407325588568194, 'topic_1': 0.07878619314815642, 'topic_2':
0.059930242486004544, 'topic_3': 0.05709988941503822, 'topic_4': 0.08464541618333908, 'topic_5':
0.09054618718342437, 'topic_6': 0.05194108109530855, 'topic_7': 0.04256951085724599, 'topic_8':
0.11481800070851007, 'topic_9': 0.12452550408399587, 'topic_10': 0.06945168139904374, 'topic_11':

```

```

0.08604101562772065, 'topic_12': 0.06557202192653713}
Chat: 219 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_8801812755493-1505678107@g.us.chat.log ):
Topics: {'topic_0': 0.10755448462141719, 'topic_1': 0.07451485418195508, 'topic_2':
0.09941335468052341, 'topic_3': 0.04000676168394397, 'topic_4': 0.05257159158689061, 'topic_5':
0.12283656670180661, 'topic_6': 0.05391919755644798, 'topic_7': 0.07896779523309297, 'topic_8':
0.11577286339218663, 'topic_9': 0.10470942786711791, 'topic_10': 0.04294360573377272, 'topic_11':
0.05746812969982169, 'topic_12': 0.04932136706102267}
Chat: 220 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_916207231114-1544876477@g.us.chat.log ):
Topics: {'topic_0': 0.04345727873208573, 'topic_1': 0.06301145286150302, 'topic_2':
0.07475513822420411, 'topic_3': 0.05961444823241363, 'topic_4': 0.07312113956414668, 'topic_5':
0.1415160659200424, 'topic_6': 0.09241844039578441, 'topic_7': 0.05892515129037771, 'topic_8':
0.13337427919273925, 'topic_9': 0.04727529370022266, 'topic_10': 0.07362526431677065, 'topic_11':
0.06324786190376894, 'topic_12': 0.07565818566594083}
Chat: 221 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_917011284590-1486113958@g.us.chat.log ):
Topics: {'topic_0': 0.036085728581094705, 'topic_1': 0.04858574298152947, 'topic_2':
0.09708431609643403, 'topic_3': 0.04962740690047081, 'topic_4': 0.08962740592580513, 'topic_5':
0.09004403793084842, 'topic_6': 0.0704607373679084, 'topic_7': 0.05629407947101953, 'topic_8':
0.09284757812443478, 'topic_9': 0.05796075654995061, 'topic_10': 0.17046073824631977, 'topic_11':
0.08483573562444438, 'topic_12': 0.05608573619973976}
Chat: 222 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_917016354308-1488093290@g.us.chat.log ):
Topics: {'topic_0': 0.14654730597590973, 'topic_1': 0.04745941359862448, 'topic_2':
0.047620997324180576, 'topic_3': 0.03711313179224849, 'topic_4': 0.03446880547805907, 'topic_5':
0.04838207019401173, 'topic_6': 0.05079423023946215, 'topic_7': 0.04012768450937373, 'topic_8':
0.06423191183989917, 'topic_9': 0.1388047024033656, 'topic_10': 0.033428820758128544, 'topic_11':
0.12371054539044848, 'topic_12': 0.1873103804962909}
Chat: 223 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_917058803589-1528353496@g.us.chat.log ):
Topics: {'topic_0': 0.01931560520261195, 'topic_1': 0.02329938289834723, 'topic_2':
0.01601994249261008, 'topic_3': 0.020492629865327905, 'topic_4': 0.026505177370155154, 'topic_5':
0.017432367459445968, 'topic_6': 0.030328790352801702, 'topic_7': 0.01774703348223303, 'topic_8':
0.019819149663896213, 'topic_9': 0.016019949186835244, 'topic_10': 0.02683045187145215, 'topic_11':
0.030276344678840796, 'topic_12': 0.7359131754754447}
Chat: 224 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_917066389716@c.us.chat.log ):
Topics: {'topic_0': 0.016452991452994007, 'topic_1': 0.183119660242012, 'topic_2':
0.04978632336231092, 'topic_3': 0.030605553176581878, 'topic_4': 0.04978632436285097, 'topic_5':
0.01645300221679387, 'topic_6': 0.01645299145299473, 'topic_7': 0.016453000887698292, 'topic_8':
0.3106337417829582, 'topic_9': 0.016452997348018458, 'topic_10': 0.030341879803205712, 'topic_11':
0.058119658242533845, 'topic_12': 0.20534187566904696}
Chat: 225 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_917096666815-1563845218@g.us.chat.log ):
Topics: {'topic_0': 0.16303417451542854, 'topic_1': 0.038034188034189044, 'topic_2':
0.03803418803418867, 'topic_3': 0.03803418803418891, 'topic_4': 0.03803418803418875, 'topic_5':
0.09358971972534691, 'topic_6': 0.03803418803418913, 'topic_7': 0.038034193062230985, 'topic_8':
0.07970086205402067, 'topic_9': 0.03803421743177831, 'topic_10': 0.03803419243870069, 'topic_11':

```

```

0.12136752135477301, 'topic_12': 0.23803417924677653}
Chat:      226 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_917340270078-1563980314@g.us.chat.log ):
Topics:    {'topic_0': 0.07692307692307694, 'topic_1': 0.07692307692307694, 'topic_2':
0.07692307692307694, 'topic_3': 0.07692307692307694, 'topic_4': 0.07692307692307694, 'topic_5':
0.07692307692307694, 'topic_6': 0.07692307692307694, 'topic_7': 0.07692307692307694, 'topic_8':
0.07692307692307694, 'topic_9': 0.07692307692307694, 'topic_10': 0.07692307692307694, 'topic_11':
0.07692307692307694, 'topic_12': 0.07692307692307694}
Chat:      227 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_917504961079-1519030093@g.us.chat.log ):
Topics:    {'topic_0': 0.12009608168319345, 'topic_1': 0.05861899093166613, 'topic_2':
0.0481764213397659, 'topic_3': 0.062162152946109946, 'topic_4': 0.0682405139479855, 'topic_5':
0.06791002226710213, 'topic_6': 0.06173632348800948, 'topic_7': 0.07714645190905028, 'topic_8':
0.08234217959845351, 'topic_9': 0.10419795782696077, 'topic_10': 0.04350277038151134, 'topic_11':
0.07195887130100728, 'topic_12': 0.1339112623791884}
Chat:      228 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_917504961079-1519136340@g.us.chat.log ):
Topics:    {'topic_0': 0.18520155518207412, 'topic_1': 0.02262486782610281, 'topic_2':
0.10544696501220428, 'topic_3': 0.02329823770772048, 'topic_4': 0.019557384549924922, 'topic_5':
0.024378854985657603, 'topic_6': 0.022624867813357184, 'topic_7': 0.028566387034428934, 'topic_8':
0.03489480388015865, 'topic_9': 0.38169877647227174, 'topic_10': 0.020121171851274103, 'topic_11':
0.06264484263688674, 'topic_12': 0.06894128504793791}
Chat:      229 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_917504961079-1519185392@g.us.chat.log ):
Topics:    {'topic_0': 0.0560803980583795, 'topic_1': 0.06548866197952495, 'topic_2':
0.08930067771778426, 'topic_3': 0.08980125690300114, 'topic_4': 0.05889225105944477, 'topic_5':
0.08057746576619836, 'topic_6': 0.08354782183820984, 'topic_7': 0.07890731258496927, 'topic_8':
0.08006419239009482, 'topic_9': 0.058293517131800046, 'topic_10': 0.052425695156014016, 'topic_11':
0.10856692675548454, 'topic_12': 0.09805382265909379}
Chat:      230 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_917504961079-1519185435@g.us.chat.log ):
Topics:    {'topic_0': 0.05857398669660506, 'topic_1': 0.046717212300443726, 'topic_2':
0.10499503082078955, 'topic_3': 0.07770885560371087, 'topic_4': 0.07499099717572826, 'topic_5':
0.09872122170629967, 'topic_6': 0.06936125748712424, 'topic_7': 0.05525736391432623, 'topic_8':
0.08620546227808527, 'topic_9': 0.05907013471661865, 'topic_10': 0.06757870668585177, 'topic_11':
0.08907150721627358, 'topic_12': 0.11174826339814398}
Chat:      231 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_917896159081-1506248485@g.us.chat.log ):
Topics:    {'topic_0': 0.03329722328407452, 'topic_1': 0.12343805323557305, 'topic_2':
0.033297224144975754, 'topic_3': 0.05231129339035217, 'topic_4': 0.07555073746092096, 'topic_5':
0.08186200188448675, 'topic_6': 0.07320336050927824, 'topic_7': 0.11357686040259152, 'topic_8':
0.11589658541436794, 'topic_9': 0.047659738299880086, 'topic_10': 0.06968218411877183, 'topic_11':
0.0633594476595457, 'topic_12': 0.1168652901951813}
Chat:      232 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_917900186560-1550952626@g.us.chat.log ):
Topics:    {'topic_0': 0.04389072727091081, 'topic_1': 0.0692118299722892, 'topic_2':
0.060508819749011726, 'topic_3': 0.09062018800889554, 'topic_4': 0.08237129824796606, 'topic_5':
0.08392590361566603, 'topic_6': 0.06714103560281341, 'topic_7': 0.06952833552190893, 'topic_8':
0.0871666689156991, 'topic_9': 0.046936882418709414, 'topic_10': 0.0777645168851838, 'topic_11':

```



```

0.07471831730577597, 'topic_12': 0.14621547648517383}
Chat:      233 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_917985455537-1488972204@g.us.chat.log ):
Topics:    {'topic_0': 0.044191105457922424, 'topic_1': 0.061727488295998466, 'topic_2':
0.06954414517277827, 'topic_3': 0.047514450421164055, 'topic_4': 0.061679569387086394, 'topic_5':
0.09535550157071432, 'topic_6': 0.0771158956108014, 'topic_7': 0.06410836805106478, 'topic_8':
0.08529263799886835, 'topic_9': 0.05502126225468267, 'topic_10': 0.054959845648637366, 'topic_11':
0.056452917504201734, 'topic_12': 0.22703681262607983}
Chat:      234 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918011717118-1513346963@g.us.chat.log ):
Topics:    {'topic_0': 0.07874276920079391, 'topic_1': 0.06723596212992784, 'topic_2':
0.05238048858949209, 'topic_3': 0.05663583302545564, 'topic_4': 0.052794348320113725, 'topic_5':
0.07956669670873764, 'topic_6': 0.059707569414135686, 'topic_7': 0.06563733427827824, 'topic_8':
0.14519866182596894, 'topic_9': 0.08826379698249369, 'topic_10': 0.056956224664827096, 'topic_11':
0.09676509308688339, 'topic_12': 0.10011522177289471}
Chat:      235 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918074820605@c.us.chat.log ):
Topics:    {}
Chat:      236 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918112269746-1515737968@g.us.chat.log ):
Topics:    {}
Chat:      237 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918154995870-1558890931@g.us.chat.log ):
Topics:    {'topic_0': 0.06626180836707188, 'topic_1': 0.07941970310391316, 'topic_2':
0.06626181328390693, 'topic_3': 0.06626180836707189, 'topic_4': 0.06626180836707182, 'topic_5':
0.11012138749199837, 'topic_6': 0.06626181798224136, 'topic_7': 0.06626183121052331, 'topic_8':
0.08380568848788714, 'topic_9': 0.07941973236531667, 'topic_10': 0.06626181114886881, 'topic_11':
0.0750337202386911, 'topic_12': 0.10836706958543804}
Chat:      238 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918178289603-1561741836@g.us.chat.log ):
Topics:    {'topic_0': 0.12733062819447533, 'topic_1': 0.04616325208598339, 'topic_2':
0.03793498423603968, 'topic_3': 0.04222898055617815, 'topic_4': 0.04680644855749857, 'topic_5':
0.06737118872076567, 'topic_6': 0.040066479011964445, 'topic_7': 0.04356906316683479, 'topic_8':
0.06125626702362053, 'topic_9': 0.16833101753538413, 'topic_10': 0.04778867600506932, 'topic_11':
0.1759527468209471, 'topic_12': 0.09520026808524089}
Chat:      239 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918197180324-1564038241@g.us.chat.log ):
Topics:    {'topic_0': 0.019784461723552377, 'topic_1': 0.022171829358233537, 'topic_2':
0.021313432446276138, 'topic_3': 0.018736451741936958, 'topic_4': 0.03203064152243398, 'topic_5':
0.061190693098304536, 'topic_6': 0.044202245146745464, 'topic_7': 0.037546606968697896, 'topic_8':
0.02274779512959812, 'topic_9': 0.02934251840104135, 'topic_10': 0.02782732253886658, 'topic_11':
0.027728273187415785, 'topic_12': 0.6353777287368971}
Chat:      240 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918197180324-1564039245@g.us.chat.log ):
Topics:    {'topic_0': 0.029733934870169853, 'topic_1': 0.01892337716094954, 'topic_2':
0.021038825334823007, 'topic_3': 0.018923396404676926, 'topic_4': 0.044761909270096996, 'topic_5':
0.031385843785330235, 'topic_6': 0.06967889620103876, 'topic_7': 0.035394188741453864, 'topic_8':
0.0200612691247155, 'topic_9': 0.027629869798432394, 'topic_10': 0.018923381625532726, 'topic_11':
0.03376741920776667, 'topic_12': 0.6297776884750139}

```

```

Chat:      241 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918507097725-1557250497@g.us.chat.log ):
Topics:    {'topic_0': 0.052138809952132194, 'topic_1': 0.05748069109202055, 'topic_2':
0.07003411351771144, 'topic_3': 0.08045076823745254, 'topic_4': 0.05854906465823214, 'topic_5':
0.08072994528266207, 'topic_6': 0.06335676277773815, 'topic_7': 0.053606757939288215, 'topic_8':
0.07263157872096979, 'topic_9': 0.06221171001574433, 'topic_10': 0.06829799531282522, 'topic_11':
0.05989009813588039, 'topic_12': 0.22062170435734293}
Chat:      242 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918795192508-1483164938@g.us.chat.log ):
Topics:    {'topic_0': 0.060044560716573485, 'topic_1': 0.0659676147945699, 'topic_2':
0.09008363424913997, 'topic_3': 0.04898511211212835, 'topic_4': 0.05165836583518389, 'topic_5':
0.05284989380140835, 'topic_6': 0.06144809673784676, 'topic_7': 0.06888589318899975, 'topic_8':
0.09897438192342024, 'topic_9': 0.05992105227605312, 'topic_10': 0.0531152989399928, 'topic_11':
0.08066338847608849, 'topic_12': 0.2074027069485982}
Chat:      243 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918795744594-1541928672@g.us.chat.log ):
Topics:    {'topic_0': 0.08942714529471198, 'topic_1': 0.047539739283694676, 'topic_2':
0.05331896972027045, 'topic_3': 0.04647185538150262, 'topic_4': 0.04722373553197657, 'topic_5':
0.0582298252919738, 'topic_6': 0.06873857416092, 'topic_7': 0.07558313089766276, 'topic_8':
0.06110968900599558, 'topic_9': 0.10119943351233254, 'topic_10': 0.0532630817688601, 'topic_11':
0.10285922188553158, 'topic_12': 0.19503559826456918}
Chat:      244 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918826310112-1553089301@g.us.chat.log ):
Topics:    {'topic_0': 0.053685485742359866, 'topic_1': 0.08425873363819779, 'topic_2':
0.06773149830616877, 'topic_3': 0.07343068926021473, 'topic_4': 0.08430926785298828, 'topic_5':
0.07800673600911319, 'topic_6': 0.08679257334469584, 'topic_7': 0.07268856270872165, 'topic_8':
0.09835091939387433, 'topic_9': 0.06070849850873194, 'topic_10': 0.06957914174421866, 'topic_11':
0.06717601274482969, 'topic_12': 0.10328188074588683}
Chat:      245 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918850732033-1560277858@g.us.chat.log ):
Topics:    {'topic_0': 0.056710585051898084, 'topic_1': 0.07568075614760882, 'topic_2':
0.04967660685578647, 'topic_3': 0.06324798517689655, 'topic_4': 0.08055954133097654, 'topic_5':
0.08102638627847264, 'topic_6': 0.09412437882926081, 'topic_7': 0.06796413235220554, 'topic_8':
0.12620450037169734, 'topic_9': 0.06703720827645293, 'topic_10': 0.07618627171072226, 'topic_11':
0.07364039928878968, 'topic_12': 0.08794124832923635}
Chat:      246 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918850823474-1551458263@g.us.chat.log ):
Topics:    {'topic_0': 0.11532958587194257, 'topic_1': 0.06283613791521403, 'topic_2':
0.04511960172059006, 'topic_3': 0.05102513070700103, 'topic_4': 0.05072686925255872, 'topic_5':
0.048492740161314496, 'topic_6': 0.0588489530563622, 'topic_7': 0.07469427405410115, 'topic_8':
0.07031690656792583, 'topic_9': 0.14141227276137552, 'topic_10': 0.05223441778613994, 'topic_11':
0.1477286942429743, 'topic_12': 0.08123441590249995}
Chat:      247 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918866126299-1502878641@g.us.chat.log ):
Topics:    {'topic_0': 0.08915886159640703, 'topic_1': 0.07097857235801902, 'topic_2':
0.048479139824500014, 'topic_3': 0.04972516884505606, 'topic_4': 0.04521560031106942, 'topic_5':
0.10888280056898682, 'topic_6': 0.04703801830843177, 'topic_7': 0.08312885412912344, 'topic_8':
0.1446568731806423, 'topic_9': 0.08051225283531359, 'topic_10': 0.05137996032758357, 'topic_11':
0.06139827796640056, 'topic_12': 0.11944561974847161}

```

```

Chat:      248 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918874547144-1539766758@g.us.chat.log ):
Topics:    {'topic_0': 0.0395801630647862, 'topic_1': 0.05029758013630012, 'topic_2':
0.046950363383721015, 'topic_3': 0.04728529064453342, 'topic_4': 0.05213782428338217, 'topic_5':
0.09224424795235447, 'topic_6': 0.06553480752160319, 'topic_7': 0.17220278280690085, 'topic_8':
0.05763404022094724, 'topic_9': 0.03818358732810304, 'topic_10': 0.08100632268259401, 'topic_11':
0.04401811011849995, 'topic_12': 0.2129248798562748}
Chat:      249 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918903675398-1525952582@g.us.chat.log ):
Topics:    {'topic_0': 0.04066149463496427, 'topic_1': 0.05730305380221398, 'topic_2':
0.050437746383418997, 'topic_3': 0.07847890508768662, 'topic_4': 0.07321712728047358, 'topic_5':
0.08430172892308054, 'topic_6': 0.06937727144410089, 'topic_7': 0.06767298302378692, 'topic_8':
0.08139039935211403, 'topic_9': 0.043893223880769223, 'topic_10': 0.05710435574137674, 'topic_11':
0.05914167209037979, 'topic_12': 0.23702003835563554}
Chat:      250 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_918956783671-1552825100@g.us.chat.log ):
Topics:    {'topic_0': 0.11740829893726085, 'topic_1': 0.04519287823707516, 'topic_2':
0.048596533440932295, 'topic_3': 0.04615198649085423, 'topic_4': 0.05171763456272455, 'topic_5':
0.05677357075373087, 'topic_6': 0.0652056704765933, 'topic_7': 0.087936118861292, 'topic_8':
0.09136305892574162, 'topic_9': 0.13476291718614114, 'topic_10': 0.04688097617919136, 'topic_11':
0.12827367421182886, 'topic_12': 0.07973668173664065}
Chat:      251 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_919102711964-1562017267@g.us.chat.log ):
Topics:    {'topic_0': 0.04922803093467578, 'topic_1': 0.08674235902851689, 'topic_2':
0.06160754316415135, 'topic_3': 0.07704612980883382, 'topic_4': 0.1361463940363263, 'topic_5':
0.100374249685251, 'topic_6': 0.05566534904424003, 'topic_7': 0.08628000898285207, 'topic_8':
0.05969481391087862, 'topic_9': 0.05561884510018507, 'topic_10': 0.07799349228359494, 'topic_11':
0.057571973065742896, 'topic_12': 0.09603081095475047}
Chat:      252 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_919137566247-1548228361@g.us.chat.log ):
Topics:    {'topic_0': 0.01953601953755662, 'topic_1': 0.01694342694342805, 'topic_2':
0.019165652673045933, 'topic_3': 0.0215730639453029, 'topic_4': 0.01879527901410378, 'topic_5':
0.021017497581562277, 'topic_6': 0.022684162305442037, 'topic_7': 0.01861009483739155, 'topic_8':
0.025832311911119895, 'topic_9': 0.017498995157589185, 'topic_10': 0.022869355641988602, 'topic_11':
0.021387869852261574, 'topic_12': 0.7540862705992044}
Chat:      253 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_919321629786-1491165067@g.us.chat.log ):
Topics:    {'topic_0': 0.03540128261110507, 'topic_1': 0.04772650739894055, 'topic_2':
0.06539471250829637, 'topic_3': 0.04969070283225255, 'topic_4': 0.0666990637498312, 'topic_5':
0.07913388608370885, 'topic_6': 0.0475744399932681, 'topic_7': 0.05112898931307058, 'topic_8':
0.056242151436950036, 'topic_9': 0.03931422873066043, 'topic_10': 0.047196628678453645, 'topic_11':
0.04916634195651721, 'topic_12': 0.3653310647069444}
Chat:      254 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_919504171173-1521911266@g.us.chat.log ):
Topics:    {'topic_0': 0.04079887554416003, 'topic_1': 0.05564747462461545, 'topic_2':
0.08560593032780667, 'topic_3': 0.0647587432241596, 'topic_4': 0.06789947207431335, 'topic_5':
0.1704503038520041, 'topic_6': 0.06661849786870915, 'topic_7': 0.051509892381326736, 'topic_8':
0.13011744562141844, 'topic_9': 0.043957758900286026, 'topic_10': 0.08629979822721887, 'topic_11':
0.05418203642086317, 'topic_12': 0.08215377093312273}

```

```

Chat:      255 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_919567635886-1544363903@g.us.chat.log ):
Topics:    {'topic_0': 0.10031261052459478, 'topic_1': 0.07891340677557108, 'topic_2':
0.09072466407173353, 'topic_3': 0.04109095163099943, 'topic_4': 0.07061717963118042, 'topic_5':
0.05589636768441083, 'topic_6': 0.06004187407017827, 'topic_7': 0.03686083876676645, 'topic_8':
0.050758296534139484, 'topic_9': 0.10760957505303698, 'topic_10': 0.03897589420583232, 'topic_11':
0.09088549357620897, 'topic_12': 0.17731284747534767}
Chat:      256 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_919623418138-1490536344@g.us.chat.log ):
Topics:    {'topic_0': 0.040836128534243454, 'topic_1': 0.06640437165482424, 'topic_2':
0.06402212292571126, 'topic_3': 0.07103228590616034, 'topic_4': 0.06123796902303944, 'topic_5':
0.11254057222085481, 'topic_6': 0.0625749359293793, 'topic_7': 0.0702893703999883, 'topic_8':
0.12185524433094312, 'topic_9': 0.040830925886998985, 'topic_10': 0.060535508408654765, 'topic_11':
0.0669007636300811, 'topic_12': 0.160939801149122}
Chat:      257 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_919672741666-1560305831@g.us.chat.log ):
Topics:    {'topic_0': 0.01603756263870802, 'topic_1': 0.01797444382328864, 'topic_2':
0.015446411734250659, 'topic_3': 0.017606929646676307, 'topic_4': 0.024293706100704873, 'topic_5':
0.016831105678940538, 'topic_6': 0.06099364290086476, 'topic_7': 0.016884245020208737, 'topic_8':
0.017461022463636598, 'topic_9': 0.01607582966711375, 'topic_10': 0.01754862636844941, 'topic_11':
0.015918822861170773, 'topic_12': 0.7469276510960041}
Chat:      258 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_919737834397-1560511400@g.us.chat.log ):
Topics:    {'topic_0': 0.04551857701954454, 'topic_1': 0.04814326255584926, 'topic_2':
0.044206287358468835, 'topic_3': 0.04814329034216908, 'topic_4': 0.054048774774393626, 'topic_5':
0.10391749789210779, 'topic_6': 0.05741695247118686, 'topic_7': 0.07570976271801058, 'topic_8':
0.08325840422896215, 'topic_9': 0.04223777570309729, 'topic_10': 0.0813631969406546, 'topic_11':
0.05202324567873651, 'topic_12': 0.2640129723168182}
Chat:      259 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_919828556027-1547698927@g.us.chat.log ):
Topics:    {'topic_0': 0.03639903091955166, 'topic_1': 0.05862125460145156, 'topic_2':
0.038621256164117664, 'topic_3': 0.10702541040821671, 'topic_4': 0.04941489637673862, 'topic_5':
0.0863990016330886, 'topic_6': 0.05417679899049907, 'topic_7': 0.04466793299647311, 'topic_8':
0.08777263065516218, 'topic_9': 0.0363990402267041, 'topic_10': 0.048303775881848775, 'topic_11':
0.058621239589728175, 'topic_12': 0.2935777315564199}
Chat:      260 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_919860598106-1522141649@g.us.chat.log ):
Topics:    {'topic_0': 0.08646586816248225, 'topic_1': 0.05578652106400105, 'topic_2':
0.04980926937903799, 'topic_3': 0.054513482953868084, 'topic_4': 0.05791265425369448, 'topic_5':
0.0687381838452539, 'topic_6': 0.08732341905180095, 'topic_7': 0.05695056843208764, 'topic_8':
0.10925624991575146, 'topic_9': 0.09436198007273347, 'topic_10': 0.059950605505935395, 'topic_11':
0.10810464826167464, 'topic_12': 0.11082654910167865}
Chat:      261 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_923056474175@c.us.chat.log ):
Topics:    {}
Chat:      262 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_923056733587-1508906176@g.us.chat.log ):
Topics:    {'topic_0': 0.07168730043711467, 'topic_1': 0.08356341393133504, 'topic_2':
0.03838834945895704, 'topic_3': 0.0421897718220467, 'topic_4': 0.09052983242364276, 'topic_5':

```

```

0.06529532343637737, 'topic_6': 0.09210085826491568, 'topic_7': 0.06250356726481522, 'topic_8':
0.056731460918970654, 'topic_9': 0.11532573388313642, 'topic_10': 0.08076367777180483, 'topic_11':
0.04296003106761566, 'topic_12': 0.1579606793192741}
Chat: 263 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_923106438145@c.us.chat.log ):
Topics: {}
Chat: 264 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_923129261430-1532194872@g.us.chat.log ):
Topics: {'topic_0': 0.05327310267605306, 'topic_1': 0.12871557142389844, 'topic_2':
0.05082213107688019, 'topic_3': 0.05449861237451109, 'topic_4': 0.07123956418449906, 'topic_5':
0.06152198312350628, 'topic_6': 0.062464279856747774, 'topic_7': 0.12112606564689571, 'topic_8':
0.08522079536672744, 'topic_9': 0.053616823097215355, 'topic_10': 0.07180952671296853, 'topic_11':
0.09723814668454749, 'topic_12': 0.08845339777554857}
Chat: 265 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_923147117253-1520597965@g.us.chat.log ):
Topics: {'topic_0': 0.026734976476886677, 'topic_1': 0.04337653794664126, 'topic_2':
0.03253124006350164, 'topic_3': 0.034360464224222384, 'topic_4': 0.05233738387769343, 'topic_5':
0.05144701448025628, 'topic_6': 0.04183558378018351, 'topic_7': 0.030412490598380207, 'topic_8':
0.04568961571302136, 'topic_9': 0.02809759059258853, 'topic_10': 0.029672466046653984, 'topic_11':
0.040519444046229584, 'topic_12': 0.5429851921537416}
Chat: 266 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_923147117253-1528994513@g.us.chat.log ):
Topics: {'topic_0': 0.024137197715399232, 'topic_1': 0.04131732120145096, 'topic_2':
0.03717037470345404, 'topic_3': 0.04693201305657511, 'topic_4': 0.39527577693462596, 'topic_5':
0.14694535969145922, 'topic_6': 0.04893971815862447, 'topic_7': 0.02413720297378734, 'topic_8':
0.026345678496789825, 'topic_9': 0.03768384298280068, 'topic_10': 0.02647460933671971, 'topic_11':
0.024137200915447885, 'topic_12': 0.1205037038328661}
Chat: 267 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_923169964268-1494772094@g.us.chat.log ):
Topics: {'topic_0': 0.0669668646623524, 'topic_1': 0.06540922554652327, 'topic_2': 0.0671226276815263,
'topic_3': 0.05606343243883839, 'topic_4': 0.07677993314697325, 'topic_5': 0.09569342616591547,
'topic_6': 0.10240544011215635, 'topic_7': 0.06523007763780352, 'topic_8': 0.09897694492212278,
'topic_9': 0.06696692009394291, 'topic_10': 0.07125034722136413, 'topic_11': 0.058235895485439,
'topic_12': 0.10889886488504105}
Chat: 268 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_923214324132-1552761743@g.us.chat.log ):
Topics: {'topic_0': 0.022971374341431622, 'topic_1': 0.023355798571347515, 'topic_2':
0.018135064228343205, 'topic_3': 0.020169280451273277, 'topic_4': 0.016646968431771735, 'topic_5':
0.02126107447802626, 'topic_6': 0.019039193925437098, 'topic_7': 0.017515961491666572, 'topic_8':
0.01796972073445806, 'topic_9': 0.03314553623056882, 'topic_10': 0.018135064960944698, 'topic_11':
0.017830680081768458, 'topic_12': 0.7538242820729633}
Chat: 269 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_923455024491-1564300348@g.us.chat.log ):
Topics: {'topic_0': 0.04980896801855082, 'topic_1': 0.06447524891975595, 'topic_2':
0.06942920838755386, 'topic_3': 0.07065499406901332, 'topic_4': 0.06591352068971504, 'topic_5':
0.08941206929875004, 'topic_6': 0.0753785861662651, 'topic_7': 0.05209044368587657, 'topic_8':
0.12514583780503083, 'topic_9': 0.05791344554218194, 'topic_10': 0.07512542623025688, 'topic_11':
0.05667906836058258, 'topic_12': 0.14797318282646738}
Chat: 270 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\

```

```

safechat_94750947291@c.us.chat.log ):
Topics: {}
Chat: 271 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_971528904770-1563885486@g.us.chat.log ):
Topics: {'topic_0': 0.04287738851965587, 'topic_1': 0.05400430543636397, 'topic_2':
0.057236853516961415, 'topic_3': 0.05570384569733732, 'topic_4': 0.09345219108388512, 'topic_5':
0.09245767068408328, 'topic_6': 0.0633254302791292, 'topic_7': 0.054922743181652696, 'topic_8':
0.14241177693494217, 'topic_9': 0.05569417496726856, 'topic_10': 0.05416495446543665, 'topic_11':
0.06559606484296546, 'topic_12': 0.16815260039031787}
Chat: 272 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_97338381710-1487691608@g.us.chat.log ):
Topics: {'topic_0': 0.044305699936810765, 'topic_1': 0.17119961811327167, 'topic_2':
0.05188145463017948, 'topic_3': 0.05891606712755748, 'topic_4': 0.04430569571340702, 'topic_5':
0.051881448312670914, 'topic_6': 0.059457208635236646, 'topic_7': 0.0670329852241557, 'topic_8':
0.18770397385644347, 'topic_9': 0.07650264818755066, 'topic_10': 0.051881462587499684, 'topic_11':
0.05945721347240052, 'topic_12': 0.07547452420281599}
Chat: 273 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_97338381710-1497431182@g.us.chat.log ):
Topics: {'topic_0': 0.05324141488893543, 'topic_1': 0.06948880660986108, 'topic_2':
0.057434284945126396, 'topic_3': 0.07547743163615109, 'topic_4': 0.057434289235309854, 'topic_5':
0.08540677420515796, 'topic_6': 0.05953072020463971, 'topic_7': 0.06893934821516742, 'topic_8':
0.16179658645562947, 'topic_9': 0.0648749255010761, 'topic_10': 0.06277742284932743, 'topic_11':
0.09189479538432023, 'topic_12': 0.0917031998692969}
Chat: 274 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_97338381710-1500435936@g.us.chat.log ):
Topics: {'topic_0': 0.037905982905986596, 'topic_1': 0.21568377545651732, 'topic_2':
0.04901707592243355, 'topic_3': 0.08075684314942039, 'topic_4': 0.03790599417300206, 'topic_5':
0.06443252855067698, 'topic_6': 0.03790598973353378, 'topic_7': 0.07358468766813518, 'topic_8':
0.1490171256477331, 'topic_9': 0.06012821259559478, 'topic_10': 0.052335707405863706, 'topic_11':
0.06568372130423454, 'topic_12': 0.07564235548686812}
Chat: 275 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_97338381710-1502880819@g.us.chat.log ):
Topics: {'topic_0': 0.05060624130391869, 'topic_1': 0.10706616722895672, 'topic_2':
0.05060625378642102, 'topic_3': 0.07799643215956914, 'topic_4': 0.09324190020868661, 'topic_5':
0.06869407102682004, 'topic_6': 0.05448220520298404, 'topic_7': 0.06223417242646564, 'topic_8':
0.11675614674646838, 'topic_9': 0.056420241704924205, 'topic_10': 0.0829059817140383, 'topic_11':
0.08032193378944134, 'topic_12': 0.09866825270130537}
Chat: 276 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_97338381710-1504447471@g.us.chat.log ):
Topics: {'topic_0': 0.04395604395604695, 'topic_1': 0.2344322069842176, 'topic_2':
0.043956096530134874, 'topic_3': 0.11538447435714191, 'topic_4': 0.04395604395604652, 'topic_5':
0.0598290351637652, 'topic_6': 0.043956053709685056, 'topic_7': 0.09157516175215029, 'topic_8':
0.09951166020297036, 'topic_9': 0.05982905984412993, 'topic_10': 0.05982905977367554, 'topic_11':
0.05982905981399017, 'topic_12': 0.043956043956045666}
Chat: 277 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_97338381710-1507218698@g.us.chat.log ):
Topics: {'topic_0': 0.05085470085470425, 'topic_1': 0.05085476830340088, 'topic_2':
0.08418795680951165, 'topic_3': 0.050854713257879096, 'topic_4': 0.050854700854703776, 'topic_5':
0.07307704038485054, 'topic_6': 0.06752135396116138, 'topic_7': 0.08418802723972237, 'topic_8':

```

```

0.22863239828025209, 'topic_9': 0.07307690540357403, 'topic_10': 0.05085470085470416, 'topic_11':
0.08418803294083309, 'topic_12': 0.05085470085470279}
Chat:      278 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_97338381710-1511318890@g.us.chat.log ):
Topics:    {'topic_0': 0.050807201032864034, 'topic_1': 0.06932575639878871, 'topic_2':
0.0637701720560598, 'topic_3': 0.09299048052483773, 'topic_4': 0.06377019651916427, 'topic_5':
0.08071216006687759, 'topic_6': 0.05173313703522399, 'topic_7': 0.08512762336157832, 'topic_8':
0.14623631541599147, 'topic_9': 0.04895539979396388, 'topic_10': 0.07596290193600758, 'topic_11':
0.07790893982111764, 'topic_12': 0.09269971603752478}
Chat:      279 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_97338381710-1520734066@g.us.chat.log ):
Topics:    {'topic_0': 0.06304384352946693, 'topic_1': 0.09422663789647394, 'topic_2':
0.059818032964628766, 'topic_3': 0.07234108874141229, 'topic_4': 0.07863523180453942, 'topic_5':
0.07971049762895734, 'topic_6': 0.06842018391704047, 'topic_7': 0.06304384037113993, 'topic_8':
0.12529739282495567, 'topic_9': 0.0764392564309748, 'topic_10': 0.07379652903615862, 'topic_11':
0.07702233690333594, 'topic_12': 0.06820512795091446}
Chat:      280 / 280 ( C:/Users/Geo/Desktop/george/WebWhatsapp-Wrapper-master/sample\
safechat_994557753110-1563735846@g.us.chat.log ):
Topics:    {'topic_0': 0.04582604434166089, 'topic_1': 0.06105435559016121, 'topic_2':
0.08071559725161476, 'topic_3': 0.05172457950013373, 'topic_4': 0.06780784744791842, 'topic_5':
0.1062154132020377, 'topic_6': 0.10200914026214025, 'topic_7': 0.07327697430230999, 'topic_8':
0.06787030846588751, 'topic_9': 0.04480382944050457, 'topic_10': 0.110794412301285, 'topic_11':
0.05185414611928949, 'topic_12': 0.13604735177505897}

```

Πίνακας 4. Τελικά αποτελέσματα WhatsApp με LDA.

Κεφάλαιο 7. K-means clustering (ομαδοποίηση)

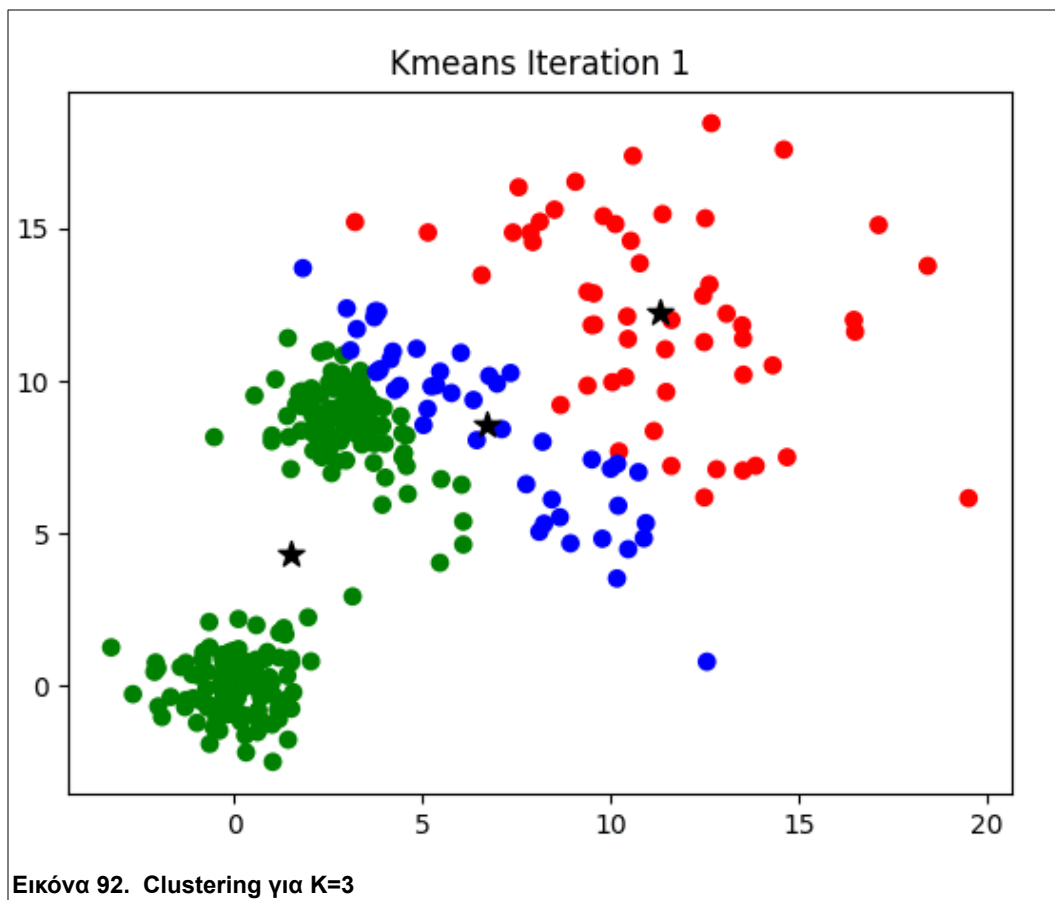
Ο K-means είναι ένας από τους πιο απλούς και πιο δημοφιλείς μη επιτηρούμενους αλγόριθμους μηχανικής μάθησης. Συνήθως, οι μη επιτηρούμενοι αλγόριθμοι βγάζουν συμπεράσματα από σύνολα δεδομένων χρησιμοποιώντας μόνο φορείς εισόδου χωρίς αναφορά σε γνωστά ή επισημασμένα αποτελέσματα.

Ο στόχος του K-means είναι απλός: η ομαδοποίηση παρόμοιων δεδομένων και η ανακάλυψη υποκείμενων μοτίβων. Για να επιτευχθεί αυτός ο στόχος, ο K-means αναζητά για έναν καθορισμένο αριθμό (k) ομάδων (clusters) σε ένα σύνολο δεδομένων.

Ως cluster ορίζεται μια συλλογή σημείων δεδομένων συγκεντρωμένα μεταξύ τους λόγω ορισμένων ομοιοτήτων.

Ο K-means είναι υποχρεωμένος να συγκλίνει σε μια λύση μετά από μερικές επαναλήψεις, και αυτό το πετυχαίνει με τα εξής 4 βασικά βήματα:

1. Αρχικοποίηση των Centroids των Cluster
2. Εκχώρηση των datapoints στα Cluster
3. Ενημέρωση των Centroids των Cluster
4. Επανάληψη των βημάτων 2-3 μέχρι να επιτευχθεί η κατάσταση διακοπής.



1. Αρχικοποίηση των Centroids των Cluster

Αρχικά, ορίζουμε στο μοντέλο των αριθμό K των cluster που θα πρέπει να δημιουργήσει. Αν υποθέσουμε ότι ο αριθμός $K = 3$ τότε το μοντέλο θα πάρει 3 datapoints από το σύνολο δεδομένων. Αυτά τα datapoints ονομάζονται centroids συμπλέγματος.

Υπάρχουν διάφοροι τρόποι για να αρχικοποιήσουμε τα κεντροειδή. Μπορούμε να τα επιλέξουμε είτε τυχαία, είτε να ταξινομήσουμε το σύνολο δεδομένων, να τα χωρίσουμε σε K τμήματα και να επιλέξουμε ένα datapoint από κάθε τμήμα ως centroid.

2. Εκχώρηση των datapoints στα Cluster

Σε αυτό το βήμα το μοντέλο εκτελεί υπολογισμούς από μόνο του και εκχωρεί το κάθε datapoint σε ένα cluster. Το μοντέλο θα υπολογίσει την απόσταση μεταξύ του datapoint & όλων των centroids και θα το αντιστοιχίσει στο σύμπλεγμα με το πλησιέστερο centroids. Και σε αυτήν την περίπτωση υπάρχουν διάφοροι τρόποι με τους οποίους υπολογίζετε αυτή η απόσταση και όλοι έχουν τα υπέρ τους και τα κατά τους.

3. Ενημέρωση των Centroids των Cluster

Επειδή τα αρχικά κεντροειδή επιλέχθηκαν αυθαίρετα, το μοντέλο τα ενημερώνει με νέες τιμές συμπλέγματος. Η νέα τιμή μπορεί να βρίσκεται ή μπορεί και όχι στο σύνολο δεδομένων. Στην πραγματικότητα, αν αυτό συνέβαινε θα ήταν μια σύμπτωση. Και αυτό συμβαίνει επειδή το ενημερωμένο centroid του cluster είναι ο μέσος όρος των μέσω των τιμών όλων των datapoints μέσα σε αυτό το σύμπλεγμα.

4. Κριτήριο Διακοπής

Δεδομένου ότι τα βήματα 2 και 3 εκτελούνται επαναληπτικά, η διαδικασία θα συνεχίζεται επ' άπειρον αν δεν θέσουμε κάποιο κριτήριο διακοπής. Το κριτήριο διακοπής λέει στον αλγόριθμο πότε πρέπει να σταματήσει να ενημερώνει τα clusters. Θα πρέπει να σημειωθεί ότι ο καθορισμός ενός κριτηρίου διακοπής δεν θα επέστρεφε απαραίτητα τα καλύτερα clusters, αλλά για να βεβαιωθούμε ότι επιστρέφει εύλογα καλά clusters, ή το σημαντικότερο, να είναι σε θέση τουλάχιστον να επιστρέψει μερικά clusters, θα πρέπει να έχουμε ένα κριτήριο διακοπής.

Υπάρχουν διάφοροι τρόποι για να ορίσουμε κάποιο κριτήριο διακοπής. Επιπλέον θα μπορούσαμε να ορίσουμε ακόμη και πολλαπλές συνθήκες όπου εάν πληρούνται, θα σταματήσουν την επανάληψη του αλγορίθμου επιστρέφοντας μας τα αποτελέσματα. Ορισμένες από τις συνθήκες διακοπής είναι οι ακόλουθοι:

1. Τα datapoints που αντιστοιχούν σε κάποιο συγκεκριμένο cluster παραμένουν τα ίδια. (χρονοβόρο)
2. Τα κεντροειδή (centroids) παραμένουν τα ίδια. (χρονοβόρο)
3. Η απόσταση των datapoints από το centroid τους είναι ελάχιστη. (Ένα threshold που θέτουμε)
4. Έχει επιτευχθεί ένας δεδομένος αριθμός επαναλήψεων (ανεπαρκείς επαναλήψεις μας δίνουν ανεπαρκή αποτελέσματα, οπότε επιλέγουμε τον αριθμό επανάληψης με σύνεση)

Κεφάλαιο 8. TF-IDF

Ο όρος tf-idf (term frequency-inverse document frequency) σημαίνει συχνότητα όρων - αντίστροφη συχνότητα εγγράφων, και το βάρος tf-idf είναι ένα βάρος το οποίο χρησιμοποιείται συχνά στην ανάκτηση πληροφοριών και την εξόρυξη κειμένου (text mining). Αυτό το βάρος είναι ένα στατιστικό μέτρο που χρησιμοποιείται για να αξιολογήσει το πόσο σημαντική είναι μια λέξη σε ένα έγγραφο (document) σε μια συλλογή (collection) ή ένα corpus. Η σημασία αυξάνεται ανάλογα με τον αριθμό των φορών που εμφανίζεται μια λέξη στο document, αλλά αντισταθμίζεται από τη συχνότητα της λέξης στο corpus. Παραλλαγές της tf-idf στάθμισης χρησιμοποιούνται συχνά ως βασικό εργαλείο από τις μηχανές αναζήτησης για τη βαθμολόγηση και την ταξινόμηση της συνάφειας ενός document σε σχέση με μια αναζήτηση ενός χρήστη.

Μια από τις απλούστερες λειτουργίες κατάταξης (ranking) υπολογίζεται με άθροιση του tf-idf για κάθε όρο αναζήτησης (query), ενώ μερικές πιο εξελιγμένες λειτουργίες κατάταξης είναι παραλλαγές αυτού του απλού μοντέλου.

Το tf-idf μπορεί να χρησιμοποιηθεί με επιτυχία για το φιλτράρισμα των stop-words σε διάφορα πεδία, όπως είναι η συνόψιση κειμένου και η ταξινόμηση (classification).

Πώς γίνεται ο υπολογισμός:

Το βάρος tf-idf αποτελείται από δύο όρους. Ο πρώτος υπολογίζει την κανονικοποιημένη Συχνότητα Όρου (TF(Term Frequency)) δηλαδή των αριθμό των φορών που εμφανίζεται μια λέξη σε ένα έγγραφο (document), διαιρούμενο με τον συνολικό αριθμό των λέξεων σε αυτό το document. Ο δεύτερος όρος είναι η αντίστροφη συχνότητα εγγράφων (IDF (Inverse Document Frequency)), που υπολογίζεται ως ο λογάριθμος του αριθμού των document στο corpus διαιρούμενο με τον αριθμό των document όπου εμφανίζεται ο συγκεκριμένος όρος.

- **TF:** Συχνότητα Όρου, η οποία μετρά τη συχνότητα εμφάνισης ενός όρου σε ένα έγγραφο. Δεδομένου ότι κάθε έγγραφο (document) έχει διαφορετικό μήκος (length), είναι πιθανό ότι ένας όρος θα εμφανιζόταν πολύ περισσότερες φορές σε μακρά έγγραφα (document) παρά σε μικρότερα έγγραφα. Επομένως, η Συχνότητα Όρου διαιρείται με το μήκος του εγγράφου (που είναι ο συνολικός αριθμός όρων στο έγγραφο), και αυτός στην ουσία είναι ένας τρόπος κανονικοποίησης.

TF (t) = (Αριθμός των φορών που εμφανίζεται ο όρος t σε ένα έγγραφο) / (Συνολικός Αριθμός όρων στο έγγραφο).

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{i,j}}$$

Εικόνα 93.

- **IDF:** Αντίστροφη Συχνότητα Εγγράφων, η οποία μετρά πόσο σημαντικός είναι ένας όρος. Κατά τον υπολογισμό του TF, όλοι οι όροι θεωρούνται εξίσου σημαντικοί. Ωστόσο, είναι γνωστό ότι ορισμένοι όροι, όπως "είναι", "και", "από" και "ότι", μπορεί να εμφανίζονται πολλές φορές, αλλά δεν έχουν μεγάλη σημασία. Επομένως, πρέπει να μειώσουμε το βάρος των συχνών αυτών όρων και να αυξήσουμε το βάρος των όρων που εμφανίζονται πιο σπάνια και αυτό το υπολογίζουμε ως εξής:

$IDF(t) = \log_e$ (Συνολικός αριθμός εγγράφων / Αριθμός εγγράφων με τον όρο t σε αυτά).

$$idf(w) = \log\left(\frac{N}{df_t}\right)$$

Εικόνα 94.

Το TF-IDF είναι ο πολλαπλασιασμός της συχνότητας όρου και της αντίστροφης συχνότητας εγγράφων.

Παράδειγμα:

Ας υποθέσουμε ότι ένα έγγραφο περιέχει 100 λέξεις και στο έγγραφο αυτό η λέξη γάτα εμφανίζεται 3 φορές. Η συχνότητα όρου (δηλ., Tf) για την λέξη γάτα είναι $(3/100) = 0,03$. Τώρα, ας υποθέσουμε ότι έχουμε 10 εκατομμύρια έγγραφα και η λέξη γάτα εμφανίζεται σε χίλια από αυτά τα έγγραφα. Η αντίστροφη συχνότητα εγγράφου (δηλ. Idf) υπολογίζεται ως $\log(10,000,000 / 1,000) = 4$. Επομένως, το βάρος Tf-idf είναι το προϊόν αυτών των ποσοτήτων: $0,03 * 4 = 0,12$.

Στην Εικόνα 95. που ακολουθεί βλέπουμε μια απεικόνιση των όσων προαναφέραμε για τον υπολογισμό TF – IDF για 2 προτάσεις, πρόταση 1 και πρόταση 2.

sentence 1	earth is the third planet from the sun				
sentence 2	Jupiter is the largest planet				
Word	TF (Sentence 1)	TF (Sentence 2)	IDF	TF*IDF (sentence 1)	TF*IDF (Sentence 2)
earth	1/8	0	$\log(2/1)=0$	0.0375	0
is	1/8	1/5	$\log(2/2)=0$	0	0
the	2/8	1/5	$\log(2/2)=0$	0	0
third	1/8	0	$\log(2/1)=0.3$	0.0375	0
planet	1/8	1/5	$\log(2/2)=0$	0	0
from	0	0	$\log(2/1)=0.3$	0	0
sun	1/8	0	$\log(2/1)=0.3$	0.0375	0
largest	0	1/5	$\log(2/1)=0.3$	0	0.06
Jupiter	0	1/5	$\log(2/1)=0.3$	0	0.06

Εικόνα 95.

Από ότι βλέπουμε:

Το TF-IDF είναι μηδέν για stop words που δεν βοηθούν στην κατανόηση του περιεχόμενου ενός εγγράφου π.χ.:

- is
- the
- from

Το TF-IDF δεν είναι μηδέν για σημαντικές λέξεις όπως είναι:

- earth
- jupiter
- Sun
- largest
- third

Όσο περισσότερα έγγραφα εισάγουμε στο σύστημα TF-IDF, τόσο πιο πολύ αυξάνεται η ακρίβεια υπολογισμού του TF-IDF.

Κεφάλαιο 9. Topic Modeling με Latent Dirichlet Allocation (LDA)

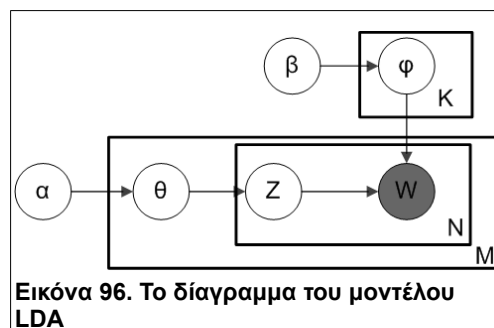
Η μοντελοποίηση θέματος (Topic Modeling) είναι η διαδικασία εντοπισμού θεμάτων σε ένα σύνολο εγγράφων (documents). Η μοντελοποίηση αυτή χρησιμοποιείται σε διάφορους τομείς όπως για παράδειγμα στις μηχανές αναζήτησης, στην αυτοματοποίηση της εξυπηρέτησης πελατών και σε οποιαδήποτε άλλη περίπτωση όπου η γνώση των θεμάτων ενός συνόλου εγγράφων είναι απαραίτητη. Για να το επιτύχουμε αυτό υπάρχουν πολλές μέθοδοι. Στην παρούσα πτυχιακή θα χρησιμοποιήσουμε μια εξ αυτών, την μοντελοποίηση θέματος Latent Dirichlet Allocation (LDA).

Το LDA (Latent Dirichlet Allocation)

Το LDA είναι μια μορφή μη εποπτευόμενης μάθησης που θεωρεί τα έγγραφα ως σάκους λέξεων (δηλ. η σειρά των λέξεων δεν έχει σημασία). Το LDA λειτουργεί αρχικά κάνοντας μια βασική υπόθεση: το έγγραφο δημιουργείται από την επιλογή αρχικά ενός συνόλου θεμάτων και έπειτα για κάθε θέμα επιλέγεται ένα σύνολο λέξεων. Και για να μπορεί να βρει θέματα, αντιστρέφει την διαδικασία αυτή ως εξής. Για κάθε έγγραφο m :

1. Υποθέτουμε ότι υπάρχουν k θέματα σε κάθε έγγραφο m ξεχωριστά.
2. Κατανέμουμε αυτά τα k θέματα στο έγγραφο m (αυτή η κατανομή είναι γνωστή ως α και μπορεί να είναι συμμετρική ή ασύμμετρη), εκχωρώντας σε κάθε λέξη ένα θέμα.
3. Για κάθε λέξη w στο έγγραφο m , υποθέτουμε ότι το θέμα της είναι λάθος, αλλά σε κάθε άλλη λέξη αντιστοιχεί το σωστό θέμα.
4. Πιθανοτικά εκχωρούμε στη λέξη w ένα θέμα που βασίζεται σε δύο πράγματα:
 - ποια θέματα βρίσκονται στο έγγραφο m .
 - πόσες φορές έχει εκχωρηθεί στην λέξη w ένα συγκεκριμένο θέμα στο σύνολο των εγγράφων (αυτή η κατανομή ονομάζεται β).
5. Η διαδικασία αυτή επαναλάβετε αρκετές φορές για κάθε έγγραφο m .

Το μοντέλο LDA



Στην Εικόνα 96, που βλέπουμε στην προηγούμενη σελίδα, παρατίθεται το διάγραμμα ενός μοντέλου LDA όπου στο διάγραμμα:

- α είναι οι κατανομές θέματος ανά έγγραφο,
- β είναι η κατανομή λέξεων ανά θέμα,
- θ είναι η κατανομή θέματος για το έγγραφο m ,
- φ είναι η κατανομή λέξεων για το θέμα k ,
- z είναι το θέμα για την νιοστή λέξη στο έγγραφο m , και
- w είναι μια συγκεκριμένη λέξη.

ΣΥΜΠΕΡΑΣΜΑΤΑ

Απώτερος στόχος της συγκεκριμένης πτυχιακής εργασίας ήταν να αποδείξουμε ότι η συλλογή δεδομένων από τις δύο εκ των πιο δημοφιλών εφαρμογών αποστολής μηνυμάτων είναι δυνατή. Τα δεδομένα που συλλέξαμε ήταν δεδομένα από τυχαία επιλεγμένες ομάδες τόσο από την εφαρμογή Telegram όσο και από την εφαρμογή WhatsApp. Εάν τα δεδομένα αυτά συλλέγονται με έναν προκαθορισμένο στόχο, τότε προσφέρονται τεράστιες δυνατότητες και θα άνοιγαν νέες πτυχές σε διάφορους τομείς της έρευνας.

Στην έρευνα προσπάθησα να συλλέξω δεδομένα από αγγλόφωνα groups για να είναι δυνατή η κατανόηση του περιεχομένου τους, δεδομένου του ότι ένα πολύ μεγάλο ποσοστό των ομάδων τόσο στο Telegram όσο και στο WhatsApp χρησιμοποιούν ινδικές διαλέκτους και γλώσσες άλλων ασιατικών χωρών. Όπως προαναφέραμε και οι δύο εφαρμογές με τις οποίες ασχοληθήκαμε έχουν μια συνεχή αύξηση στον αριθμό των χρηστών τους παγκοσμίως κάτι που κάνει την συλλογή των δεδομένων τους ένα πανίσχυρο εργαλείο έρευνας για πολλούς τομείς. Έχοντας όλα αυτά υπόψη θα αναφερθούν ορισμένοι τομείς για τους οποίους πιθανόν η συλλογή των δεδομένων των δύο αυτών εφαρμογών θα μπορούσε να φανεί χρήσιμη τα επόμενα χρόνια.

1. Μία από τις μεγαλύτερες ανησυχίες που εκφράζουν οι γονείς νεαρών χρηστών του διαδικτύου είναι η εύκολη πρόσβαση σε πορνογραφικό υλικό που κυκλοφορεί τόσο στο διαδίκτυο όσο και στα μέσα κοινωνικής δικτύωσης. Ειδικά μέσω των μέσων κοινωνικής δικτύωσης οι νεαροί χρήστες εκτίθενται σε μεγαλύτερους κινδύνους καθώς μέσω των δικτύων αυτών έχουν άμεση επικοινωνία με άλλα μέλη των ομάδων (groups) που θα μπορούσαν να έχουν ως σκοπό την έκθεση των παιδιών τόσο σε πορνογραφικό υλικό όσο και σε πιθανές ασελγείς πράξεις στα ίδια τα νεαρά άτομα, κάτι το οποίο θα μπορούσε να τραυματίσει ανεπανόρθωτα των ψυχισμό των νεαρών χρηστών, σε βαθμό που θα επηρέαζε άμεσα τη στάση και τη συμπεριφορά τους.

Η συλλογή και η ανάλυση δεδομένων μέσω κοινωνικής δικτύωσης, θα μπορούσε να αποτρέψει τους κακόβουλους χρήστες να προβούν σε αυτού του είδους κακόβουλες ενέργειες, με αποτέλεσμα την μείωση του κινδύνου έκθεσης των νεαρών χρηστών.

2. Στα επόμενα χρόνια ο τομέας του μάρκετινγκ θα γίνει ακόμα πιο ανταγωνιστικός. Όσοι επιθυμούν να καταλήξουν σε μια κατάλληλη στρατηγική μάρκετινγκ θα πρέπει να αντλήσουν βαθύτερες γνώσεις σχετικά με την αγορά και να βασίσουν τις αποφάσεις τους περισσότερο σε δεδομένα παρά σε άλλες πτυχές.

Για το σκοπό αυτό, το μέλλον του μάρκετινγκ συνδέεται στενά με την ανάλυση των δεδομένων που συγκεντρώνονται από διάφορους δικτυακούς τόπους πολυμέσων, τις πλατφόρμες κοινωνικών μέσων κλπ. Με αυτήν την στρατηγική, τα μελλοντικά οφέλη από την συλλογή δεδομένων για τον τομέα του μάρκετινγκ θα είναι τεράστια, και εκείνοι που επενδύουν σε αυτόν τον τομέα από νωρίς θα αναδειχθούν ηγέτες στην αγορά του μάρκετινγκ.

3. Ένας άλλος τομέας που θα επωφεληθεί από την άντληση και την επεξεργασία δεδομένων είναι η ανάλυση συναισθημάτων (sentiment analysis). Προς το παρόν, η τάση αυτή έχει αρχίσει να παίζει κάποιον ρόλο στην επίτευξη μιας στρατηγικής σε διάφορους τομείς της αγοράς και όχι μόνο, και στο μέλλον ο ρόλος αυτός αναμένεται να αποκτήσει μεγαλύτερη δυναμική στη λήψη αποφάσεων.

Για να δώσουμε ένα παράδειγμα, ας υποθέσουμε ότι μια εταιρεία λανσάρει ένα νέο προϊόν ή μια νέα υπηρεσία. Πώς θα αναλύσει η εταιρία την αποτελεσματικότητά του

προϊόντος και πώς θα αντλήσει ιδέες σχετικά με τον μελλοντικό σχεδιασμό του προϊόντος/υπηρεσίας; Αυτή τη στιγμή υπάρχουν εταιρείες που αναλύουν τα σχόλια χρηστών για να μάθουν κάτι καινούργιο σχετικά με τα προϊόντα τους, αλλά η πρακτική δεν είναι και τόσο διαδεδομένη. Αυτό στο μέλλον πρόκειται να αλλάξει, και η ανάλυση συναισθημάτων από δεδομένα που θα συλλέγονται από διάφορους δικτυακούς τόπους πολυμέσων και πλατφόρμες κοινωνικών μέσων θα παίζει μεγάλο ρόλο στην επίτευξη μιας πολιτικής και μιας στρατηγικής σε μια εταιρία. Οι εταιρείες που θα επενδύσουν στον τομέα αυτό, στο μέλλον θα αποκομίσουν τεράστια οφέλη όσον αφορά την ανάλυση συναισθημάτων και θα είναι σε θέση έχουν μια ευρύτερη εικόνα για τις προσδοκίες και την συνολική συμπεριφορά των πελατών τους.

Αν και τα προαναφερθέντα υπερβαίνουν το πεδίο εφαρμογής αυτής της πτυχιακής εργασίας, αυτά αναφέρονται για να δείξουμε τις δυνατότητες που μας προσφέρονται εάν μπορούμε να συλλέγουμε δεδομένα χρηστών και μάλιστα δεδομένα εφαρμογών όπως είναι το Telegram και το WhatsApp με τεράστιο αριθμό χρηστών και με παγκόσμια διασπορά.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Kiran Garimella, Gareth Tyson. WhatsApp, Doc? A First Look at WhatsApp Public Group Data. Ανακτήθηκε από <https://users.ics.aalto.fi/kiran/content/whatsapp.pdf>
2. Mansoor Iqbal. WhatsApp Revenue and Usage Statistics (2019). (19 Feb, 2019). Ανακτήθηκε από <https://www.businessofapps.com/data/whatsapp-statistics/>
3. Mansoor Iqbal. Telegram Revenue and Usage Statistics (2019). (6 Nov 2019). Ανακτήθηκε από <https://www.businessofapps.com/data/telegram-statistics/>
4. MEHVISH. WhatsApp Group vs Broadcast: What's the Difference. (05 Jul 2018). Ανακτήθηκε από <https://www.guidingtech.com/whatsapp-group-vs-broadcast/>
5. fastsms.co.uk. A Brief History of the SMS Text Message. (11 May 2017). Ανακτήθηκε από <https://fastsms.co.uk/blog/sms-messaging-history>
6. Mukul Hase (mukulhase). WebWhatsapp-Wrapper. (18th May 2018). Ανακτήθηκε από <https://github.com/mukulhase/WebWhatsapp-Wrapper>
7. Mikhail Salnikov. Text Clustering with K-means and tf-idf. (6 Aug 2018). Ανακτήθηκε από <https://medium.com/@MSalnikov/text-clustering-with-k-means-and-tf-idf-f099bcf95183>
8. ProWebScraper. 3 Predictions about Future of Web Scraping. Ανακτήθηκε από <https://prowebscraper.com/blog/future-of-web-scraping/>
9. ROBOT NINJA. Jason, Based in Canberra Australia. Team lead and coffee aficionado. An Introduction to Using Selenium-Docker Containers for End-to-End Testing. (December 6, 2017). Ανακτήθηκε από <https://robotninja.com/blog/introduction-using-selenium-docker-containers-end-end-testing/>
10. Alexei Novikov. Visualizing a cloud of skills with Python. (22 Feb 2016). Ανακτήθηκε από <http://www.dzhaworks.com/blog/visualizing-a-cloud-of-skills-with-python/>
11. Jiayu Yi. Introduction to the Telegram API. (24 Oct 2017). Ανακτήθηκε από <https://towardsdatascience.com/introduction-to-the-telegram-api-b0cd220dbed2>
12. Azika Amelia. K-Means Clustering: From A to Z. Everything you need to know about K-means clustering. (27 Sep 2018). Ανακτήθηκε από <https://towardsdatascience.com/k-means-clustering-from-a-to-z-f6242a314e9a>
13. tf-idf :: A Single-Page Tutorial - Information Retrieval and Text Mining. Ανακτήθηκε από <http://www.tfidf.com/>

14. Tyler Doll. LDA Topic Modeling: An Explanation. (25 Jun 2018). Ανακτήθηκε από <https://towardsdatascience.com/lda-topic-modeling-an-explanation-e184c90aadcd>
15. Aneesha Bakharia. Topic Modeling with Scikit Learn. (1 Sep 2016). Ανακτήθηκε από <https://medium.com/mlreview/topic-modeling-with-scikit-learn-e80d33668730>
16. Nikolaos Lykousas, Constantinos Patsakis. Large-scale analysis of grooming in modern social networks. (16 April 2020). Ανακτήθηκε από <https://medium.com/mlreview/topic-modeling-with-scikit-learn-e80d33668730>
17. Nikolaos Lykousas, Vicenc, Gomez, Constantinos Patsakis. Adult content in Social Live Streaming Services: Characterizing deviant users and relationships . (27 Jun 2018). Ανακτήθηκε από <https://medium.com/mlreview/topic-modeling-with-scikit-learn-e80d33668730>