



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
UNIVERSITY OF PIRAEUS

Τμήμα Οργάνωσης και Διοίκησης Επιχειρήσεων

Διδακτορική Διατριβή Κωνσταντίνου Η. Μαραγκού με θέμα:

«Εξαγωγή, ανάλυση και απεικόνιση ποιοτικών πληροφοριών από αδόμητα
επιχειρηματικά δεδομένα»

Επιβλέπων καθηγητής: **Πέτρος Ε. Μαραβελάκης**
Αναπληρωτής καθηγητής Τμήματος Οργάνωσης και
Διοίκησης Επιχειρήσεων Πανεπιστημίου Πειραιώς

Τριμελής Επιτροπή: **Αθανάσιος Λαγοδήμος**
Καθηγητής Τμήματος Οργάνωσης και Διοίκησης
Επιχειρήσεων Πανεπιστημίου Πειραιώς

Πέτρος Ε. Μαραβελάκης
Αναπληρωτής Καθηγητής Τμήματος Οργάνωσης και
Διοίκησης Επιχειρήσεων Πανεπιστημίου Πειραιώς

Μιχαήλ Σφακιανάκης
Καθηγητής Τμήματος Οργάνωσης και Διοίκησης
Επιχειρήσεων Πανεπιστημίου Πειραιώς

Επταμελής Επιτροπή: **Νικόλαος Γεωργόπουλος**
(επιπλέον της τριμελούς) Καθηγητής Τμήματος Οργάνωσης και Διοίκησης
Επιχειρήσεων Πανεπιστημίου Πειραιώς

Νικόλαος Ζάχαρης
Καθηγητής Τμήματος Πληροφορικής και Υπολογιστών
Πανεπιστημίου Δυτικής Αττικής

Ευαγγελία Κοπανάκη
Επίκουρη Καθηγήτρια Τμήματος Οργάνωσης και
Διοίκησης Επιχειρήσεων Πανεπιστημίου Πειραιώς

Γρηγόριος Χονδροκούκης
Καθηγητής Τμήματος Βιομηχανικής Διοίκησης Και
Τεχνολογίας Πανεπιστημίου Πειραιώς

Πειραιάς 2022



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΣΧΟΛΗ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΧΕΙΡΗΜΑΤΙΚΩΝ ΚΑΙ ΔΙΕΘΝΩΝ ΣΠΟΥΔΩΝ
ΤΜΗΜΑ ΟΡΓΑΝΩΣΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ

ΒΕΒΑΙΩΣΗ ΕΚΠΟΝΗΣΗΣ ΔΙΔΑΚΤΟΡΙΚΗΣ ΔΙΑΤΡΙΒΗΣ

Δηλώνω υπεύθυνα ότι η διδακτορική διατριβή για τη λήψη του διδακτορικού τίτλου, του Τμήματος Οργάνωσης και Διοίκησης Επιχειρήσεων του Πανεπιστημίου Πειραιώς, με τίτλο "Εξαγωγή, ανάλυση και απεικόνιση ποιοτικών πληροφοριών από αδόμητα επιχειρηματικά δεδομένα" έχει συγγραφεί από εμένα αποκλειστικά και στο σύνολό της. Δεν έχει υποβληθεί ούτε έχει εγκριθεί στο πλαίσιο κάποιου άλλου διδακτορικού, μεταπτυχιακού προγράμματος ή προπτυχιακού τίτλου σπουδών, στην Ελλάδα ή στο εξωτερικό, ούτε είναι εργασία ή τμήμα εργασίας ακαδημαϊκού ή επαγγελματικού χαρακτήρα.

Δηλώνω επίσης υπεύθυνα ότι οι πηγές στις οποίες ανέτρεξα για την εκπόνηση της συγκεκριμένης εργασίας, αναφέρονται στο σύνολό τους, κάνοντας πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο: Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του διπλώματός μου.

Υπογραφή Διδάκτορα

Όνοματεπώνυμο: Μαραγκός Κωνσταντίνος

Ημερομηνία: 27 Ιανουαρίου 2022

Κείμενο αφιέρωσης και ευχαριστιών

Θα ήθελα να ευχαριστήσω θερμά για τη βοήθεια που μου παρείχαν κατά τη διάρκεια εκπόνησης της διδακτορικής μου διατριβής:

- Τον κύριο **Αθανάσιο Κουρεμένο**, ομότιμο καθηγητή του τμήματος Οργάνωσης και Διοίκησης Επιχειρήσεων.
- Τον καθηγητή του τμήματος Οργάνωσης και Διοίκησης Επιχειρήσεων κύριο **Πέτρο Μαραβελάκη**, ο οποίος ήταν και ο επιβλέπων καθηγητής της διδακτορικής διατριβής.
- Τους καθηγητές του τμήματος Οργάνωσης και Διοίκησης Επιχειρήσεων κ.κ. **Μιχαήλ Σφακιανάκη**, Κοσμήτορα της Σχολής Οικονομικών, Επιχειρηματικών, Διεθνών και Ευρωπαϊκών Σπουδών και **Αθανάσιο Λαγοδήμο**, οι οποίοι ήταν μέλη της τριμελούς επιτροπής παρακολούθησης του παρόντος διδακτορικού.

Χωρίς την αμέριστη συμπαράσταση των παραπάνω δεν θα ήταν δυνατή η ολοκλήρωση αυτής της διδακτορικής διατριβής.

Παράλληλα θα ήθελα να ευχαριστήσω τα μέλη της επταμελούς επιτροπής που ανέλαβε τη βάσανο να κρίνει το παρόν διδακτορικό. Τους ευχαριστώ γιατί με τις εύστοχες παρατηρήσεις τους βελτιώθηκε η παρούσα Θέση.

Τέλος θα ήθελα να ευχαριστήσω την προϊσταμένη της γραμματείας του τμήματος Οργάνωσης και Διοίκησης Επιχειρήσεων κυρία **Κυριακή Σταυριανίδου** και τις διοικητικές υπαλλήλους της γραμματείας για την αμέριστη βοήθεια τους σε ό,τι διοικητικό θέμα με απασχόλησε κατά τη διάρκεια της εκπόνησης της διδακτορικής διατριβής.

Κωνσταντίνος Η. Μαραγκός

Περίληψη

Ένα από τα ανοικτά προβλήματα που προσπαθεί να επιλύσει η ακαδημαϊκή κοινότητα και που μαστίζει τον τομέα του εμπορίου και της πολιτικής, είναι η ανάλυση αδόμητων δεδομένων. Από εικόνες και βίντεο, έως ήχο και κείμενο, τα αδόμητα δεδομένα αποτελούν πλειοψηφία έναντι των δομημένων. Παρά το γεγονός αυτό και ενώ τα αδόμητα δεδομένα περιέχουν σημαντικό πλούτο πληροφορίας συγκριτικά με τα δομημένα, η εξαγωγή αυτής με ακρίβεια και ουσία, παραμένει πρόκληση για την ενημερωμένη και επιδραστική λήψη αποφάσεων.

Η παρούσα διατριβή εστιάζει στην ανάλυση κειμένου, βασιζόμενη στο συναίσθημα που εμπεριέχεται σε αυτό. Η μεθοδολογική προσέγγιση που παρουσιάζεται, δίνει τη δυνατότητα εντοπισμού των συναισθηματικών κανονικοτήτων που προσπαθεί να εγείρει ο συγγραφέας και ύστερα, επιχειρεί την κατηγοριοποίηση των κειμένων, με βάση τα συναισθήματα ως διανυσματικές συνιστώσες. Η μέθοδος παραμένει εξίσου αποτελεσματική ανεξαρτήτως περιεχομένου (ειδησιογραφία, marketing, πολιτική και λοιπά) και μήκους κειμένου.

Στα πλαίσια της έρευνας χρησιμοποιείται λεξικό για τον εντοπισμό των συναισθημάτων των λέξεων, καθώς επίσης και διάφορα μοντέλα μη εποπτευόμενης μηχανικής μάθησης. Συγκεκριμένα γίνεται χρήση της LDA (Latent Dirichlet Allocation), η οποία σε πολλές περιπτώσεις ανακαλύψαμε ότι παράγει «συναισθηματικό απόσταγμα», όπως θα δούμε. Επίσης, γίνεται χρήση των μοντέλων Mahalanobis, One Class SVM και Isolation Forest, με σκοπό τον εντοπισμό ακραίων κειμένων (anomalies). Η εφαρμογή των προαναφερθέντων πραγματοποιείται επί ενός περίπλοκου συνόλου δεδομένων, το οποίο αποτελείται κατά κύριο μέρος από ειδησιογραφικά άρθρα και κατά δεύτερο λόγο, από προπαγανδιστικά κείμενα τους Ισλαμικού Χαλιφάτου (ISIS). Τα προπαγανδιστικά κείμενα συμπεριλαμβάνονται, καθώς το συναίσθημα αποτελεί σημαντικό πυλώνα μιας αποτελεσματικής προπαγάνδας. Η βιβλιογραφική έρευνα πεδίου περιέχει εκτενή αναφορά σε αυτό, καθώς και άλλα συστατικά στοιχεία της προπαγάνδας ανά τις δεκαετίες. Είναι άξιο αναφοράς ότι τα ευρήματα της ανάλυσης, υποδεικνύουν την ύπαρξη κοινών πρακτικών μεταξύ της δυτικής ειδησιογραφίας και της τρομοκρατικής προπαγάνδας, με σημαντικές ομοιότητες στα συναισθήματα που οι συγγραφείς των δύο πλευρών προσπαθούν να εγείρουν.

Abstract

One of the open problems that the academic community is trying to solve and that plagues the commercial and politics sector, is the analysis of unstructured data. From images and video, to sound and text, unstructured data are the majority compared to structured. Despite that fact and even though unstructured data contain a wealth of information, contrary to structured data, the extraction of said information with accuracy and meaningfulness, is still a challenge in the context of impactful decision making.

This thesis focuses on text analysis, based on the emotions that underlie it. The presented methodological approach gives the ability to spot emotional patterns that the author is trying to evoke and then, attempts to classify the documents of the corpus, based on their emotion vectors. This method is effective regardless of the context (news articles, marketing, politics, etc) as well as the length of the document.

For this research we use a lexicon that allows to identify the emotions from words, and also various unsupervised machine learning models. Specifically, LDA (Latent Dirichlet Allocation) is used, which we found out that in many occasions can produce “distilled emotions”, as we shall see. We also use Mahalanobis, One Class SVM and Isolation Forest, in an effort to identify anomalous documents. The application of the above is performed on a complex dataset, that consists mainly of news articles and then, by ISIS’ propaganda texts. The propaganda texts are included because emotions are a key pillar of an effective propaganda. The state-of-the-art research addresses that extensively, as well as other elements of propaganda along the decades. It is worth noting that the findings of the analysis suggest the existence of common practices between western news agencies and terrorist propaganda, with important similarities in the emotions that the authors are trying to evoke, from both sides.

Πίνακας Περιεχομένων

Πίνακας Περιεχομένων.....	6
Περιεχόμενες Εικόνες.....	9
Περιεχόμενοι Πίνακες.....	10
1 Εισαγωγή.....	12
1.1 Γενικά.....	12
1.2 Ο άνθρωπος στην εποχή της κοινωνικής δικτύωσης.....	16
1.3 Εισαγωγή στην ανάλυση υποκειμενικής γνώμης (sentiment analysis).....	20
1.4 Βασικές έννοιες στην ανάλυση υποκειμενικής γνώμης.....	27
1.5 Ταξινόμια της ανάλυσης υποκειμενικής γνώμης (sentiment analysis).....	30
1.6 Κατηγοριοποίηση εγγράφου σε σχέση με το Συναίσθημα (Emotion Classification of Documents).....	32
2 Βιβλιογραφική έρευνα πεδίου και μεθοδολογικό πλαίσιο ανάλυσης των δεδομένων της διατριβής.....	37
2.1 Εισαγωγή.....	37
2.2 Βιβλιογραφική έρευνα πεδίου αναφορικά με την προπαγάνδα του ISIS.....	37
2.3 Βιβλιογραφική έρευνα αναφορικά με την αξιοποίηση της απόστασης Mahalanobis και της απόστασης Hellinger στην ανάλυση υποκειμενικής γνώμης.....	47
2.4 Βιβλιογραφική μελέτη σε σχέση με τον εντοπισμό ανωμαλιών σε κείμενο με μη εποπτευόμενη μάθηση.....	52
2.5 Εισαγωγή στο μεθοδολογικό πλαίσιο της Θέσης.....	57
2.6 Το μοντέλο Latent Dirichlet Allocation (LDA).....	58
2.7 Η αξιοποίηση των λεξικών στην ανάλυση υποκειμενικής γνώμης.....	61
2.8 Γενικά περί μέτρησης της απόστασης στη μηχανική μάθηση.....	65
2.8.1 Η απόσταση Mahalanobis.....	68
2.8.2 Η απόσταση Hellinger.....	70
2.8.3 Εφαρμογή πάνω στη μέτρηση αποστάσεων.....	71
3 Η ανάλυση συναισθήματος στο χώρο της πολιτικής.....	72
3.1 Γενικά για τον υβριδικό πόλεμο.....	72
3.2 Γενικά για την προπαγάνδα.....	74
3.3 Εισαγωγή στο φαινόμενο του ISIS.....	81
3.4 Κατανοώντας την προπαγάνδα του ISIS.....	88
3.5 Μεθοδολογικό πλαίσιο.....	90
3.5.1 Συλλογή Δεδομένων.....	90
3.5.2 Προετομασία δεδομένων.....	90
3.5.3 Προεπεξεργασία.....	91

3.5.4	Η εύρεση του θέματος μέσω της Latent Dirichlet Allocation (LDA-Topic Modelling) ..	91
3.5.5	Το Λεξικό	91
3.5.6	Κανονικοποίηση.....	92
3.6	Αποτελέσματα	94
3.6.1	Βασική ανάλυση αποτελεσμάτων	94
3.6.2	Ανάλυση της πολικότητας σε επίπεδο περιοδικού	97
3.6.3	Ανάλυση συναισθήματος σε επίπεδο έκδοσης περιοδικού	97
3.6.4	Αποτελέσματα από την ανάλυση μέσω της LDA.....	100
3.6.5	Ανάλυση συναισθήματος στις λέξεις-θέματα (topic words) από την LDA.....	102
3.7	Συμπεράσματα.....	102
4	Ανίχνευση ανωμαλιών με τη χρήση μη εποπτευόμενης Μηχανικής Μάθησης και βάση το συναίσθημα	104
4.1	Εισαγωγή.....	104
4.2	Συλλογή Δεδομένων	105
4.3	Προετοιμασία Δεδομένων.....	106
4.4	Προεπεξεργασία	107
4.5	Εξαγωγή Συναισθήματος	108
4.6	Διερευνητική Ανάλυση	109
4.7	Εντοπισμός Ανωμαλιών	113
4.8	Αποτελέσματα	116
4.8.1	Αθροιστικά Δεδομένα.....	117
4.8.2	Κανονικοποιημένα Δεδομένα.....	118
4.8.3	Ερμηνεία	118
5	Συμπεράσματα.....	119
5.1	Ανασκόπηση	119
5.2	Μελλοντική έρευνα	121
6	Εφαρμογές.....	122
7	Βιβλιογραφία.....	124
8	Παράρτημα Α – Πλήθος λέξεων και κάλυψη ανά περιοδικό και έκδοση	147
9	Παράρτημα Β – Πλήθος λέξεων και διαγράμματα κάλυψης	148
10	Παράρτημα Γ – Μέση εκτίμηση συναισθήματος και πολικότητας ανά περιοδικό (Overall). 151	
11	Παράρτημα Δ – Μέση εκτίμηση συναισθήματος και πολικότητας σε επίπεδο περιοδικού. 152	
12	Παράρτημα Ε – Εκτίμηση Συναισθήματος και πολικότητας ανά έκδοση περιοδικού..... 154	
13	Παράρτημα ΣΤ – Διαγράμματα συναισθήματος και πολικότητας ανά έκδοση περιοδικού. 159	
13.1	Dabiq.....	159
13.2	Daral-Islam	174

13.3	Istok.....	184
13.4	Konstantiniyye	188
13.5	Rumiyah	195
14	Παράρτημα Ζ – Επαναλήψεις της LDA. Προσδιοριζόμενες λέξεις ανά επανάληψη.....	208
14.1	1 ^η επανάληψη της LDA από 10	208
14.2	2 ^η επανάληψη της LDA από 10	210
14.3	3 ^η επανάληψη της LDA από 10	212
14.4	4 ^η επανάληψη της LDA από 10	214
14.5	5 ^η επανάληψη της LDA από 10	216
14.6	6 ^η επανάληψη της LDA από 10	218
14.7	7 ^η επανάληψη της LDA από 10	220
14.8	8 ^η επανάληψη της LDA από 10	222
14.9	9 ^η επανάληψη της LDA από 10	224
14.10	10 ^η επανάληψη της LDA από 10	226
15	Παράρτημα Η – Σχετικά Ιστογράμματα Συναισθημάτων (ISIS / Media)	228
16	Παράρτημα Θ – Θηκογράμματα Συναισθημάτων.....	230
17	Παράρτημα Ι – Ιστογράμματα Πλήθους Λέξεων	232
17.1	abcnews.go.com	232
17.2	www.cnn.com	233
17.3	www.huffingtonpost.com.....	234
17.4	www.bbc.co.uk	235
17.5	tass.com	236
17.6	www.dw.com	237
17.7	www.aljazeera.com	238
17.8	europe.chinadaily.com.cn.....	239
17.9	www.rte.ie	240
17.10	Al Hayat Media Center	241
18	Παράρτημα Κ – Δείκτες Επίδοσης Μοντέλων ανά Αναδίπλωση (Αθροιστικά Συναισθήματα) 242	
18.1	Mahalanobis – Εκπαίδευση	242
18.2	Mahalanobis – Τεστ	243
18.3	ocSVM – Εκπαίδευση.....	244
18.4	ocSVM – Τεστ.....	245
18.5	Isolation Forest – Εκπαίδευση	246
18.6	Isolation Forest – Τεστ	247
18.7	Συγκεντρωτικά Στατιστικά	247

19	Παράρτημα Λ – Δείκτες Επίδοσης Μοντέλων ανά Αναδίπλωση (Κανονικοποιημένα Συναισθήματα)	248
19.1	Mahalanobis – Εκπαίδευση	248
19.2	Mahalanobis – Τεστ	249
19.3	ocSVM – Εκπαίδευση.....	250
19.4	ocSVM – Τεστ	251
19.5	Isolation Forest – Εκπαίδευση	252
19.6	Isolation Forest – Τεστ	253
19.7	Συγκεντρωτικά Στατιστικά	253

Περιεχόμενες Εικόνες

Εικόνα 1	Πλήθος συσκευών IoT (σε δις) κατ' έτος. Πηγή δεδομένων: NTCA.....	12
Εικόνα 2	Η αναλογία δομημένων και αδόμητων δεδομένων στο σύνολο των υπό ανάλυση δεδομένων.	15
Εικόνα 3	Η διεπαφή στο ηλεκτρονικό κατάστημα της Amazon. Φαίνεται η αξιολόγηση που απορρέει από άλλους χρήστες καθώς και οι προτάσεις προϊόντων που ενδιαφέρουν τον επισκέπτη με βάση την κατηγοριοποίηση που υλοποιεί το λογισμικό της Amazon.	20
Εικόνα 4	Δηλώσεις του Ιάπωνα πρωθυπουργού από τα μέσα κοινωνικής δικτύωσης με αφορμή τον σεισμό της 13 Φεβρουαρίου 2021 στην περιοχή της Fukushima.	24
Εικόνα 5	Ανάρτηση των The Japan Times αναφορικά με το σεισμό των 7,3 ρίχτερ στη περιοχή της Fukushima στις 13 Φεβρουαρίου 2021.	25
Εικόνα 6	Συνηθισμένα Emojis.....	27
Εικόνα 7	Ανάρτηση του προέδρου Biden στο μέσο κοινωνικής δικτύωσης Twitter.	28
Εικόνα 8	Ανάρτηση του CNN αναφορικά με το σεισμό της 12 ^{ης} Φεβρουαρίου 2021 στην Ιαπωνία.	29
Εικόνα 9	Ανάρτηση πολίτη αναφορικά με το σεισμό της 12 ^{ης} Φεβρ. 2021 στην Ιαπωνία	29
Εικόνα 10	Ανάρτηση του Ιάπωνα πρωθυπουργού για το σεισμό της 12 ^{ης} Φεβρ. 2021 στην Ιαπωνία.....	29
Εικόνα 11	Κατηγορίες Συναισθημάτων	32
Εικόνα 12	Τα βασικά συναισθήματα κατά Ekman. Η εικόνα έχει ληφθεί από το research-gate	34
Εικόνα 13	Ο κύκλος των συναισθημάτων του Plutchnik.....	34
Εικόνα 14	Τα βασικά συναισθήματα κατα OCC	35
Εικόνα 15	Μαχητής του ISIS. Φανερή η αντιγραφή της ιδεασ από την πρώτη εικόνα του πίνακα 8	88
Εικόνα 16	εξωφυλλο του dabiq, με αφορμη τρομοκρατικη επιθεση στο παρισι.....	89
Εικόνα 17	Μεταβολή του πλήθους λέξεων ανά έκδοση του Dabiq και η κάλυψη από το EmoLex	95
Εικόνα 18	Πλήθος λέξεων ανά έκδοση του Rumiyah και κάλυψη από το EmoLex	96
Εικόνα 19	Μέση πολικότητα ανά περιοδικό υπολογισμένη σε όλες τις εκδόσεις του.	97
Εικόνα 20	Μέσος ανά συναίσθημα υπολογισμένος σε όλη τη συλλογή.....	98
Εικόνα 21	Μέση ένταση ανά συναίσθημα σε επίπεδο περιοδικού.....	99
Εικόνα 22	Αραχνοειδές διάγραμμα για όλες τις λέξεις-θέματα της LDA.....	102
Εικόνα 23	Ποσοστό άρθρων Media έναντι ISIS	106
Εικόνα 24	Ποσοστό άρθρων ανά πηγή.....	107
Εικόνα 25	Ποσοστό άρθρων ανά πηγή μετά την αφαίρεση άρθρων με <30 λέξεις	108
Εικόνα 26	Ιστογράμματα ισχύος των τριών κύριων συναισθημάτων (αθροιστικά αριστερά, κανονικοποιημένα δεξιά)	111

Εικόνα 27 Πλήθος λέξεων μετά την προεπεξεργασία και πλήθος λέξεων που ενυπάρχουν στο EmoLex ανά έκδοση του Dabiq.....	148
Εικόνα 28 Πλήθος λέξεων μετά την προεπεξεργασία και πλήθος λέξεων που ενυπάρχουν στο EmoLex ανά έκδοση του Daral-Islam.....	148
Εικόνα 29 Πλήθος λέξεων μετά την προεπεξεργασία και πλήθος λέξεων που ενυπάρχουν στο EmoLex ανά έκδοση του Istok	149
Εικόνα 30 Πλήθος λέξεων μετά την προεπεξεργασία και πλήθος λέξεων που ενυπάρχουν στο EmoLex ανά έκδοση του Konstantiniyye	149
Εικόνα 31 Πλήθος λέξεων μετά την προεπεξεργασία και πλήθος λέξεων που ενυπάρχουν στο EmoLex ανά έκδοση του Rumiyaħ.....	150
Εικόνα 32 Μέση εκτίμηση συναισθήματος υπολογισμένη σε όλα τα περιοδικά.....	151
Εικόνα 33 Μέση πολικότητα στο σύνολο των περιοδικών	152
Εικόνα 34 Μέση εκτίμηση συναισθήματος ανά περιοδικό	152
Εικόνα 35 Μέση τιμή πολικότητας ανά περιοδικό.....	153

Περιεχόμενοι Πίνακες

Πίνακας 1 Το επιθυμητό εξαγόμενο της ανάλυσης θεμάτων σε κείμενα	58
Πίνακας 2 Σχετικές συχνότητες εμφάνισης όρων ανά κείμενο (παράδειγμα)	59
Πίνακας 3 Τα λεξικά εξόρυξης γνώμης εν συντομία	62
Πίνακας 4 Τα λεξικά συναισθημάτων εν συντομία.....	63
Πίνακας 5 Πίνακας υπολογισμού της απόστασης Hellinger μεταξύ δυο κατανομών πιθανοτήτων...	72
Πίνακας 6 Βρετανικές αφίσες του 1 ^{ου} Παγκοσμίου Πολέμου με σκοπό να καλλιεργήσουν κατάλληλο κλίμα στους πολίτες ώστε να καταταγούν οι αντρες στα όπλα	77
Πίνακας 7 Γερμανικές προπαγανδιστικές αφίσες του 2ου παγκοσμίου πολέμου	78
Πίνακας 8 Προπαγανδιστικές αφίσες του 2ου παγκόσμιου πολέμου από τις ΗΠΑ.....	79
Πίνακας 9 Βρετανικές προπαγανδιστικές αφίσες από το 2ο Παγκόσμιο Πόλεμο.....	80
Πίνακας 10 Προπαγανδιστικές αφίσες από την περίοδο του ψυχρού πολέμου.....	80
Πίνακας 11 Προπαγανδιστικές αφίσες από τους πολέμους στην Κορέα και το Βιετνάμ.....	81
Πίνακας 12 Στοιχεία των εκδόσεων του al-Hayat Media Center	90
Πίνακας 13 Κανονικοποίηση διανυσμάτων των λέξεων του Λεξικού EmoLex	94
Πίνακας 14 Πίνακας αραχνοειδών διαγραμμάτων τεσσάρων διαφορετικών εκδόσεων των περιοδικών.....	100
Πίνακας 15 Άρθρα ανά πηγή	107
Πίνακας 16 Συνδιακύμανση Αθροιστικών Συναισθημάτων Media.....	112
Πίνακας 17 Συνδιακύμανση Αθροιστικών Συναισθημάτων ISIS	112
Πίνακας 18 Συνδιακύμανση Κανονικοποιημένων Συναισθημάτων Media.....	113
Πίνακας 19 Συνδιακύμανσης Κανονικοποιημένων Συναισθημάτων ISIS.....	113
Πίνακας 20 Μέσος Ορος δεικτων επίδοσης μοντελου (αθροιστικα συναισθηματα).....	115
Πίνακας 21 Μεσος Ορος δεικτων επιδοσης μοντελου (κανονικοποιημενα δεδομενα).....	115
Πίνακας 22 Πλήθος λέξεων και κάλυψη από το EmoLex ανά περιοδικό και έκδοση	147
Πίνακας 23 Μέσες τιμές εκτίμησης συναισθήματος και πολικότητας ανά περιοδικό.....	151

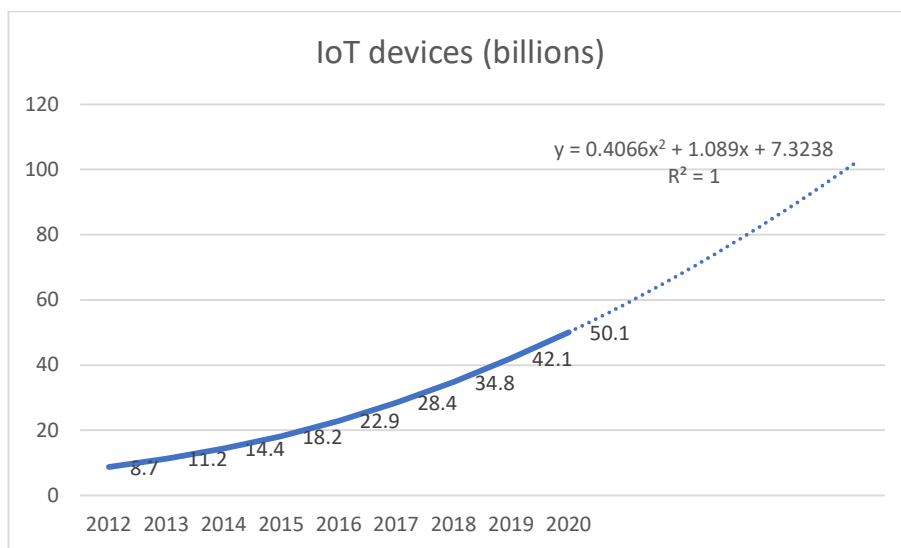
Κρείσσον ελέσθαι ψεύδος, ἢ ἀληθές κακόν.

Μένανδρος, 4ος αιώνας π.Χ.

1 Εισαγωγή

1.1 Γενικά

Τα τελευταία χρόνια σημειώνεται μια ραγδαία ανάπτυξη των Τεχνολογιών Πληροφορικής και Επικοινωνιών (εν συντομία ΤΠΕ ή ICT). Το Διαδίκτυο των Πραγμάτων (Internet of Things, σύντομα IoT) έχει ενσωματωθεί σε πολλές από τις συσκευές που χρησιμοποιούμε και επηρεάζει δραστικά τον τρόπο με τον οποίο ενεργούμε και αποφασίζουμε. Μερικές από τις συσκευές αυτές είναι τα έξυπνα κινητά τηλέφωνα (smartphones), οι κάμερες παρακολούθησης (CCTV), οι διάφοροι αισθητήρες (που αξιοποιούνται στην παρακολούθηση της κυκλοφορίας, των αγροτικών δραστηριοτήτων κλπ), οι έξυπνες οικιακές συσκευές κλπ. Σύμφωνα με δεδομένα από τον National Telecommunications Cooperative Association (NTCA), στο διάστημα από 2012 έως 2020 το πλήθος των συσκευών IoT αυξήθηκε κατά 475%. Με βάση τα στοιχεία αυτά, το πλήθος των IoT συσκευών αναμένεται κατά το έτος 2025 να είναι της τάξης των 100 δις (Εικόνα 1).



ΕΙΚΟΝΑ 1 ΠΛΗΘΟΣ ΣΥΣΚΕΥΩΝ ΙΟΤ (ΣΕ ΔΙΣ) ΚΑΤ' ΕΤΟΣ. ΠΗΓΗ ΔΕΔΟΜΕΝΩΝ: NTCA

Η χρήση των παραπάνω συσκευών έχει σαν αποτέλεσμα τη δημιουργία δεδομένων μεγάλου όγκου. Τα δεδομένα αυτά υφίστανται επεξεργασία από τις συσκευές IoT και συνήθως, επιχειρησιακές αποφάσεις λαμβάνονται σε πραγματικό χρόνο. Ταυτόχρονα, και η επικοινωνία των ανθρώπων, μέσω του διαδικτύου, αποτελεί μια πηγή μεγάλου όγκου δεδομένων, τα οποία αξίζει να αναλυθούν.

Είναι κοινά αποδεκτό (NIST, 2019) μεταξύ των εκπροσώπων:

- Των Επιχειρήσεων,
- Της Ακαδημαϊκής κοινότητας και
- Των Κυβερνήσεων

ότι η ανάλυση των δεδομένων αυτών μπορεί να δράσει καταλυτικά στη δημιουργία ανάπτυξης και καινοτομίας.

Τα δεδομένα μεγάλου όγκου (Big Data) αποθηκεύονται αλλά και αντλούνται μέσω εκτεταμένων βάσεων δεδομένων και έχει παρατηρηθεί ότι το μέγεθός τους αυξάνεται εκθετικά με τον χρόνο. Η επεξεργασία των δεδομένων αυτών δεν μπορεί να γίνει με τους παραδοσιακούς τρόπους και επιβάλλεται η αξιοποίηση μεθοδολογιών που ανήκουν στην Εξόρυξη Δεδομένων, τη Μηχανική Μάθηση και την Τεχνητή Νοημοσύνη. Τα βασικά χαρακτηριστικά τους είναι:

- **Ο όγκος (volume).** Όταν αναφερόμαστε στον όγκο των δεδομένων αναφερόμαστε στο μέγεθος του συνόλου των δεδομένων που δημιουργούνται σε ημερήσια βάση από τα μέσα κοινωνικής δικτύωσης, τις επιχειρηματικές δραστηριότητες κλπ. Το μέγεθος αυτό υπολογίζεται σε πολλαπλάσια της μονάδας Bytes.
- **Η ποικιλότητα (variety).** Η παράμετρος αυτή σχετίζεται με το βαθμό αξιοποίησης δεδομένων διαφορετικού τύπου που προέρχονται από διαφορετικές πηγές. Έτσι, ενώ στο παρελθόν τα δεδομένα που αντλούσαμε προέρχονταν από μια βάση δεδομένων και είχαν συγκεκριμένη μορφή (π.χ. πίνακα), σήμερα τα δεδομένα που επεξεργαζόμαστε και αναλύουμε μπορούν να προέρχονται από emails, αναρτήσεις σε μέσα κοινωνικής δικτύωσης κλπ ενώ η μορφή τους μπορεί να αφορά εικόνα, κείμενο, βίντεο κ.α.
- **Η ταχύτητα (velocity).** Το χαρακτηριστικό αυτό αφορά το ρυθμό με τον οποίο δημιουργούνται τα δεδομένα αυτά σε πραγματικό χρόνο. Το χαρακτηριστικό αυτό εμπεριέχει διάφορες εκφάνσεις ρυθμών μεταβολής των δεδομένων.
- **Η μεταβλητότητα (variability).** Η έννοια αυτή μπορεί να αφορά είτε την ασυνέπεια που εμπεριέχεται στα δεδομένα μεγάλου όγκου λόγω της ύπαρξης ανωμαλιών στις παρατηρήσεις και ακραίων τιμών είτε να αφορά τη διαφορετικότητα των διαστάσεων αυτών των ίδιων των δεδομένων λόγω της ποικιλότητας της προέλευσης τους.

- **Η ακρίβεια (veracity).** Το στοιχείο αυτό, αφορά γενικά στην ακρίβεια και ειλικρίνεια ενός συνόλου δεδομένων. Αυτό είναι επιπλέον σημαντικό στον χώρο των μεγάλων δεδομένων, με το νόημα του όρου να συμπεριλαμβάνει έννοιες όπως, η ποιότητα και αξιοπιστία της πηγής των δεδομένων, ο τύπος των δεδομένων και ο τρόπος επεξεργασίας τους.

Όλα τα παραπάνω απαιτούν κλιμακωτές αρχιτεκτονικές για αποδοτική αποθήκευση, διαχείριση και ανάλυση των δεδομένων αυτών (NIST, 2019).

Τα δεδομένα μεγάλου όγκου εμφανίζονται σε πολλούς και διαφορετικούς κλάδους της ανθρώπινης δραστηριότητας όπως είναι η Υγεία, η Ακαδημαϊκή κοινότητα, οι τεχνολογίες Πληροφορικής και Επικοινωνιών, το Εμπόριο, οι Μεταφορές και ο Χρηματοπιστωτικός κλάδος.

Θα μπορούσε κανείς να αναφέρει κάποιες επιχειρήσεις που έχουν ενσωματώσει την ανάλυση των δεδομένων μεγάλου όγκου στην καθημερινή τους πρακτική. Παραδείγματα είναι η Walmart, η American Express, η Uber, το Netflix κ.α. Όμως, η πλειονότητα των επιχειρήσεων (πέραν των τεχνολογικών ηγετών) που λειτουργούν σε παγκόσμιο επίπεδο φαίνεται να στερείται της κουλτούρας αλλά και των εργαλείων που αποσκοπούν στην αξιοποίηση των δεδομένων μεγάλου όγκου. Αυτό έχει ως αποτέλεσμα να αδυνατούν να δημιουργήσουν προγνωστικά μοντέλα, τέτοια που θα ενίσχυαν τον κύκλο εργασιών τους. Ένας σοβαρός λόγος για την υστέρηση αφορά τη διαφορετικότητα της επεξεργασίας και της ανάλυσης που απαιτείται για την εκμετάλλευση των δεδομένων αυτών.

Στην κατεύθυνση της θεραπείας των παραπάνω προβλημάτων βρίσκεται η **Μηχανική Δεδομένων** (Data Engineering), η **Αναλυτική** (Analytics) καθώς και η **Μηχανική Μάθηση** (Machine Learning). Η Αναλυτική αποτελεί ένα νέο επιστημονικό πεδίο που ανήκει στην ευρύτερη περιοχή της Λήψης Αποφάσεων. Αξιοποιεί τεχνικές από διάφορα επιστημονικά πεδία, όπως η Τεχνητή Νοημοσύνη και η Στατιστική, και στοχεύει στην συστηματική επεξεργασία, διαχείριση και ανάλυση των δεδομένων με σκοπό να αποκαλυφθούν υποκρυπτόμενα πρότυπα, σχέσεις μεταξύ των δεδομένων, ή τάσεις εξέλιξης και να οδηγήσουν στην δυνατότητα προβλέψεων (NIST, 2019). Ειδικότερα, βασικός στόχος της Αναλυτικής είναι η ανάλυση μεγάλου όγκου αδόμητων δεδομένων. Η πραγμάτωση αυτού του στόχου είναι ιδιαίτερα σημαντική στο σημερινό γίνεσθαι, μιας και έχει παρατηρηθεί εμπειρικά ότι κάθε δεκαοκτώ μήνες διπλασιάζεται το πλήθος των διαθέσιμων δεδομένων (NIST, 2019). Η δημιουργία ενός αποτελεσματικού πλαισίου επεξεργασίας και ανάλυσης των

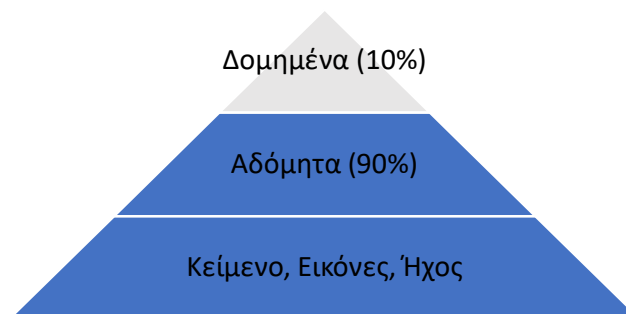
δεδομένων αυτού του είδους θα μπορούσε να αποτελέσει κρίσιμο εργαλείο στα χέρια των σημερινών διοικητών (managers) και ληπτών απόφασης (decision makers) και να αλλάξει ριζικά τον τρόπο λήψης αποφάσεων στο σύγχρονο επιχειρείν.

Αναφορικά με την έννοια των αδόμητων δεδομένων, ο σύγχρονος αναλυτής συχνά καλείται να συλλέξει, επεξεργαστεί και αναλύσει μεγάλο όγκο από δεδομένα τα οποία μπορούν να κατηγοριοποιηθούν σε δύο βασικές ομάδες, τα δομημένα και τα αδόμητα.

Λέγοντας **δομημένα δεδομένα** εννοούμε εκείνα τα δεδομένα που παράγονται με τρόπον ώστε να ακολουθούν συγκεκριμένα και προκαθορισμένα μορφότυπα. Ως εκ τούτου, τα επιμέρους στοιχεία ενός συνόλου δομημένων δεδομένων είναι όμοια μεταξύ τους. Μια συνηθισμένη πηγή προέλευσης των δεδομένων αυτών είναι ο χώρος των οικονομικών δραστηριοτήτων. Τα τιμολόγια πωλήσεων, τα χρηματοοικονομικά στοιχεία των επιχειρήσεων, τα ιστορικά στοιχεία των συνεδριάσεων ενός χρηματιστηρίου περιέχουν δομημένα δεδομένα. Ως ένα παράδειγμα δομημένων δεδομένων μπορεί να σκεφτεί κανείς το αρχείο παραγγελιών μιας επιχείρησης. Η κάθε εγγραφή σε αυτό το αρχείο περιέχει τα ίδια πεδία με τις υπόλοιπες εγγραφές (πχ. Επωνυμία πελάτη, Έδρα πελάτη, Στοιχεία υπευθύνου επικοινωνίας, κλπ) και ο τρόπος με τον οποίο καταγράφεται η πληροφορία είναι πανομοιότυπος σε όλες τις εγγραφές.

Αντίθετα, τα **αδόμητα δεδομένα** δεν ακολουθούν συγκεκριμένα μορφότυπα. Αυτά είναι συνήθως κείμενα, εικόνες και αρχεία ήχου. Παραδείγματα πηγών δεδομένων αυτής της μορφής είναι το περιεχόμενο ενός email, ηχογραφημένες / βιντεοσκοπημένες συνεντεύξεις, η επικοινωνία σε τηλεφωνικά κέντρα κλπ.

Τέλος, υπάρχει και μια ενδιάμεση κατηγορία δεδομένων που είναι τα **ημιδομημένα**. Αυτά αφορούν σε μίγμα δομημένων και αδόμητων δεδομένων (παράδειγμα οι ιστοσελίδες HTML, οι συζητήσεις μέσω κοινωνικών δικτύων και λοιπά).



ΕΙΚΟΝΑ 2 Η ΑΝΑΛΟΓΙΑ ΔΟΜΗΜΕΝΩΝ ΚΑΙ ΑΔΟΜΗΤΩΝ ΔΕΔΟΜΕΝΩΝ ΣΤΟ ΣΥΝΟΛΟ ΤΩΝ ΥΠΟ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ.

Έχει διαπιστωθεί ότι το ποσοστό των αδόμητων δεδομένων που εμφανίζεται στο σύνολο των δεδομένων μεγάλου όγκου, είναι της τάξης του 90% του συνόλου, ενώ μόνο το 10% εμφανίζεται σε μορφή δομημένων δεδομένων (Azaz, et al., 2019), (Griffith, 2017), (Deloitte, 2017). Τα παραπάνω εμφανίζονται στην Εικόνα 2.

Πολλοί ερευνητές επισημαίνουν το γεγονός ότι το Ορατό Διαδίκτυο ή Διαδίκτυο Επιφανείας έχει πεντακόσιες φορές μικρότερη «έκταση» από το Σκοτεινό Διαδίκτυο (γνωστό ως Dark Web) (Goodman, 2015), (Barton, 2016) και ότι το 37% των δεδομένων που υπάρχουν, αν αναλύονταν, θα αποκόμιζαν οι επιχειρήσεις μεγάλο κέρδος (Vesset & Schubmehl, 2016). Από τα παραπάνω είναι φανερό η σημασία να δημιουργηθεί ένα μεθοδολογικό πλαίσιο ανάλυσης των αδόμητων δεδομένων. Αναφορικά με το σύνηθες Διαδίκτυο, κρίσιμο ρόλο στη δημιουργία των δεδομένων μεγάλου όγκου, γενικότερα, και των αδόμητων δεδομένων, ειδικότερα, διαδραματίζουν οι πλατφόρμες κοινωνικής δικτύωσης όπως είναι το Facebook, το Twitter, το Instagram και το YouTube. Οι πλατφόρμες αυτές, λόγω της φιλοσοφίας που υιοθετούν σχετικά με τη συνεργατική δημιουργία του περιεχομένου τους, επιτρέπουν τη δημιουργία συζητήσεων ή/και την ανταλλαγή απόψεων. Μέσα στην επικοινωνία αυτή συμπεριλαμβάνονται, εκτός των άλλων, και ανασκοπήσεις προϊόντων και υπηρεσιών. Αυτό καθιστά τις πλατφόρμες αυτές, δεξαμενές μεγάλου όγκου πολύτιμων αδόμητων δεδομένων όπως είναι τα δεδομένα κειμένου, βίντεο εικόνας κλπ.

1.2 Ο άνθρωπος στην εποχή της κοινωνικής δικτύωσης

Η ανάπτυξη του Διαδικτύου έχει επιδράσει καθοριστικά στη διαμόρφωση του τρόπου με τον οποίο οι άνθρωποι:

- Επικοινωνούν μεταξύ τους.
- Πληροφορούνται για προϊόντα και υπηρεσίες.
- Αγοράζουν προϊόντα και υπηρεσίες.

Οι διαδικτυακές πλατφόρμες αγοράς προϊόντων (όπως η Amazon, eBay, Alibaba, κλπ) όχι μόνο επέτρεψαν στον καταναλωτή να αγοράζει χωρίς γεωγραφικούς και χρονικούς περιορισμούς (Gao, et al., 2018) αλλά κατέστησαν δυνατό στους καταναλωτές να καταθέτουν την άποψη τους αναφορικά με την συνέπεια της εταιρίας με την οποία συναλλάχθηκαν, καθώς και για την ποιότητα των παρεχόμενων προϊόντων και υπηρεσιών που αυτοί απήλασαν. Η δυνατότητα αυτή που έχει πλέον ο μέσος καταναλωτής να καταχωρεί τη γνώμη του για προϊόντα και υπηρεσίες στις εμπορικές πλατφόρμες, αλλά και στα μέσα κοινωνικής δικτύωσης, έχει αλλάξει τον τρόπο με τον οποίο οι καταναλωτές

επιλέγουν και αγοράζουν προϊόντα και υπηρεσίες. Σήμερα, οι καταναλωτές αναζητούν την άποψη που έχουν όμοιοι με αυτούς καταναλωτές, σε σχέση με προϊόντα και υπηρεσίες που προτίθενται να αγοράσουν. Ο σημερινός καταναλωτής φαίνεται να έχει αποσυνδεθεί από τη λογική που ίσχυε παλαιότερα, κατά την οποία είχε ως γνώμονα την άποψη που υιοθετούν οι ειδικοί επί του εκάστοτε θέματος, προκειμένου να επιλέξει προϊόντα και υπηρεσίες. Σήμερα, όπως ισχυρίζονται ερευνητές, ο καταναλωτής δείχνει να χρησιμοποιεί σε μεγάλο βαθμό ως πηγή πληροφόρησης τα blogs και τα μέσα κοινωνικής δικτύωσης (Wei-Fan & Lun-Wei, 2016), (Guesalaga, 2016) προκειμένου να πάρει αποφάσεις. Κατά κάποιους ερευνητές, είναι τόσο ισχυρή αυτή η μορφή επικοινωνίας, που καταλήγει στο να καθορίζουν οι πελάτες την προώθηση των αγαθών, μέσω των αποτιμήσεων που αναρτούν στα μέσα κοινωνικής δικτύωσης και να γίνονται συνδημιουργοί, συνεργάτες και προωθητές των μηνυμάτων της διαφημιστικής καμπάνιας (Hamilton, et al., 2016).

Η αλλαγή στον τρόπο που επιλέγει ο σημερινός καταναλωτής ευνοεί την ανάπτυξη του ηλεκτρονικού εμπορίου, αλλάζει τον τρόπο με τον οποίο οι επιχειρήσεις προωθούν τα προϊόντα τους και, μέσω των κοινωνικών δικτύων, οδηγεί νομοτελειακά σε ένα καινοφανή τρόπο διάδρασης των ανθρώπων αναφορικά με την επιλογή και αξιολόγηση προϊόντων και υπηρεσιών (Ravi & Ravi, 2015).

Η αλλαγή αυτή δεν έχει περάσει απαρατήρητη από τους εμπλεκόμενους με το επιχειρείν (όπως οι διαφημιστές, οι σχεδιαστές προϊόντων, κλπ), τους αποφασίζοντες στο χώρο της πολιτικής, τους ακαδημαϊκούς κλπ. Ο καθένας από αυτούς, για δικούς του λόγους, προσπάθησε να ερμηνεύσει και να αξιοποιήσει τη νέα πραγματικότητα. Οι διαφημιστές για να μπορέσουν να αξιοποιήσουν τα νέα κανάλια επικοινωνίας στην προώθηση προϊόντων, οι πολιτικοί για να περάσουν τα μηνύματα της προεκλογικής τους εκστρατείας με μεγαλύτερο επηρεασμό των ψηφοφόρων, κλπ.

Ας δούμε μερικά τέτοια παραδείγματα που επιβεβαιώνουν τα ανωτέρω:

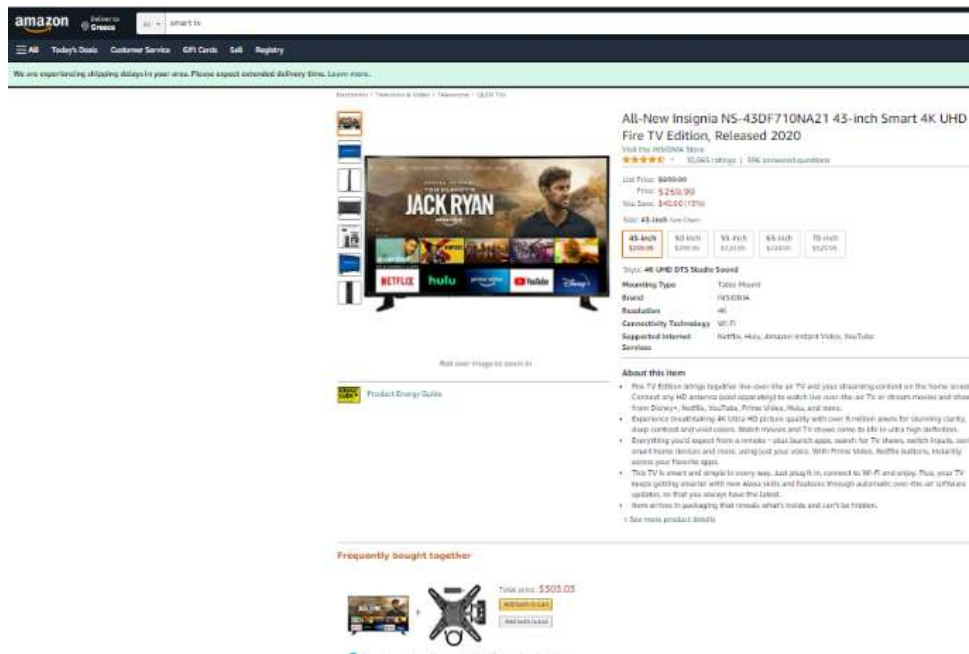
- **Στο χώρο της πολιτικής.** Είναι γνωστό σε όλους το σκάνδαλο της εταιρίας Cambridge Analytica. Σύμφωνα με τις αποκαλύψεις που είδαν το φως της δημοσιότητας, η εν λόγω εταιρία ενεπλάκη στην προεκλογική εκστρατεία του τότε υποψηφίου προέδρου των ΗΠΑ, Donald Trump (Guardian, 2018), (NewYorkTimes, 2018). Η Cambridge Analytica αφού απέσπασε τα προσωπικά δεδομένα 87 εκατομμυρίων χρηστών που ήταν δημοσιευμένα σε κοινωνικά δίκτυα όπως Facebook και Twitter, τα ανέλυσε με σκοπό να προσδιοριστούν πρότυπα (patterns) ψηφοφόρων αλλά και κρίσιμα

ζητήματα που υπόβοσκαν στις μεταξύ τους εικονικές συζητήσεις. Η ανάλυση αυτών των δεδομένων επέτρεψε στον Donald Trump να εφαρμόσει στοχευμένη προεκλογική εκστρατεία σε κάποιες πολιτείες των ΗΠΑ και να αποκτήσει προβάδισμα έναντι της Hillary Clinton στις προεδρικές εκλογές του 2016 (Grover, et al., 2017). Η Ομοσπονδιακή Επιτροπή Εμπορίου των ΗΠΑ επέβαλε πρόστιμο ύψους 5 δις δολαρίων στο Facebook για τη διαρροή των προσωπικών δεδομένων (Nuñez, 2019), (Wells, et al., 2016). Ο Donald Trump, όπως υποστηρίζουν ερευνητές, δεν περιορίστηκε στην προεκλογική εκμετάλλευση των δυνατοτήτων που του έδιναν τα μέσα κοινωνικής δικτύωσης αλλά φρόντισε και μετεκλογικά να έχει έντονη παρουσία στα μέσα κοινωνικής δικτύωσης ώστε να εξασφαλίζει την μέγιστη δυνατή προσοχή και αποδοχή των θέσεων του από τους πολίτες (Zhang, et al., 2018).

Είναι λιγότερο γνωστό ότι και ο προκάτοχος του Donald Trump, Barack Obama, είχε επίσης εκμεταλλευτεί τα εργαλεία της Τεχνητής Νοημοσύνης (Artificial Intelligence, AI) και τα μέσα κοινωνικής δικτύωσης, τόσο κατά τη διάρκεια της προεκλογικής του εκστρατείας όσο και μετεκλογικά. Συγκεκριμένα, κατά τη διάρκεια της προεκλογικής του εκστρατείας το 2012, με αντίπαλο τον Mitt Romney, προσέλαβε ως τεχνογνώστη τον καθηγητή Rayid Ghani, ειδικό σε θέματα τεχνητής νοημοσύνης. Ο τελευταίος, ηγούμενος μιας ομάδας ειδικών, ανέλυσε δεδομένα που προέρχονταν από τα κοινωνικά δίκτυα προκειμένου ο B. Obama να αποκτήσει προβάδισμα έναντι του αντιπάλου του και να κατορθώσει να εκλεγεί Πρόεδρος των ΗΠΑ (Mitchum, 2013), (Greenwald, 2013). Όπως δείχνουν σχετικές έρευνες, οι τακτικές αυτές δεν περιορίζονται στις ΗΠΑ (Vergeer & Hermans, 2013), ούτε αφορούν μόνο την πολιτική. Σύμφωνα με πολλούς ερευνητές οι πλατφόρμες κοινωνικής δικτύωσης γενικότερα και το Twitter ειδικότερα, αξιοποιούνται, πέρα από τους πολιτικούς, και από τους χρήστες και τους δημοσιογράφους, ώστε να δημιουργήσει ο καθένας από αυτούς το δικό του αφήγημα στο διαδίκτυο, αναφορικά με το ενδεδειγμένο πλαίσιο διαχείρισης κάποιων κοινωνικών ή οικονομικών προβλημάτων (Choi & Park, 2014), (Moody-Ramirez & Cole, 2018), (Van Der Meer & Verhoeven, 2013), (Merry, 2016). Μάλιστα, υπάρχουν ερευνητές που ισχυρίζονται ότι τα MME έχουν την τάση να υπερτονίζουν την αρνητική πλευρά των προβλημάτων, μέσω των μέσων κοινωνικής δικτύωσης, προκειμένου να αποκτήσουν μεγαλύτερη επισκεψιμότητα οι αναρτήσεις τους (Eshbaugh-Soha, 2010), (Dunaway, 2013), (Soroka & McAdams, 2015).

- **Στο χώρο της διαφήμισης και των πωλήσεων.** Οι πρακτικές που αξιοποιούν τεχνητή νοημοσύνη και εξασφαλίζουν στις επιχειρήσεις προηγμένες δυνατότητες προώθησης προϊόντων χρησιμοποιούνται σε μεγάλο βαθμό, ειδικά από μεγάλες επιχειρήσεις (Wertz, 2018). Τα κοινωνικά δίκτυα και οι εμπορικές πλατφόρμες έχουν επιτρέψει στις επιχειρήσεις να:
 - Απευθύνονται στους εν δυνάμει πελάτες τους με πολυμεσικά μηνύματα.
 - Αναπτύσσουν αλληλεπιδραστικά κανάλια επικοινωνίας μεταξύ τους (Mangold & Faulds, 2009).

Μέσω αυτής της πολυεπίπεδης επικοινωνίας, οι επιχειρήσεις μπορούν να αυξήσουν σημαντικά το βαθμό εμπιστοσύνης που αποδίδουν οι πελάτες στην επιχείρηση και τα προϊόντα της (Karlan & Haenlein, 2010) και να αποκτήσουν μια συνείδηση που να τους απαγκιστρώνει από την εμπιστοσύνη που παραδοσιακά έχουν σε επιχειρήσεις που είναι γεωγραφικά κοντινές τους (Gao, et al., 2018). Η ύπαρξη των μέσων κοινωνικής δικτύωσης φαίνεται να επηρεάζει ισχυρά τις πωλήσεις (Guesalaga, 2016) και να επιτρέπει στις επιχειρήσεις να δημιουργούν στενότερες σχέσεις με τους πελάτες τους (Rapp, et al., 2013). Πολλοί ερευνητές ισχυρίζονται ότι ένα μεγάλο μέρος των πωλήσεων που γίνονται παγκόσμια, εξαρτάται από τις αξιολογήσεις προϊόντων και υπηρεσιών που προέρχονται από άλλους καταναλωτές (Peng, et al., 2014). Παράλληλα, η ανάλυση των δεδομένων που αντλούνται από το διαδίκτυο επιτρέπει στις επιχειρήσεις να κατηγοριοποιούν τους πελάτες τους και να κατανοούν καλύτερα τις ανάγκες τους και τελικά, να είναι σε θέση να δημιουργήσουν προϊόντα και υπηρεσίες εστιασμένα στις ανάγκες και τις απαιτήσεις των πελατών τους (Kauffmann, et al., 2019). Από τα παραπάνω καταδεικνύεται η σημασία που έχει για τις επιχειρήσεις η αποτελεσματική ανάλυση των διαδικτυακών δεδομένων. Γνωστότερο ίσως παράδειγμα επιχείρησης που έχει ενσωματώσει επιτυχημένα τις πρακτικές αυτές στη λειτουργία της, είναι η Amazon (Εικόνα 3), ενώ στην ίδια κατεύθυνση κινούνται και διάφορες μεγάλες επιχειρήσεις όπως η Alibaba, η Expedia κλπ.



ΕΙΚΟΝΑ 3 Η ΔΙΕΠΑΦΗ ΣΤΟ ΗΛΕΚΤΡΟΝΙΚΟ ΚΑΤΑΣΤΗΜΑ ΤΗΣ AMAZON. ΦΑΙΝΕΤΑΙ Η ΑΞΙΟΛΟΓΗΣΗ ΠΟΥ ΑΠΟΡΡΕΙ ΑΠΟ ΑΛΛΟΥΣ ΧΡΗΣΤΕΣ ΚΑΘΩΣ ΚΑΙ ΟΙ ΠΡΟΤΑΣΕΙΣ ΠΡΟΪΟΝΤΩΝ ΠΟΥ ΕΝΔΙΑΦΕΡΟΥΝ ΤΟΝ ΕΠΙΣΚΕΠΤΗ ΜΕ ΒΑΣΗ ΤΗΝ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΠΟΥ ΥΛΟΠΟΙΕΙ ΤΟ ΛΟΓΙΣΜΙΚΟ ΤΗΣ AMAZON.

Η υιοθέτηση παρόμοιων προσεγγίσεων παρουσιάζεται και σε άλλες πλατφόρμες, όπως η eBay, η Google και οι Uber Technologies, με αποτέλεσμα να μεταβάλλεται ριζικά ο τρόπος προώθησης των προϊόντων αλλά και να γίνεται εφικτή η μείωση του κόστους συναλλαγής και παραγωγής (Rangaswamy, et al., 2020).

1.3 Εισαγωγή στην ανάλυση υποκειμενικής γνώμης (sentiment analysis)

Ένα μεγάλο κομμάτι από τις πνευματικές διεργασίες του Ανθρώπου καταγράφεται σε μορφή κειμένου. Το κείμενο αυτό άλλοτε καταγράφει αντικειμενικά κάποια πραγματικότητα, ενώ άλλοτε εκφράζει την υποκειμενική γνώμη, τη στάση ή τα συναισθήματα του δημιουργού του κειμένου, για κάποιο θέμα σε κάποια χρονική στιγμή.

Για παράδειγμα, στις 24 Φεβρουαρίου 2021 αναρτάται ένα ρεπορτάζ του J. Ronciano στην ιστοσελίδα του Forbes (Ronciano, 2021). Ο συγγραφέας του άρθρου σημειώνει μεταξύ άλλων:

1. Η μετοχή της Tesla έκλεισε την Τρίτη 23/2/2021 με απώλειες 2,19%, παρότι ενδοσυνεδριακά έφτασε να χάνει άνω του 10%, διευρύνοντας τις απώλειές της αναφορικά με την τιμή που είχε στις 8 Φεβρουαρίου, την ημέρα που, για πρώτη

φορά, ο Elon Mask αποκάλυψε ότι προχώρησε σε επένδυση ύψους 1,5 δισ. δολαρίων στο bitcoin. Το ποσό αυτό αντιστοιχούσε στο 20% των κεφαλαίων της αυτοκινητοβιομηχανίας του.

2. Όπως αναφέρει ο αναλυτής της Wedbush Νταν Άιβς "Ο κλάδος των ηλεκτρικών οχημάτων βρίσκεται στα πρώτα στάδια του παιχνιδιού... επισημαίνοντας ότι η Tesla κυριαρχεί στην αγορά, ... "Η αγορά θα έχει πολλούς νικητές σε ολόκληρο τον κόσμο - ειδικά καθώς το 'πράσινο παλιρροιακό κύμα' που τροφοδοτεί ο Αμερικανός πρόεδρος Τζο Μπάιντεν διαφαίνεται στον ορίζοντα, στις ΗΠΑ.

Στο εν λόγω κείμενο διακρίνει κανείς κάποιες πληροφορίες που έχουν αντικειμενικό χαρακτήρα, όπως είναι τα στοιχεία που αφορούν τις απώλειες που υπέστη η μετοχή της Tesla ή το ύψος του κεφαλαίου που επενδύθηκε σε bitcoin. Αντίθετα, στο δεύτερο τμήμα του κειμένου ο συγγραφέας του άρθρου καταγράφει την υποκειμενική άποψη του αναλυτή Νταν Άιβς χωρίς να είναι φανερό αν συμφωνεί με αυτήν ή όχι.

Είναι φανερό ότι ο άνθρωπος ως καταναλωτής, επενδυτής, ψηφοφόρος κλπ καλείται να πάρει αποφάσεις βασιζόμενος τόσο σε αντικειμενικά δεδομένα όσο και σε υποκειμενικές απόψεις και εκτιμήσεις άλλων ανθρώπων. Ο Liu επισημαίνει ότι η υποκειμενική άποψη ενός υποκειμένου εξαρτάται από ποικίλους παράγοντες όπως τα ενδιαφέροντα, το μορφωτικό επίπεδο, η ιδεολογία του γράφοντος, ο χρόνος στον οποίο εκφράστηκε η γνώμη αυτή κλπ (Liu, 2017). Όπως τονίζουν οι Hu & Liu "λόγω της ποικιλότητας των παραπάνω παραγόντων, είναι σημαντικό να εξετάζεται μια συλλογή από υποκειμενικές γνώμες παρά μια γνώμη από ένα συγκεκριμένο πρόσωπο με απώτερο σκοπό να δημιουργηθεί ένα είδος συγκερασμού των επιμέρους υποκειμενικών απόψεων" (Hu & Liu, 2004). Έτσι, είναι πιθανό να μας ενδιαφέρει η αξιοποίηση των μηνυμάτων χρηστών που εμφανίζονται σε κάποιο μέσο κοινωνικής δικτύωσης προκειμένου να εκτιμήσουμε τις πολιτικές τάσεις που επικρατούν στον πληθυσμό του δείγματος, να προσδιορίσουμε το βαθμό ικανοποίησης χρηστών σχετικά με προϊόντα και υπηρεσίες ή να αναλύσουμε οικονομικά και άλλα στοιχεία προκειμένου να προβλέψουμε την εξέλιξη κάποιων κλάδων της οικονομίας μέσα στο χρόνο (Yadollahi, et al., 2017).

Προβλήματα της κατηγορίας αυτής καλείται να αντιμετωπίσει η Ανάλυση Υποκειμενικής Γνώμης (Sentiment analysis). Όπως αναφέρουν διάφοροι ερευνητές του χώρου, «Συνήθως, η ανάλυση υποκειμενικής γνώμης οδηγεί στην απόφαση αν ένα κείμενο διάκειται θετικά,

αρνητικά ή ουδέτερα σχετικά με κάποιο θέμα. Δεν περιορίζεται όμως πάντα στην εύρεση της πολικότητας του εγγράφου αφού συχνά το ζητούμενο είναι να προσδιοριστούν τα συναισθήματα που κυριαρχούν στον δημιουργό του κειμένου». (Liu, 2017), (Mohammad, 2016).

Η ανάλυση της υποκειμενικής γνώμης, γνωστή και ως εξόρυξη γνώμης (opinion mining), χρησιμοποιεί ποικίλες τεχνικές όπως επεξεργασία φυσικής γλώσσας (Natural Language Processing, σύντομα NLP), νευρωνικά δίκτυα, στατιστική και άλλα, προκειμένου να αυτοματοποιηθεί η εξαγωγή και η ταξινόμηση της υποκειμενικής γνώμης που περιέχεται σε σύνολα κειμένων που εμπεριέχουν τη διατύπωση υποκειμενικής γνώμης υποκειμένων (Basant, et al., 2015) και να προσδιοριστεί η άποψη ή τα συναισθήματα που έχουν διαμορφώσει ομάδες διαδικτυακών χρηστών αναφορικά με προϊόντα, υπηρεσίες ή θέματα που αφορούν την κοινωνική, την οικονομική ή/και την πολιτική ζωή (Kang & Park, 2014).

Η αυτοματοποιημένη ανάλυση κειμένων με σκοπό την εξαγωγή της υποκειμενικής γνώμης, της στάσης και των συναισθημάτων που εμπεριέχουν τα κείμενα αυτά βοηθά, εκτός των άλλων, στην κατηγοριοποίηση των υποκειμένων και τη δημιουργία μοντέλων πρόβλεψης ενώ έχει μεγάλο πεδίο εφαρμογής σε πολλές περιπτώσεις, όπως:

- **Στη Δημόσια Υγεία.** Έχουν αναπτυχθεί εργαλεία για τη μηχανοποιημένη αποτίμηση διαδικτυακών αναρτήσεων σε μέσα κοινωνικής δικτύωσης με σκοπό τον αυτοματοποιημένο εντοπισμό ατόμων που παρουσιάζουν καταθλιπτικό προφίλ με αυτοκτονικές τάσεις (Rao, 2019), (Birjali, et al., 2017), (Cherry, et al., 2012), ή δείχνουν να είναι θύματα διαδικτυακού εκφοβισμού (cyber-bulling) (Nahar, et al., 2012), (Nahar, et al., 2014), (Dadvar, et al., 2014), (Kontostathis, et al., 2013). Τα εργαλεία αυτά, όπως καταλαβαίνει κανείς, καλούνται να ελέγξουν παραβατικές συμπεριφορές ή να εμποδίσουν τη διάπραξη εγκλημάτων. Σε μια εντελώς διαφορετική αξιοποίησή τους, τα εργαλεία της ανάλυσης υποκειμενικής γνώμης αξιοποιούνται για να προσδιοριστούν εκείνα τα χαρακτηριστικά που καθορίζουν τις υγιείς κοινωνίες (Gohil, et al., 2018), (Schwartz, et al., 2013), (Paul & Dredze, 2011).
- **Στην Πολιτική.** Στο χώρο της Πολιτικής διακρίνει κανείς ένα μεγάλο ενδιαφέρον των εμπλεκόμενων (πολιτικών, αναλυτών, συμβούλων κλπ) για την καταγραφή της υποκειμενικής γνώμης σε σχέση με τα Κοινά (Ratkiewicz, et al., 2011), (Metaxas & Mustafaraj, 2012) (Choi & Park, 2014), (Dunaway, 2013), (Greenwald, 2013). Αυτό είναι απόρροια της επιθυμίας των παραπάνω εμπλεκόμενων να έχουν εικόνα της

υποκειμενικής γνώμης και των τάσεων, όπως καταγράφονται στα μέσα κοινωνικής δικτύωσης, αναφορικά με εκλογές, κοινωνικά προβλήματα, εθνικά ή διεθνή ζητήματα (Haselmayer & Jenny, 2017), (Ansaria, et al., 2020), (Mohammad, et al., 2014). Αν αναλογιστεί κανείς ότι υπάρχουν ερευνητές που υποστηρίζουν ότι τα μέσα κοινωνικής δικτύωσης μπορούν να μεταβάλουν σημαντικά την πρόθεση ψήφου των ψηφοφόρων (Oliveira, et al., 2017) ή να βοηθήσουν στην πόλωση (Ott, 2017) τότε είναι αναμενόμενο το ενδιαφέρον που δείχνουν οι πολιτικοί και αναλυτές για τις υποκειμενικές απόψεις που εμφανίζονται στις πλατφόρμες κοινωνικής δικτύωσης.

- **Στη Διαχείριση του προφίλ της επιχείρησης (Brand management) και τη διαφήμιση.** Πολλοί ερευνητές εκφράζουν την άποψη ότι οι αγορές προϊόντων και υπηρεσιών που κάνει ένας καταναλωτής σήμερα βασίζονται, σε πολύ μεγάλο βαθμό, στην διαδικτυακή προβολή τους καθώς και στην άποψη που διατυπώνουν άλλοι καταναλωτές για το προϊόν αυτό (Guesalaga, 2016), (Ravi & Ravi, 2015), (Pathak & Pathak-Shelat, 2017). Παράλληλα, πολλές έρευνες αναδεικνύουν τη σημασία που έχει για μια επιχείρηση να είναι σε θέση να κατανοεί την άποψη που έχουν οι πελάτες της, αναλύοντας τις αναρτήσεις τους στα μέσα κοινωνικής δικτύωσης. Η ανάλυση αυτή είναι σε θέση να σκιαγραφήσει την άποψη που έχει ο καταναλωτής για τα προϊόντα αλλά και για το προφίλ, της επιχείρησης (Griesser & Gupta, 2019), (Yaakub & Jinglan-Zhang, 2013), (Pathak & Pathak-Shelat, 2017). Παράλληλα, δίνει την ευκαιρία στην επιχείρηση να εφαρμόσει ένα προωθημένο σύστημα διαχείρισης σχέσεων με τον πελάτη και να απαντά με άμεσο και αυτοματοποιημένο τρόπο στις ερωτήσεις και στα παράπονα των πελατών της (Caruano, et al., 2020), (Moghaddam, 2015), (Gupta, et al., 2012).
- **Στην εκπαίδευση.** Η εξ αποστάσεως εκπαίδευση έχει μπει στην καθημερινότητα όλων, ειδικά λόγω της πανδημίας του κορονοϊού και με βεβαιότητα μπορεί να πει κανείς ότι θα αποτελεί πλέον μια σημαντική συνιστώσα παροχής της εκπαίδευσης. Ένα από τα μειονεκτήματα που εμφανίζει η ψηφιακή εξ αποστάσεως εκπαίδευση (σύγχρονη και ασύγχρονη) είναι η αδυναμία του δάσκαλου ή του εκπαιδευτή να έχει επίγνωση των συναισθημάτων που κυριαρχούν στην εικονική του τάξη. Η κατάσταση γίνεται δυσμενέστερη στις περιπτώσεις που η εκπαίδευση παρέχεται, μέσω της προσέγγισης των διαδικτυακών προγραμμάτων ασύγχρονης εκπαίδευσης (Massive Open Online Courses ή MOOCs) που προσφέρονται ελεύθερα σε μεγάλο αριθμό

συμμετεχόντων. Συχνά, οι συμμετέχοντες σε ένα τέτοιο event ανέρχονται σε δεκάδες χιλιάδες. Πολλοί ερευνητές ισχυρίζονται ότι η αξιοποίηση της ανάλυσης υποκειμενικής γνώμης στις περιπτώσεις αυτές μπορεί να εξασφαλίσει τον αυτόματο εντοπισμό των συναισθημάτων των συμμετεχόντων αναφορικά με την ποιότητα της παρεχόμενης εκπαίδευσης ή επιμόρφωσης (Barron-Estrada, et al., 2019), (Oramas, et al., 2018), (Mac Kim & Calvo, 2010), (Suero-Montero & Suhonen, 2014).

- **Στη διαχείριση καταστροφών και καταστάσεων έκτακτης ανάγκης.** Τα μέσα κοινωνικής δικτύωσης επιτρέπουν την άμεση επικοινωνία μεταξύ των αρμόδιων φορέων και των ενδιαφερομένων για την αντιμετώπιση εκτάκτων αναγκών. Ερευνητές που έχουν μελετήσει τη συμπεριφορά και τα συναισθήματα των πολιτών κατά τη διάρκεια καταστροφών και έκτακτων καιρικών φαινομένων, μέσω της ανάλυσης των αναρτήσεων στο Twitter, πιστεύουν ότι η ανάλυση των συναισθημάτων των πληγέντων από μια καταστροφή μας επιτρέπει να αποτιμήσουμε γρήγορα τη σοβαρότητα της κατάστασης, να αντιδράσουμε με αποτελεσματικότητα και να προσφέρουμε βοήθεια και ανακούφιση στα θύματα της καταστροφής (Sakaki, et al., 2010), (Carley, et al., 2016), (Beigi, et al., 2016), (Nepalli, et al., 2017), (Ragini, et al., 2018).

Στις εικόνες που ακολουθούν (Εικόνες 4 και 5) φαίνεται η αξιοποίηση των μέσων κοινωνικής δικτύωσης για την αντιμετώπιση των συνεπειών του ισχυρού σεισμού που έλαβε χώρα στις 13 Φεβρουαρίου 2021 στις ακτές της Φουκουσίμα στην Ιαπωνία.

The screenshot shows a webpage with the following content:

- Navigation: Home > Policy > Cabinet Decisions and Other Announcements > 2021 > Instructions from the Prime Minister Regarding the Earthquake off coast of Fukushima Prefecture (23:09)
- Section Header: Cabinet Decisions and Other Announcements
- Title: Instructions from the Prime Minister Regarding the Earthquake off coast of Fukushima Prefecture (23:09)
- Date: February 13, 2021
- Related Link: Japanese
- Buttons: Tweet, Share
- List of instructions:
 - 1) Identify the damages as soon as possible;
 - 2) Coordinate closely with local governments and make utmost effort in emergency disaster response including rescuing affected people, putting human lives first and working as one; and
 - 3) Provide timely and accurate information to the public such as evacuation and damages.

ΕΙΚΟΝΑ 4 ΔΗΛΩΣΕΙΣ ΤΟΥ ΙΑΠΩΝΑ ΠΡΩΘΥΠΟΥΡΓΟΥ ΑΠΟ ΤΑ ΜΕΣΑ ΚΟΙΝΩΝΙΚΗΣ ΔΙΚΤΥΩΣΗΣ ΜΕ ΑΦΟΡΜΗ ΤΟΝ ΣΕΙΣΜΟ ΤΗΣ 13 ΦΕΒΡΟΥΑΡΙΟΥ 2021 ΣΤΗΝ ΠΕΡΙΟΧΗ ΤΗΣ FUKUSHIMA.



ΕΙΚΟΝΑ 5 ΑΝΑΡΤΗΣΗ ΤΩΝ THE JAPAN TIMES ΑΝΑΦΟΡΙΚΑ ΜΕ ΤΟ ΣΕΙΣΜΟ ΤΩΝ 7,3 ΡΙΧΤΕΡ ΣΤΗ ΠΕΡΙΟΧΗ ΤΗΣ FUKUSHIMA ΣΤΙΣ 13 ΦΕΒΡΟΥΑΡΙΟΥ 2021.

- **Στην χρηματο-οικονομική.** Από το 1993, με τη δημοσίευση του πρώτου άρθρου που έθετε τον προβληματισμό αναφορικά με την επίδραση που έχουν οι ειδήσεις πάνω στη μεταβλητότητα των χρηματιστηριακών προϊόντων (Engle & Ng, 1993), υπάρχει ζωηρό ενδιαφέρον στην επιστημονική κοινότητα αναφορικά με την αξιοποίηση της ανάλυσης υποκειμενικής γνώμης στο χώρο των Χρηματοπιστωτικών αγορών. Τα κείμενα που εμφανίζονται στο διαδίκτυο και αφορούν αντικειμενικά δεδομένα ή υποκειμενικές απόψεις σχετικά με την πορεία της Οικονομίας ή του Χρηματιστηρίου φαίνεται ότι επηρεάζουν, σε μεγάλο βαθμό, τόσο τη διάθεση όσο και τις αποφάσεις των υποκειμένων. Όπως αναφέρουν οι Ling et al. «η ανάλυση υποκειμενικής γνώμης καθώς και ο προσδιορισμός των συναισθημάτων από κείμενα που περιλαμβάνουν ειδήσεις μπορεί να βοηθήσει στην μοντελοποίηση και πρόβλεψη που αφορά τις τιμές, τη μεταβλητότητα, τον όγκο των συναλλαγών των χρηματοπιστωτικών προϊόντων αλλά και των πιθανών κινδύνων, αναφορικά με επενδυτικές κινήσεις» (Luo, et al., 2018). Αρκετοί ερευνητές επισημαίνουν ότι η ανάλυση της υποκειμενικής γνώμης, στην περίπτωση αυτή, διαφέρει σημαντικά από την ανάλυση που διενεργείται σε γενικές κριτικές πελατών για καθημερινά προϊόντα και υπηρεσίες (Liu, 2012), (Ma, et al., 2017) ενώ επιπρόσθετα, τα αναλυόμενα κείμενα που αφορούν χρηματο-οικονομικά θέματα παρουσιάζουν περίπλοκη και δυσνόητη γραφή (Luo, et al., 2018). Τα συνήθη ερευνητικά ερωτήματα που τίθενται, στην περίπτωση αυτή,

αφορούν την επίδραση που έχουν οι ειδήσεις στην διαμόρφωση της αξίας των μετοχών (Kazemian, et al., 2016), του όγκου των συναλλαγών (Engelberg & Parsons, 2011), στη μεταβλητότητα (Rekabsaz, et al., 2017), καθώς και στον προσδιορισμό των πιθανών κινδύνων (Norpp & Hanbury, 2015).

- **Στην κοινωνιολογία και την ψυχολογία.** Η ανάλυση υποκειμενικής γνώμης δεδομένων που εμπεριέχονται στα μέσα κοινωνικής δικτύωσης οδηγεί σε ενδιαφέροντα συμπεράσματα τόσο για προβλήματα που αφορούν την ψυχολογία όσο και την κοινωνιολογία. Διάφοροι ερευνητές ισχυρίζονται ότι η επικοινωνία μέσω του Twitter διευκολύνει τόσο την πολιτική πόλωση αναφορικά με κοινωνικά προβλήματα (Ott, 2017), (Morales, et al., 2015), (Hong & Kim, 2016) όσο και την επικοινωνία μεταξύ ατόμων που έχουν ίδιες αντιλήψεις (Merry, 2016) με αποτέλεσμα να δημιουργούνται εικονικοί τόποι συνάντησης με αυξημένη ηχώ πολωμένων απόψεων (Hong & Kim, 2016). Παράλληλα, υπάρχουν έρευνες που αναδεικνύουν τη σημαντικότητα της ανάλυσης υποκειμενικής γνώμης για τον προσδιορισμό των χαρακτηριστικών της προσωπικότητας των υποκειμένων και των συναισθημάτων που βιώνουν, με βάση το περιεχόμενο των αναρτήσεων τους καθώς και την κατανόηση των διαφορών ανάμεσα στα φύλα, μέσω της ανάλυσης της γλώσσας που χρησιμοποιούν στις αναρτήσεις τους οι άνδρες και οι γυναίκες (Joseph, et al., 2016), (Grijalva, et al., 2015), (Minamikawa & Yokohama, 2011).
- **Στη Λογοτεχνία.** Η λογοτεχνία είναι ένας χώρος όπου τα κείμενα βρίθουν από ήρωες οι οποίοι έχουν κάποια προσωπικότητα, θέτουν στόχους ως απόρροια των φιλοδοξιών ή των επιθυμιών που έχουν και κυριαρχούνται από πάθη. Όπως είναι αναμενόμενο, η ανάλυση υποκειμενικής γνώμης έχει χρησιμοποιηθεί από διάφορους ερευνητές για να δοθεί απάντηση σε ποικίλα ερευνητικά ερωτήματα. Ο συνηθέστερος τρόπος αξιοποίησης της ανάλυσης υποκειμενικής γνώμης στη λογοτεχνία αφορά την κατηγοριοποίηση των κειμένων μιας συλλογής με βάση κάποια κριτήρια όπως τα κυρίαρχα συναισθήματα που υποβόσκουν στο κείμενο (Liu, 2015), (Yu, 2008). Παράλληλα, ερευνητές έχουν επιχειρήσει να δημιουργήσουν αυτοματοποιημένους τρόπους κατάταξης κειμένων με βάση:
 - Το είδος του τέλους σε σχέση με τους πρωταγωνιστές (Zehe, et al., 2016).
 - Τη δομή του κειμένου (Reagan, et al., 2016).

- Τη γεωγραφική κατανομή των συναισθημάτων, όπως αυτή καταγράφεται σε κείμενα (Heuser, et al., 2016).
- Τη σκιαγράφιση του χαρακτήρα και των συναισθημάτων του ήρωα ή της ηρωίδας κάποιου λογοτεχνικού έργου (Jacobs, 2019), (Jhavar & Mirza, 2018), (Kim & Klinger, 2018), (Kim & Klinger, 2019).

Παράλληλα, δεν λείπουν έρευνες που έχουν ως στόχο:

- Να εντοπισθούν οι ωραιότεροι στίχοι από το έργο του Σαίξπηρ (Jacobs, et al., 2017).
- Να παραχθεί μουσική με βάση τα συναισθήματα που κυριαρχούν σε ένα κείμενο (Davis & Mohammad, 2014).

Από τα παραπάνω γίνεται φανερό ότι το επιστημονικό πεδίο της Αναλυτικής, γενικότερα, και της ανάλυσης υποκειμενικής γνώμης είναι ένα πολλά υποσχόμενο πεδίο που μπορεί να προσφέρει εργαλεία ανάλυσης προβλημάτων σε πολλούς χώρους της ανθρώπινης δραστηριότητας.

1.4 Βασικές έννοιες στην ανάλυση υποκειμενικής γνώμης

Αν επισκεφθεί κανείς τα μέσα κοινωνικής δικτύωσης θα παρατηρήσει ότι οι αναρτήσεις αποτελούνται αποκλειστικά από κείμενα, βίντεο, εικόνες ή αρχεία ήχου τα οποία αποτυπώνουν την άποψη κάποιου για ένα θέμα. Σημαντική υποκατηγορία εικόνων είναι τα [emojis](#). Αυτά είναι εικόνες που εισάγει ο χρήστης στα κείμενα του προκειμένου να εκφράσει συναισθήματα (Εικόνα 6).



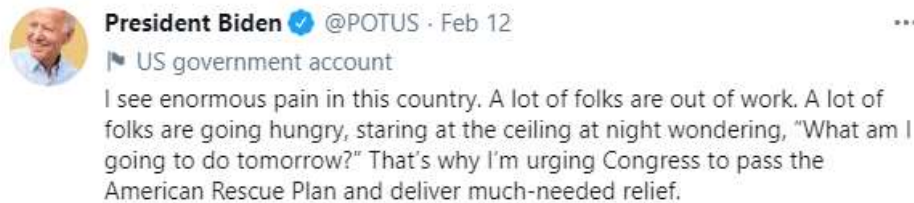
ΕΙΚΟΝΑ 6 ΣΥΝΗΘΙΣΜΕΝΑ ΕΜΟJIS

Είναι φανερό ότι η ανάλυση δεδομένων με τόσο μεγάλο βαθμό ανομοιογένειας εμφανίζει πολυποίκιλα προβλήματα. Η παρούσα Θέση έχει ως στόχο την ανάλυση αμιγούς κειμένου.

Όπως έχει ήδη αναφερθεί, ένα κείμενο (παραδοσιακό ή ψηφιακό) ενδέχεται να αποτυπώνει την υποκειμενική άποψη ενός ή περισσότερων υποκειμένων. Συνηθέστερη περίπτωση τέτοιου κειμένου αποτελεί το κείμενο ενός tweet κάποιου χρήστη της πλατφόρμας Twitter. Όπως τονίζει ο Liu, η υποκειμενική γνώμη αυτή έχει τα εξής βασικά χαρακτηριστικά γνωρίσματα:

- Τον κτήτορα της γνώμης.
- Τη χρονική στιγμή που εκφράζεται η γνώμη αυτή.
- Την πολικότητα που εκφράζει ή τα συναισθήματα που εμπεριέχει.

Ας χρησιμοποιήσουμε ως παράδειγμα την ανάρτηση του προέδρου των ΗΠΑ που περιέχεται στην Εικόνα 7.



ΕΙΚΟΝΑ 7 ΑΝΑΡΤΗΣΗ ΤΟΥ ΠΡΟΕΔΡΟΥ ΒΙΔΕΝ ΣΤΟ ΜΕΣΟ ΚΟΙΝΩΝΙΚΗΣ ΔΙΚΤΥΩΣΗΣ TWITTER.

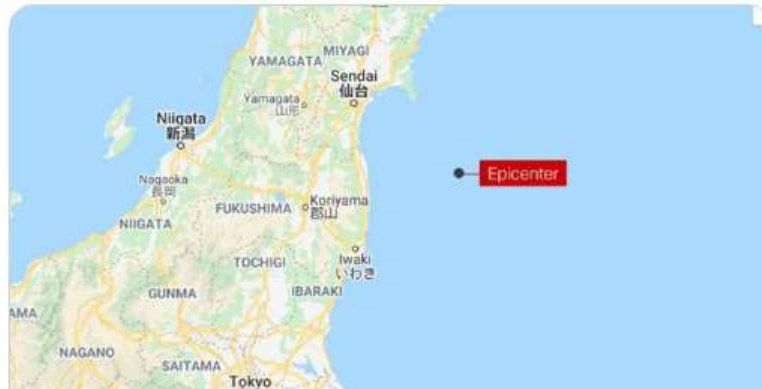
Ιδιοκτήτης του εν λόγω κειμένου είναι ο Joe Biden. Κατά συνέπεια, το κείμενο θεωρείται ότι εκφράζει την υποκειμενική γνώμη και τα συναισθήματά του. Η ανάρτηση του κειμένου έγινε στις 12 Φεβρουαρίου 2021. Στο περιεχόμενο του κειμένου ο πρόεδρος Biden φαίνεται να αποδέχεται την ύπαρξη αρνητικού κλίματος στην οικονομική κατάσταση των ΗΠΑ.

Μια ανάρτηση σε μέσο κοινωνικής δικτύωσης δεν αφορά πάντα υποκειμενική γνώμη. Για παράδειγμα, στην Εικόνα 8, που ακολουθεί, εμφανίζεται μια ανάρτηση του CNN που πληροφορεί το κοινό για ένα γεγονός. Συγκεκριμένα, αφορά τον σεισμό που έπληξε τις ανατολικές ακτές της Ιαπωνίας στις 13 Φεβρουαρίου 2021. Πέρα από το γεγονός ότι το κείμενο περιγράφει ένα αδιαμφισβήτητο γεγονός δεν μας επιτρέπει να αντιληφθούμε αν ο συντάκτης του κειμένου έχει θετική, αρνητική ή αδιάφορη θέση πάνω στο γεγονός.



CNN Breaking News @cnnbrk · 2h

A 7.1 magnitude earthquake struck off the coast of eastern **Japan** late Saturday evening but there were no immediate reports of casualties or damage



ΕΙΚΟΝΑ 8 ΑΝΑΡΤΗΣΗ ΤΟΥ CNN ΑΝΑΦΟΡΙΚΑ ΜΕ ΤΟ ΣΕΙΣΜΟ ΤΗΣ 12^{ΗΣ} ΦΕΒΡΟΥΑΡΙΟΥ 2021 ΣΤΗΝ ΙΑΠΩΝΙΑ.

Από το γεγονός ότι η υποκειμενική γνώμη έχει ιδιοκτήτη απορρέει ότι η υποκειμενική γνώμη μεταξύ διαφορετικών ιδιοκτητών είναι δυνατό να διαφέρει σε σημαντικότητα. Για παράδειγμα, στις Εικόνες 9 και 10 που ακολουθούν, εμφανίζονται οι αναρτήσεις ενός απλού πολίτη και του Ιάπωνα πρωθυπουργού αντίστοιχα για τον σεισμό της 12^{ης} Φεβρουαρίου 2021.



TAERIRAHMAWATI C* @BooNon_98 · 2h

tw // earthquake

oh no~ this is really bad news
A magnitude 7.1 earthquake has occurred in **Japan** at night.

to j-carats, please stay safe!!

ΕΙΚΟΝΑ 9 ΑΝΑΡΤΗΣΗ ΠΟΛΙΤΗ ΑΝΑΦΟΡΙΚΑ ΜΕ ΤΟ ΣΕΙΣΜΟ ΤΗΣ 12^{ΗΣ} ΦΕΒΡ. 2021 ΣΤΗΝ ΙΑΠΩΝΙΑ



PM's Office of Japan @JPN_PMO

Instructions from the Prime Minister Regarding the Earthquake off coast of Fukushima Prefecture (23:09)
japan.kantei.go.jp/99_suga/decisi...
(February 13)



Instructions from the Prime Minister Regarding the Earthquak...
Cabinet Decisions / Direction - Prime Minister of Japan and His Cabinet
japan.kantei.go.jp

6:20 μ.μ. · 13 Φεβ 2021 · Twitter Web App

ΕΙΚΟΝΑ 10 ΑΝΑΡΤΗΣΗ ΤΟΥ ΙΑΠΩΝΑ ΠΡΩΘΥΠΟΥΡΓΟΥ ΓΙΑ ΤΟ ΣΕΙΣΜΟ ΤΗΣ 12^{ΗΣ} ΦΕΒΡ. 2021 ΣΤΗΝ ΙΑΠΩΝΙΑ

1.5 Ταξινόμια της ανάλυσης υποκειμενικής γνώμης (sentiment analysis)

Η ανάλυση υποκειμενικής γνώμης χρησιμοποιεί διαδικασίες ανάλυσης φυσικής γλώσσας (Natural Language Processing, σύντομα NLP) προκειμένου να αυτοματοποιηθεί η αποτύπωση, ή η κατηγοριοποίηση της υποκειμενικής γνώμης που προέρχεται από ανασκοπήσεις κειμένων που περιέχουν την έκφραση υποκειμενικής γνώμης διαφόρων ανθρώπων (Basant, et al., 2015)

Σύμφωνα με πολλούς ερευνητές (Liu, 2015), (Liu, 2012) (Yadollahi, et al., 2017), (Rana & Cheah, 2016), η ανάλυση υποκειμενικής γνώμης (ή sentiment analysis) μέσω αρχείων κειμένου διακρίνεται στα παρακάτω επίπεδα ανάλυσης:

- 1) **Επίπεδο εγγράφου (document level)**. Στην περίπτωση αυτή, το ζητούμενο είναι να εξαχθούν από όλο το κείμενο οι λέξεις που συνιστούν έκφραση γνώμης ή συναισθήματος και να υπολογιστεί η πολικότητα (polarity) ή το συναισθηματικό προφίλ όλου του κειμένου. Αυτό το είδος της ανάλυσης αξιοποιείται σε κοινωνικές και ψυχολογικές μελέτες που αφορούν κοινωνικά δίκτυα, έρευνες αγοράς που στοχεύουν στην αποτίμηση της ικανοποίησης πελατών καθώς και σε ανάλυση συναισθημάτων ασθενών κ.α. (Yadollahi et al., 2017).
- 2) **Επίπεδο πρότασης (sentence level)**. Στην περίπτωση αυτή, η ανάλυση στοχεύει στον προσδιορισμό της πολικότητας ή συναισθήματος ανά πρόταση. Κατά την ανάλυση αυτή είναι σημαντικό να καθοριστεί από την αρχή αν η πρόταση-στόχος είναι αντικειμενική ή υποκειμενική. Αυτό το είδος ανάλυσης χρησιμοποιείται πολύ σε εφαρμογές που έχουν να κάνουν με αποτίμηση τιτιβισμών (tweets), αναρτήσεις του Facebook, σχόλια, μικρά μηνύματα κλπ.
- 3) **Επίπεδο οντότητας (entity / aspect level)**. Αφορά τον προσδιορισμό της γνώμης που έχουν άτομα για κάποιο συγκεκριμένο προϊόν ή χαρακτηριστικό προϊόντος, ή γενικότερα μιας οντότητας (Talafha, et al., 2019), (Yadollahi, et al., 2017), (Rana & Cheah, 2016), (Kiritchenko, et al., 2014). Προκειμένου να διενεργηθεί ανάλυση συναισθήματος σε επίπεδο χαρακτηριστικού πρέπει να γίνει πρώτα η απομόνωση όλων των οντοτήτων (entities) και των αντίστοιχων χαρακτηριστικών τους (aspects/features). Στη συνέχεια, το ζητούμενο είναι να προσδιοριστεί η πολικότητα των απόψεων της εξεταζόμενης ομάδας υποκειμένων για κάθε χαρακτηριστικό και να δημιουργηθεί μια σύνοψη των εν λόγω

απόψεων (Rana & Cheah, 2016) οι οποίες, όπως αναφέρουν οι Hu & Liu (2004) μπορούν να εκφράζονται στο κείμενο είτε ρητά είτε έμμεσα μέσω σχετικών επιθέτων ή επιρρημάτων (Hu & Liu, 2004).

Η παρούσα Θέση εστιάζει στην ανάλυση υποκειμενικής γνώμης σε επίπεδο εγγράφου (Document sentiment classification/ document-level sentiment analysis). Η ανάλυση αυτή αποτελεί και την πιο παλαιά και διαδεδομένη μορφή ανάλυσης κειμένου με σκοπό την κατηγοριοποίηση του και την εξαγωγή υποκειμενικής γνώμης.

Το υπό ανάλυση αντικείμενο είναι ένα έγγραφο που θεωρείται ότι εμπεριέχει κάποια υποκειμενική γνώμη. Μέσω αλγορίθμων επιζητείται η κατηγοριοποίηση του με βάση δύο προσανατολισμούς ή πολικότητες (orientations/polarities). Αυτές είναι η θετική και η αρνητική πολικότητα (ή υποκειμενική γνώμη) (βλέπε Pang and Lee, 2008; Liu, 2012). Όπως αναφέρει για το θέμα ο Liu «στην ανάλυση υποκειμενικής γνώμης σε επίπεδο εγγράφου κάθε έγγραφο θεωρείται ως μια οντότητα και δεν εξετάζονται επιμέρους απόψεις (aspects) που, τυχόν, να εμπεριέχονται στο έγγραφο» (Liu, 2015). Κατά τη διάρκεια αυτής της ανάλυσης μπορούν να αξιοποιηθούν αλγόριθμοι εποπτευόμενης μάθησης ή μη εποπτευόμενης μάθησης.

Η ανάλυση και κατηγοριοποίηση σε επίπεδο εγγράφου προϋποθέτει την ισχύ των παρακάτω υποθέσεων (Liu, 2015):

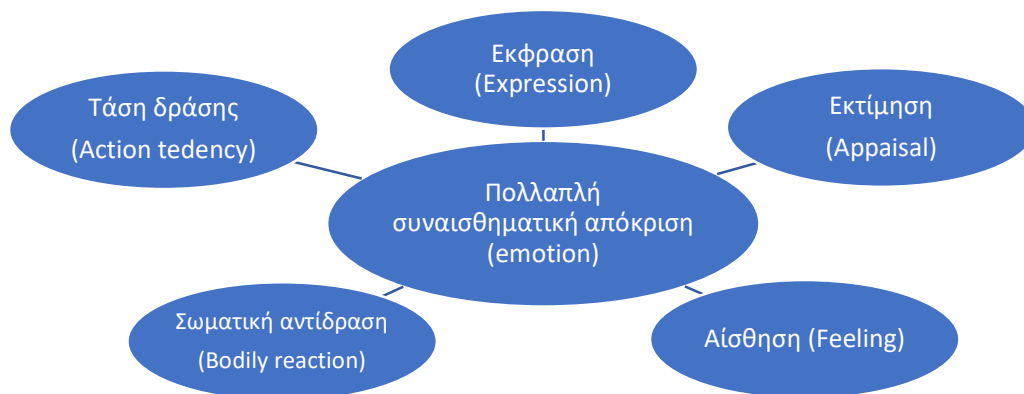
- Το υπό εξέταση έγγραφο εκφράζει άποψη πάνω σε μία γενική οντότητα (entity) και αφορά ένα μόνο κτήτορα απόφασης.
- Η χρονική στιγμή στην οποία εκφράζεται η υποκειμενική γνώμη θεωρείται δεδομένη ή αδιάφορη.

Ανάλογα με το εξαγόμενο από την ανάλυση, αν γίνεται εκτίμηση της πολικότητας του εγγράφου μέσω μιας δίτιμης λογικής μεταβλητής (θετικό/ αρνητικό) τότε το πρόβλημα μπορεί να θεωρηθεί ότι εμπίπτει στην κατηγορία των προβλημάτων κατηγοριοποίησης. Αντίθετα, όταν το εξαγόμενο της ανάλυσης είναι μια μεταβλητή κλίμακας διάταξης τότε μπορεί να θεωρηθεί ότι ανήκει στα προβλήματα παλινδρόμησης.

1.6 Κατηγοριοποίηση εγγράφου σε σχέση με το Συναίσθημα (Emotion Classification of Documents)

Διάφοροι ερευνητές ισχυρίζονται ότι τα άτομα που προσχωρούν σε μεγάλες ομάδες χάνουν συχνά την ικανότητα να σκέπτονται και να ενεργούν λογικά με αποτέλεσμα να πράττουν μέσα από την κυριαρχία του συναισθήματος και της 'λογικής' του πλήθους (Le Bon, 1895), (McClelland, 1989), (Barbalet, 2004). Στη σημερινή εποχή παρατηρείται οι άνθρωποι να προσχωρούν σε πολυπληθείς εικονικές κοινότητες, μέσω των πλατφορμών κοινωνικής δικτύωσης. Τα μέλη των κοινοτήτων αυτών μέσω μηνυμάτων κειμένου ή πολυμέσων υφίστανται κοινά ερεθίσματα και αναπτύσσουν παρόμοια αντανακλαστικά. Ως εκ τούτου, έχει ενδιαφέρον η έρευνα των συναισθημάτων και των απόψεων από τα οποία διακατέχονται τα μέλη των «κοινωνιών» αυτών. Η εν λόγω Θέση προτείνει ένα μεθοδολογικό πλαίσιο το οποίο είναι σε θέση να αναλύει το περιεχόμενο κειμένων και να προσδιορίζει τα συναισθήματα που καλλιεργούνται μέσω των κειμένων αυτών. Το μεθοδολογικό πλαίσιο αυτό αξιοποιεί μοντέλα μηχανικής μάθησης και τεχνικές εξαγωγής και ανάλυσης συναισθήματος. Όπως αναγνωρίζεται από πολλούς ερευνητές, η κατηγοριοποίηση των εγγράφων μέσω των συναισθημάτων ή της διάθεσης που υποκρύπτονται σε αυτά είναι ιδιαίτερα δύσκολη. Αυτό οφείλεται βασικά στα εξής (Liu, 2015), (Alm, 2008):

- Υπάρχουν πολλές κατηγορίες και τύποι συναισθημάτων και διαθέσεων (Εικόνα 11).
- Εμφανίζονται σημαντικές ομοιότητες μεταξύ των διάφορων συναισθημάτων και των διαθέσεων με αποτέλεσμα να μην είναι εύκολη η διάκριση του ενός από το άλλο.
- Δεν είναι εύκολη η καταγραφή και διάκριση των συναισθημάτων μέσω κειμένου.



ΕΙΚΟΝΑ 11 ΚΑΤΗΓΟΡΙΕΣ ΣΥΝΑΙΣΘΗΜΑΤΩΝ

Στο σημείο αυτό αξίζει να σημειωθούν οι διάφορες συναισθηματικές έννοιες. Σύμφωνα με το λεξικό Merriam-Webster (Merriam & Webster, 2021):

- **Θυμικό** (affect): Αφορά την συνειδητή και υποκειμενική πτυχή ενός συναισθήματος που δεν σχετίζεται με σωματικές αλλαγές. Κατ' άλλους ορίζεται σαν μια πρωτογενής νευροφυσιολογική κατάσταση η οποία δεν έχει ως στόχο κάποιο αντικείμενο ή υποκείμενο (Russell, 2003)
- **Συναίσθημα** (emotion): Το συναίσθημα ορίζεται ως μια συνειδητή νοητική αντίδραση που κατευθύνεται προς ένα αντικείμενο και συνοδεύεται από φυσιολογικές και συμπεριφορικές αλλαγές στο σώμα του υποκειμένου.
- **Διάθεση** (Mood): Η διάθεση αφορά ένα υποβόσκον συναίσθημα, μια διανοητική κατάσταση πριν από μια πράξη.
- **Αίσθημα** (Feeling): Αφορά τις βασικές φυσικές αισθήσεις με τις οποίες το σώμα αντιλαμβάνεται την επαφή με το δέρμα, τη θερμοκρασία κλπ.

Όπως επισημαίνει ο Liu (2017), η παραπάνω ταξινόμηση δημιουργεί συχνά σύγχυση στον μελετητή και ορθώνει δυσκολίες στην υλοποίηση της ανάλυσης συναισθήματος (Liu, 2017).

Βασική προϋπόθεση για την ανάλυση συναισθήματος αποτελεί η αποδοχή ενός θεωρητικού πλαισίου το οποίο καθορίζει τα είδη των συναισθημάτων. Οι βασικές θεωρίες που συναντώνται είναι οι κάτωθι:

- Το μοντέλο του **Paul Ekman** (Ekman, 1999) (Εικόνα 12) θεωρεί ότι υπάρχουν έξι βασικές κατηγορίες συναισθημάτων που προέρχονται από διακριτά νευρικά ερεθίσματα και, ως εκ τούτου, τα συναισθήματα είναι ανεξάρτητα. Τα βασικά αυτά συναισθήματα είναι:
 - Ευτυχία (happiness).
 - Δυστυχία (sadness).
 - Οργή (anger).
 - Απέχθεια (disgust).
 - Έκπληξη (surprise)
 - Φόβος (fear)

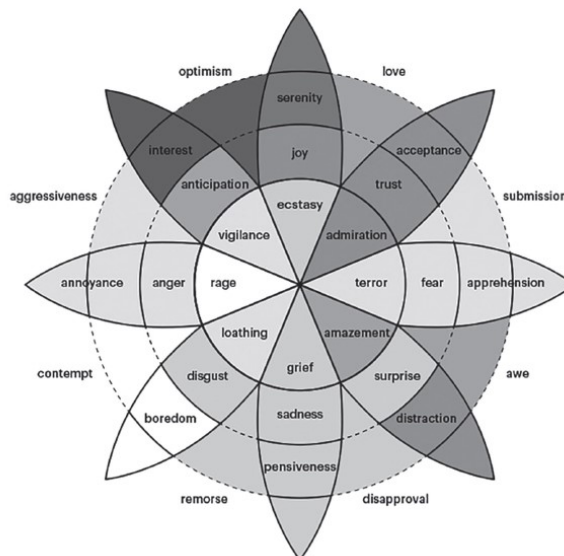
Συνδυασμοί αυτών των συναισθημάτων δημιουργούν σύνθετα συναισθήματα όπως είναι η ενοχή, η υπερηφάνεια κλπ.



ΕΙΚΟΝΑ 12 ΤΑ ΒΑΣΙΚΑ ΣΥΝΑΙΣΘΗΜΑΤΑ ΚΑΤΑ ΕΚΜΑΝ. Η ΕΙΚΟΝΑ ΕΧΕΙ ΛΗΦΘΕΙ ΑΠΟ ΤΟ RESEARCH-GATE

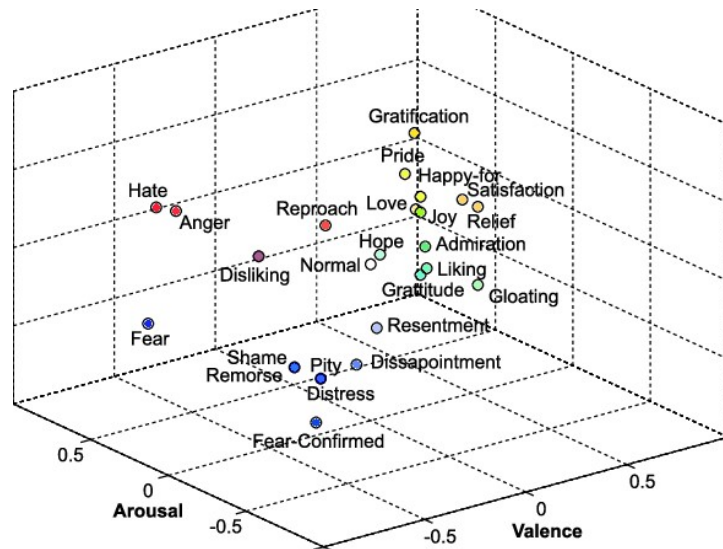
• Το μοντέλο του **Robert Plutchik** (Plutchik, 1980), (Εικόνα 13) υποθέτει ότι υπάρχουν οκτώ βασικά συναισθήματα τα οποία εμφανίζονται ανά ζεύγη και οι συνδυασμοί τους δημιουργούν σύνθετα συναισθήματα. Τα βασικά, κατά Plutchnik, συναισθήματα είναι τα έξι βασικά συναισθήματα του Ekman προσαυξημένα με τα συναισθήματα Αποδοχή (acceptance)/εμπιστοσύνη (trust) και προσμονή (anticipation). Κατά συνέπεια, τα ζεύγη των βασικών συναισθημάτων είναι:

- Ευτυχία/ Δυστυχία.
- Εμπιστοσύνη/ Απέχθεια.
- Οργή/ Φόβος.
- Έκπληξη/ Προσμονή



ΕΙΚΟΝΑ 13 Ο ΚΥΚΛΟΣ ΤΩΝ ΣΥΝΑΙΣΘΗΜΑΤΩΝ ΤΟΥ ΠΛΟΥΤΣΝΙΚ

- Το μοντέλο των **Ortony, Clore, and Collins (OCC)** (Ortony, et al., 1990) (Εικόνα 14) υποστηρίζει ότι η αναλογία των βασικών συναισθημάτων δεν είναι αυτή που εμφανίζεται από τον Ekman και τον Plutchik. Δέχονται 22 βασικά συναισθήματα που καλύπτουν μια πολύ ευρύτερη αναπαράσταση συναισθημάτων όπως ανακούφισης, φθόνου, μομφής, εκτίμησης, ντροπής, οίκτου, απογοήτευσης, θαυμασμού, ελπίδας, θλίψης, ικανοποίησης, γοητείας, συμπάθειας και αντιπάθειας.



ΕΙΚΟΝΑ 14 ΤΑ ΒΑΣΙΚΑ ΣΥΝΑΙΣΘΗΜΑΤΑ ΚΑΤΑ OCC

Σε επίπεδο ανάλυσης συναισθήματος κειμένου, οι βασικές προσεγγίσεις μέσω των οποίων γίνεται η ανάλυση συναισθήματος από κείμενα είναι οι παρακάτω:

- **Προσέγγιση δημιουργίας κανόνων (RC).** Η προσέγγιση αυτή βασίζεται στη δημιουργία γραμματικών και λογικών κανόνων που πρέπει να ακολουθηθούν για να γίνει ο προσδιορισμός των συναισθημάτων (Acheampong, et al., 2020). Η προσέγγιση απαιτεί την αναγνώριση λέξεων-κλειδιών (key recognition KR) και μεθόδους λεξιλογικής συγγένειας που παρέχεται από λεξικά συναισθήματος. Όπως αναφέρουν οι Acheampong et al. (2020), τα πλέον γνωστά είναι:
 - Το WordNet-Affect (Strapparava & Valitutti, 2004).
 - Το EmoSenticNet (Poria, et al., 2014).
 - Το DepecheMood (Staiano, et al., 2014).
 - Το SentiWordNet (Esuli & Sebastiani, 2014).

- ο και το λεξικό του Εθνικού Συμβουλίου Έρευνας του Καναδά (NRC) (Mohammad & Turney, 2010)

Αυτά τα λεξικά συναισθημάτων περιέχουν λέξεις αναζήτησης συναισθημάτων ή λέξεις-κλειδιά όπως χαρούμενος/θυμωμένος, αγάπη/μίσος, έκπληξη/απογοήτευση κ.λπ. Το ζητούμενο στις περιπτώσεις αυτές είναι να βρεθεί σε ένα κείμενο το πλήθος ή το ποσοστό εμφανίσεων των λέξεων αυτών. Συνήθως, η προσέγγιση αυτή ενισχύεται με την μέθοδο λεξικής συγγένειας (LA) (Acheampong, et al., 2020), η οποία σε δεύτερο επίπεδο αναθέτει πιθανολογικές συγγένειες στις τυχαίες λέξεις συναισθημάτων.

- **Η προσέγγιση μηχανικής μάθησης (ML).** Στην προσέγγιση αυτή γίνεται εκτεταμένη χρήση αλγόριθμων μηχανικής μάθησης για να κατηγοριοποιηθεί ένα κείμενο σε διάφορες κατηγορίες συναισθημάτων. Συχνά, οι αλγόριθμοι εποπτευόμενης μάθησης δείχνουν να υπερέχουν σε πολλά σημεία από τους αντίστοιχους της μη εποπτευόμενης μάθησης
- **Υβριδική προσέγγιση (HA).** Η προσέγγιση αυτή συνδυάζει αφενός την προσέγγιση δημιουργίας κανόνων (RC) και αφετέρου την μηχανική μάθηση (ML). Εκμεταλλευόμενη τα πλεονεκτήματα καθεμίας από τις προσεγγίσεις μειώνει τις επιδράσεις των μειονεκτημάτων τους. Έχει διαπιστωθεί ότι ο συνδυασμός των υβριδικών μοντέλων με μη εποπτευόμενη βαθιά μάθηση (deep learning) και τυπικούς κανόνες δημιουργεί πλαίσιο ιδιαίτερα ικανοποιητικό (Huang, et al., 2019), (Ghanbari-Adivi & Mosleh, 2019), (Chatterjee, et al., 2019)

2 Βιβλιογραφική έρευνα πεδίου και μεθοδολογικό πλαίσιο ανάλυσης των δεδομένων της διατριβής.

2.1 Εισαγωγή

Η παρούσα θέση εστιάζει στο πρόβλημα της ανάλυσης κειμένων με σκοπό τον εντοπισμό των υποβοσκοτών συναισθημάτων. Για το σκοπό αυτό αναλύονται κείμενα που αφορούν ηλεκτρονικά περιοδικά του ISIS καθώς και κείμενα από μέσα μαζικής ενημέρωσης του Δυτικού κόσμου. Ο στόχος της ανάλυσης είναι διττός. Κατά πρώτον, επιχειρείται να προσδιορισθεί το σύνολο των συναισθημάτων, που στόχευε να αφυπνίσει ο ISIS στους αναγνώστες του. Κατά δεύτερο, επιχειρείται η σύγκριση των κειμένων αυτών με κείμενα που απαντώνται σε δυτικά μέσα μαζικής ενημέρωσης προκειμένου να διαπιστωθεί αν τα κείμενα που ανήκουν σε περιοδικά του ISIS περιέχουν χαρακτηριστικά με βάση τα οποία θα μπορούσαν να θεωρηθούν ως ανώμαλα δεδομένα μέσα στη γενικότερη συλλογή των κειμένων που αποτελείται τόσο από κείμενα του Δυτικού όσο και από κείμενα του ISIS.

Η δομή του 2^{ου} κεφαλαίου έχει ως εξής:

- Στην ενότητα 2.2 γίνεται μια βιβλιογραφική αναφορά μελετών σχετικών με την προπαγάνδα του ISIS.
- Στην ενότητα 2.3 παρατίθεται βιβλιογραφική αναφορά μελετών σχετικών με την αξιοποίηση των μέτρων απόστασης Hellinger και Mahalanobis σε θέματα ανάλυσης. Όπως θα διαπιστωθεί, στη συνέχεια, τα μέτρα αυτά χρησιμοποιούνται στην παρούσα Θέση για τον εντοπισμό των ανώμαλων δεδομένων.
- Στην ενότητα 2.4 παρατίθεται βιβλιογραφική αναφορά μελετών σχετικών με τον εντοπισμό ανώμαλων δεδομένων κειμένου.
- Στις ενότητες από 2.5 έως 2.8 γίνεται περιγραφή μεθοδολογικών εργαλείων και αλγορίθμων που αξιοποιούνται στο προτεινόμενο πλαίσιο ανάλυσης συναισθήματος.

2.2 Βιβλιογραφική έρευνα πεδίου αναφορικά με την προπαγάνδα του ISIS

Ο ISIS αποτελεί μια οργάνωση που έχει απασχολήσει σημαντικά το Δυτικό Κόσμο, τόσο σε επίπεδο πολιτικής όσο και σε επίπεδο Μαζικών Μέσων Ενημέρωσης. Οι ενέργειες του ISIS (πολεμικές επιχειρήσεις, προπαγανδιστικές καμπάνιες, τρομοκρατικές ενέργειες) προβάλλονταν σε μεγάλο βαθμό από τα παραδοσιακά κανάλια επικοινωνίας των ΜΜΕ

(τηλεόραση, ημερήσιος τύπος κλπ). Παράλληλα, ο ISIS φρόντιζε να προβάλλει τόσο τις επιτυχίες του στα πεδία των μαχών όσο και το ιδεολογικό του πιστεύω μέσα από το Ίντερνετ ή άλλες μορφές επικοινωνίας. Έχοντας αυτοπροσδιοριστεί ως οργάνωση που είχε στόχο την επίθεση (Carafella, et al., 2019) προσπάθησε να ενδυθεί το ρόλο του εκπληρωτή της ισλαμικής προφητείας περί επικράτησης (Gambhir, et al., 2016). Ο ISIS, έχοντας υιοθετήσει ένα φιλόδοξο και πολυεπίπεδο στρατηγικό πλαίσιο, στόχευε στην ταυτόχρονη επιβολή και ενίσχυση της παρουσίας του τόσο στα πεδία των μαχών όσο και παγκόσμια. Πέρα από τις μάχες, ο ISIS επένδυσε σε μεγάλο βαθμό τόσο στην κατήχηση παιδιών στο Ιράκ και τη Συρία όσο και στην προσέλκυση μαχητών από το εξωτερικό (Gambhir, et al., 2016). Ο δεύτερος στόχος του υλοποιήθηκε μέσα από μια εκτεταμένη αξιοποίηση των ΤΠΕ και τα αποτελέσματα της ήταν θεαματικά (Gambhir, 2016). Το φαινόμενο της διάχυσης της προπαγάνδας του ISIS έχει γίνει αντικείμενο έρευνας πολλών ερευνητών και, παράλληλα, αποτελεί την πηγή πολλών ερευνητικών ερωτημάτων της παρούσας Θέσης. Για το λόγο αυτό παραθέτουμε μια εκτεταμένη βιβλιογραφική έρευνα σχετικά με επιστημονικές μελέτες και άρθρα που είχαν σαν στόχο να μελετήσουν την προπαγάνδα του ISIS.

Οι Cunningham et al. (2014) εξετάζουν το ενδεχόμενο ο ISIS να αποτελεί μια πιο ακραία ριζοσπαστική οργάνωση από την Al-Qaeda μέσα από τη διαδικτυακή παρουσία του στο Twitter. Επισημαίνουν την αξία που θα είχε η ανάλυση των λογαριασμών χρηστών και των κυρίαρχων θεμάτων και ιδεών για να κατανοηθεί το αφήγημα της οργάνωσης. Παράλληλα, επισημαίνουν τη μετατόπιση του κέντρου βάρους των εχθροπραξιών στο πεδίο των δυτικών κοινωνιών (Cunningham, et al., 2014).

Ο Farwell (2014) υποστηρίζει ότι ο ISIS ξεχωρίζει για την εκλεπτυσμένη χρήση και κατανόηση των μέσων κοινωνικής δικτύωσης για την επίτευξη των στόχων της οργάνωσης. Μέσα από την επικοινωνιακή του στρατηγική στοχεύει να πείσει όλους τους Μουσουλμάνους για την αναγκαιότητα της επαναδημιουργίας του χαλιφάτου ενώ επιχειρεί να απεικονίσει τον εαυτό του ως τον υπέρμαχο της κοινωνικής δικαιοσύνης. Μέσα από το Twitter, το Facebook και το Instagram στοχεύει να επηρεάσει πολλαπλούς ακροατές όπως αντιπάλους, φίλους και δημοσιογράφους. Ο Farwell έχει την άποψη ότι η συνύπαρξη εικόνων φρίκης (όπως αποκεφαλισμοί και εκτελέσεις) με ειρηνικές εικόνες (ξεκούραση των μαχητών, μαχητές παίζουν και ταΐζουν γάτες) δημιουργούν ένα αφήγημα που θέλει τον ISIS υπέρμαχο της προώθησης της ευημερίας των ανθρώπων και όχι της δολοφονίας τους (Farwell, 2014).

Ο Fernandez (2015), αντιπαραβάλλοντας τον ISIS με την Al-Qaida για το χρονικό διάστημα από το 2006 έως το 2014, διατυπώνει την άποψη ότι ο ISIS έχει αξιοποιήσει τα μέσα κοινωνικής δικτύωσης με έναν πρωτόγνωρο τρόπο προκειμένου να στρατολογήσει νέα μέλη στις δυνάμεις του ενώ, την ίδια στιγμή, φαίνεται ότι η αντιπροπαγάνδα της Δύσης δε μπορεί να ανακόψει τη δυναμική του ISIS. Τέλος, προτείνει ένα πιο αποτελεσματικό πλαίσιο αντιπροπαγανδιστικών ενεργειών (Fernandez, 2015).

Η Fisher (2015), ξεκινώντας από το γεγονός ότι οι τζιχαντιστικές ομάδες έχουν αποδείξει ότι είναι ικανές να δημιουργούν μια μόνιμη και ιδεολογικά συνεκτική παρουσία στα κοινωνικά δίκτυα σχετικά με την προσέλκυση μαχητών και χρηματικών δωρεών, χρησιμοποίησε τεχνικές μεγάλων δεδομένων, ανάλυσης δικτύου και, μέσα από παραδείγματα στρατηγικής επικοινωνίας των Τζιχαντιστών, εντόπισε τα χαρακτηριστικά στοιχεία που επέτρεψαν στις ομάδες αυτές να διατηρήσουν μόνιμη παρουσία για το περιεχόμενό τους στο διαδίκτυο. Η Fisher πιστεύει ότι η δραστηριότητα των υποστηρικτών του ISIS μέσα στο διαδίκτυο μπορεί να παρομοιαστεί με τη συμπεριφορά ενός σμήνους πουλιών που μετακινείται και αναδιατάσσεται ανάλογα με τις συνθήκες, Ως εκ τούτου, η ερευνήτρια θεωρεί ότι η διαγραφή μεμονωμένων λογαριασμών χρηστών δεν έχει αποτέλεσμα και θα πρέπει να αντιμετωπιστεί το πρόβλημα σε ενιαία βάση (Fisher, 2015).

Οι Berger & Morgan (2015), μέσα από ένα δείγμα 20.000 λογαριασμών του Twitter, επιχειρήσαν να προσδιορίσουν το πλήθος των υποστηρικτών του ISIS, τη γεωγραφική κατανομή τους, τον τύπο του λογαριασμού τους καθώς και τον τρόπο με τον οποίο ενεργούν διαδικτυακά, τον τρόπο με τον οποίο υποστηρίζουν το ISIS και πώς συμμετέχουν στις εξαιρετικά οργανωμένες διαδικτυακές του δραστηριότητες (Berger & Morgan, 2015).

Οι Magdy et al. (2015), έχοντας ως στόχο την καλύτερη κατανόηση της προέλευσης του ISIS συνέλεξαν ένα μεγάλο αριθμό από αραβόφωνα τιτιβίσματα σχετικά με τον ISIS και τα κατηγοριοποίησαν σε δύο κατηγορίες δηλαδή φίλα προσκείμενοι και εχθρικά προσκείμενοι. Στη συνέχεια, μέσα από ιστορικά στοιχεία αναζήτησαν τις ρίζες των ατόμων που ήταν φίλα προσκείμενοι προς τον ISIS. Η ανάλυση έδειξε να έχουν σχέση με την επανάσταση «Αραβική Άνοιξη» (Magdy, et al., 2015).

Οι Gates & Podder (2015), διαπιστώνουν ότι το Ισλαμικό Κράτος βασίζεται σε μεγάλο βαθμό στην στρατολόγηση ξένων μαχητών η οποία γίνεται με διάφορους τρόπους. Παρά το γεγονός

ότι η διαδικασία πρόσληψης αλλοδαπών θα μπορούσε να αποτελεί πρόβλημα για την Οργάνωση, η ανάλυση έδειξε ότι η διαχείριση του προβλήματος από τον ISIS είναι αποτελεσματική.

Ο Ferrara (2015) εστίασε στα προβλήματα που δημιουργούνται από την κλιμακούμενη συμπεριφορά των ατόμων που εκτίθενται σε κακή χρήση των μέσων κοινωνικής δικτύωσης από τρομοκρατικές οργανώσεις, οι οποίες βρίσκουν μεθόδους να παραπληροφωρήσουν, να χειραγωγήσουν και, τελικά, να ενορχηστρώσουν παράνομες ενέργειες.

Ο Prucha (2016) αναλύει δεδομένα που προέρχονται από την πλατφόρμα Telegram και ισχυρίζεται ότι στα μηνύματα διακρίνει κανείς ένα συνεκτικό σύνολο από ισχυρά αφηγήματα σε σχέση με τη jihad καθώς και μια έντονη παρουσία του ISIS στα μέσα κοινωνικής δικτύωσης και το Διαδίκτυο. Καταλήγει ότι θα πρέπει να δοθεί ιδιαίτερη έμφαση στην αποτελεσματική καταπολέμηση αυτών των πτυχών δραστηριότητας του οι οποίες, εκτός των άλλων, στοχεύουν στην επιβίωση της οργάνωσης ακόμα και αν αυτή απωλέσει τα εδάφη που έχει κυριέψει.

Οι Bodine-Baron et al. (2016) μελέτησαν αραβόφωνα δημοσιεύσεις στο Twitter και, προκειμένου να συνεισφέρουν στην προσπάθεια εντοπισμού των υποστηρικτών και των εχθρών του ISIS επιχείρησαν να αποτυπώσουν την πορεία της συζήτησης για τον ISIS. Η προσέγγιση τους βασίστηκε σε εργαλεία διαχείρισης δεδομένων μεγάλου όγκου και σε λεξικολογική ανάλυση κοινωνικών δικτύων. Οι ερευνητές ισχυρίζονται ότι οι εχθροί του ISIS τον αναφέρουν στις δημοσιεύσεις όχι ως «The Islamic State» αλλά ως «Daesh» ενώ οι υποστηρικτές του ISIS δημοσιεύουν 50% περισσότερο από τους αντιπάλους του. Παράλληλα, εστίασαν στη γεωγραφική κατανομή των θετικών και αρνητικών αναρτήσεων.

Οι Davis et al. (2016) ερεύνησαν το πρόβλημα του εντοπισμού αυτοματοποιημένων πρακτόρων στα κοινωνικά δίκτυα (social bots). Τα προγράμματα αυτά μιμούνται την ανθρώπινη συμπεριφορά με απώτερο στόχο τη χειραγώγηση συζήτησης, τη διάχυση προπαγάνδας και την παραπληροφόρηση. Στα πλαίσια της αντιμετώπισης του προβλήματος, οι ερευνητές παρουσίασαν μια δημόσια διαθέσιμη υπηρεσία με την ονομασία BotOrNot που έχει στόχο να αξιολογεί αν ένας λογαριασμός Twitter εμφανίζει ομοιότητα με τα γνωστά χαρακτηριστικά των κοινωνικών ρομπότ.

Οι Ferrara et al. (2016) παρουσίασαν ένα πλαίσιο μηχανικής μάθησης που αξιοποιούσε περί τα εξήντα χαρακτηριστικά (μεταδεδομένα και χρονικά δεδομένα) για τον εντοπισμό εξτρεμιστών χρηστών. Το πλαίσιο αυτό εφαρμόστηκε σε ένα σύνολο δεδομένων εκατομμυρίων τιτιβισμάτων που είχαν δημιουργηθεί από χρήστες του Twitter των οποίων οι λογαριασμοί είχαν ανασταλεί με μη αυτοματοποιημένο τρόπο από το Twitter λόγω της εμπλοκής τους σε εξτρεμιστικές εκστρατείες. Εφάρμοσαν έξι σενάρια πρόβλεψης (στατικά και δυναμικά) με σκοπό τον εντοπισμό αφενός των εξτρεμιστών χρηστών, κατά δεύτερο λόγο των τακτικών χρηστών που θα μπορούσαν να υιοθετήσουν εξτρεμιστικό περιεχόμενο και τέλος, των χρηστών που θα προωθήσουν τις δημοσιεύσεις των εξτρεμιστών σε δικές τους επαφές.

Οι Gialamproukidis et al. (2016) αντιμετώπισαν το πρόβλημα της παρακολούθησης της δραστηριότητας των τρομοκρατικών οργανώσεων στα μέσα κοινωνικής δικτύωσης. Θέλοντας να προσδιορίσουν τους χρήστες που επηρεάζουν περισσότερο την κοινότητα, πρότειναν ένα νέο κεντρικό δείκτη καλούμενο Mapping Entropy Betweenness (MEB) τον οποίο αξιοποίησαν σε δεδομένα του Twitter προκειμένου να προσδιορίσουν τους λογαριασμούς χρηστών που επηρέαζαν σημαντικότερα τη διάδοση της προπαγάνδας του Χαλιφάτου.

Οι Ghajar-Khosravi et al. (2016) επιχείρησαν να προσδιορίσουν τα κυριότερα θέματα και έννοιες που ανακύπτουν από δημοσιεύσεις σε μέσα κοινωνικής δικτύωσης. Χρησιμοποίησαν ως δεδομένα αναρτήσεις γυναικών στο Twitter που διέκριντο θετικά για τον ISIS και διαπίστωσαν ότι είναι δυνατή η ανάπτυξη μεθόδων εντοπισμού των αναρτήσεων που είναι φίλα προσκείμενες προς τον ISIS.

Οι Benigni & Carley (2016) μελέτησαν την κοινότητα CASOS Jihadist Twitter Community (CJTC). Η κοινότητα αυτή αποτελείται από περισσότερους από 15.000 χρήστες Twitter που υποστηρίζουν το ISIS. Ανέλυσαν τις μεθόδους εντοπισμού και παρακολούθησης των κοινοτήτων που υποστηρίζουν τις εξτρεμιστικές τζιχαντιστικές οργανώσεις και προσδιόρισαν ένα ενεργό κοινωνικό botnet που επιχειρεί να ανυψώσει την κοινωνική επιρροή των χρηστών που υποστηρίζουν την ατζέντα της Jabhat al-Nusra.

Οι Khawaja & Khan (2016) καταδεικνύουν πόσο σημαντικά η διαδικτυακή προπαγάνδα του ISIS έχει επηρεάσει προς όφελος της οργάνωσης την άποψη που έχει για αυτή το κοινό στόχος.

Οι Mahood & Rane (2017), επισημαίνουν την ικανότητα που έχει επιδείξει ο ISIS να στρατολογεί μουσουλμάνους από όλο τον κόσμο, συμπεριλαμβανομένης της Δύσης, καθώς και την προπαγάνδα του για ριζοσπαστικοποίηση μέσα από αφηγήματα που χρησιμοποιούν επιλεκτικά πτυχές του Ισλάμ. Θεωρούν επιβεβλημένη την κατανόηση του τρόπου με τον οποίο ο ISIS δημιουργεί τα αφηγήματα της προπαγάνδας του με στόχο την ανάπτυξη αποτελεσματικής αντιμετώπισης της εξάπλωσης της αποδοχής των θέσεων του.

Ο Ingram (2017) αναδεικνύει τη σημασία που έχουν για την προπαγάνδα του ISIS το αγγλόφωνο περιοδικό Dabiq το οποίο, στη συνέχεια, αντιπαραβάλλεται με το περιοδικό Inspire της Al-Qaida της Αραβικής Χερσονήσου. Ο σκοπός της έρευνας ήταν να μελετήσει τον τρόπο με τον οποίο σχεδιάζονταν τα αφηγήματα των περιοδικών που απέβλεπαν στην ριζοσπαστικοποίηση και την ιδεολογική πόλωση του ακροατηρίου τους. Τέλος, ο συγγραφέας δίνει μια ανάλυση ενός πλήθους στρατηγικών που αξιοποιούνται από το Dabiq για να ενισχύσει την οπτική των μηνυμάτων του.

Ο Ferrara (2017), επανερχόμενος στη σημασία που έχουν τα μέσα κοινωνικής δικτύωσης στη ριζοσπαστικοποίηση των ατόμων, ανέλυσε τη δραστηριότητα εικοσιπέντε χιλιάδων χρηστών κοινωνικών δικτύων προκειμένου να προσδιορίσει την κοινωνική επίδραση των μηνυμάτων στους υποστηρικτές του ISIS. Η μεθοδολογία που χρησιμοποιήθηκε αφορούσε την ανάλυση φυσικής γλώσσας και τους συνδετικούς χάρτες δυναμικής δραστηριότητας. Ο Ferrara ισχυρίζεται ότι ο τρόπος δράσης του ISIS μέσω των πλατφορμών κοινωνικής δικτύωσης οδηγεί στην άποψη πως οι χρήστες που αναμετέδιδαν τα μηνύματα του ISIS ακολουθούσαν πολύπλοκες στρατηγικές με σκοπό να επηρεάσουν άλλους χρήστες.

Οι Cao et al. (2017) εξέτασαν τη συμπεριφορά που εμφάνιζαν χρήστες των μέσων κοινωνικής δικτύωσης. Χρησιμοποίησαν την πλατφόρμα VKontakte (www.vk.com) που απετέλεσε την πρώτη πλατφόρμα που αξιοποίησε ο ISIS προκειμένου να διασπείρει την προπαγάνδα του. Μέσα από ένα σύνολο χρηστών που άγγιζε τα 350 εκατομμύρια άντλησαν δεδομένα που επιτρέπουν να ισχυριστούμε την ύπαρξη δυναμικών προτύπων στις διαδικτυακές διαδρομές που ακολουθούν τα άτομα που αναπτύσσουν, σε υψηλό βαθμό, υποστήριξη σε εξτρεμιστές

γενικά και στον ISIS ειδικά. Μέσα από την καταγραφή των γεωγραφικών και άλλων χαρακτηριστικών, οι ερευνητές πιστεύουν ότι μπορούμε να προβλέψουμε μελλοντικούς κινδύνους, χωρίς να περιοριζόμαστε μόνο στην πληροφορία που απορρέει από στατικές παραμέτρους των ατόμων όπως η εθνικότητα τους, η μεταναστευτική τους κατηγορία (πολιτικός πρόσφυγας/οικονομικός μετανάστης κλπ).

Οι Alfifi & Caverlee (2017) εξέτασαν τη συμπεριφορά μακροχρόνιων χρηστών που είχαν διαγραφεί από τα μέσα κοινωνικής δικτύωσης, μέσα από μια έρευνα σε αραβόφωνους χρήστες του Twitter. Μέσα από τη μελέτη των τιτιβισμάτων που εξέπεμπαν, των γλωσσικών χαρακτηριστικών που χρησιμοποιούσαν και της δραστηριότητάς τους διαπίστωσαν ότι, αν και φρόντιζαν να κρύβονται επιμελώς, οι μακροχρόνιοι χρήστες εμφάνιζαν ιδιότυπη συμπεριφορά που τους διαφοροποιούσε τόσο από τους συνήθεις χρήστες όσο και από τους υποστηρικτές του ISIS. Η εν λόγω συμπεριφορά θα μπορούσε να επιτρέψει τον έγκαιρο εντοπισμό των χρηστών αυτών σε πολύ πιο πρώιμο στάδιο της δραστηριότητάς τους.

Οι Klausen et al. (2018), συνθέτοντας τα προβλήματα που αντιμετώπισαν οι Gialamproukidis et al. (2016) και Ghajar-Khosravi et al. (2016), κινήθηκαν ερευνητικά στην ανάπτυξη μεθοδολογίας εντοπισμού των νέων λογαριασμών που δημιουργούσαν οι χρήστες που είχαν διαγραφεί από μια πλατφόρμα κοινωνικής δικτύωσης εξαιτίας της έντονης δραστηριότητάς τους που είχαν αναφορικά με την προπαγάνδα του ISIS. Χρησιμοποίησαν δεδομένα από το Twitter και ανέπτυξαν ένα συμπεριφορικό μοντέλο το οποίο προσδιόριζε ομάδες λογαριασμών που ανήκαν στον ίδιο χρήστη.

Οι Badaway & Ferrara (2018) χρησιμοποίησαν ένα σύνολο δεδομένων που άγγιζε τα δύο εκατομμύρια τιτιβίσματα του Twitter και είχαν δημοσιευτεί από 25.000 υποστηρικτές του ISIS. Τα δεδομένα αυτά αξιοποιήθηκαν για να αποτυπωθεί ο τρόπος με τον οποίο ο ISIS διαχέει την προπαγάνδα του μέσω των κοινωνικών δικτύων και προσελκύει νέους αραβόφωνους μαχητές τόσο από τον αραβικό κόσμο όσο και από τις υπόλοιπες περιοχές της Υφήλιου. Χρησιμοποίησαν προσέγγιση βασισμένη σε λεξικό και τη λογική του σάκου των λέξεων. Οι ερευνητές διαπίστωσαν ότι στα αραβόφωνα μηνύματα εμφανίζονταν τρία θέματα που έχουν σχέση με τη βία, τη θρησκεία και την αίρεση (Adam Badaway & Ferrara, 2018).

Ο Mansour (2018) συνέλεξε tweets σχετικά με τον ISIS από τις παρακάτω οκτώ χώρες: Βέλγιο, Μεγάλη Βρετανία, Γαλλία, ΗΠΑ, Αίγυπτος, Συρία, Σαουδική Αραβία, Τουρκία. Στη συνέχεια, τα ανέλυσε μέσω της μεθοδολογίας TF-IDF και διαπίστωσε ότι (α) οι χρήστες χρησιμοποιούσαν στα τιτιβίσματα τους, αναφορικά με τον ISIS, σχεδόν τις ίδιες λέξεις και (β) η πλειονότητα των χρηστών αντιμετώπιζαν τον ISIS ως πηγή απειλής και φόβου.

Στην ίδια κατεύθυνση κινήθηκαν οι Ozeren et al. (2018), προκειμένου να αναλύσουν τον τρόπο με τον οποίο ο ISIS προσέλκυε και στρατολογούσε γυναίκες από την περιοχή της Τουρκίας. Οι παραπάνω ερευνητές ανέλυσαν τη στρατηγική με την οποία η οργάνωση αξιοποιούσε τα μέσα κοινωνικής δικτύωσης για να διασπείρει την προπαγάνδα του και να προσελκύσει γυναίκες-συντρόφους για τους μαχητές του ISIS.

Οι Speckhard et al. (2018) υλοποίησαν μια εκστρατεία 24 ημερών μέσω διαφημίσεων στο Facebook, προκειμένου να προωθήσουν μηνύματα που να αποδομούν την προπαγάνδα του ISIS. Η ομάδα-στόχος της έρευνας-εκστρατείας αφορούσε άτομα ηλικίας 18-50 ετών που θα μπορούσαν να επηρεαστούν από το ριζοσπαστικό αφήγημα του ISIS και θα μπορούσαν να εμπλακούν σε παραπέρα τρομοκρατικές ενέργειες. Το ηλικιακό φάσμα της ομάδας-στόχου είναι εκείνο που έχει και τη μεγαλύτερη δραστηριότητα στην επικοινωνία μέσω του Facebook. Τα αποτελέσματα της έρευνας ήταν ενθαρρυντικά μια και το πολυμεσικό υλικό της εκστρατείας άγγιξε το 1,7 εκατομμύρια θεάσεις και εκατοντάδες θετικά σχόλια.

Το Jones (2018) αναφέρει ότι η εκστρατεία που έχει εξαπολύσει ο ISIS, αναφορικά με την καταστροφή μνημείων παγκόσμιας κληρονομιάς, εντάσσεται στο ευρύτερο πλαίσιο της προπαγάνδας της οργάνωσης. Με τις ενέργειες αυτές, ο ερευνητής ισχυρίζεται ότι ο ISIS απορρίπτει την υπόσταση των κρατών Συρίας και Ιράκ και προβάλλει στο ακροατήριό του την άποψη ότι εξαγνίζει την περιοχή από ειδωλολατρικά στοιχεία.

Οι Cunliffe, E., & Curini, L. (2018) υιοθετώντας την ίδια οπτική με τον Jones, καταδεικνύουν ότι ένα μέρος της προπαγάνδας του ISIS, χρησιμοποιούσε στα μηνύματα πολυμεσικό υλικό σχετικό με την καταστροφή μνημείων παγκόσμιας κληρονομιάς. Χρησιμοποιώντας ως δεδομένα αναρτήσεις του Twitter και ανάλυση υποκειμενικής γνώμης μελέτησαν τις αντιδράσεις της διαδικτυακής κοινότητας στα μηνύματα αυτά.

Οι Bengini et al. (2018), ανέλυσαν δεδομένα από την κοινότητα του Twitter σχετικά με τον ISIS και την σύγκρουση της Κριμαίας και πρότειναν ένα πλαίσιο για την ιδεολογική συσταδοποίηση των χρηστών του Twitter.

Οι Alyousef & Hasmah (2018) έχουν την άποψη ότι η παρουσία του ISIS και άλλων ακραίων οργανώσεων στα μέσα κοινωνικής δικτύωσης έχει ενισχυθεί δραματικά με αποτέλεσμα να αποτελούν κίνδυνο για τη νεολαία. Υλοποίησαν έρευνα προκειμένου να προσδιορίσουν τα αποτελέσματα που έχει η προπαγάνδα του ISIS στους νέους, μέσα από τη χρήση των κοινωνικών δικτύων.

Οι Alfifi et al. (2018), μέσω δεδομένων του 2015 από το Twitter, εξέτασαν το βαθμό στον οποίο είναι ικανός ο ISIS να διαχέει τα μηνύματα της προπαγάνδας του μέσω των υποστηρικτών του.

Οι McElreath et al. (2018), χρησιμοποιώντας την μέθοδο OSINT και επισημαίνοντας το ρόλο που έχουν τα μέσα κοινωνικής δικτύωσης στη διαδικασία ριζοσπαστικοποίησης, καταλήγουν στην άποψη ότι ο ISIS πέτυχε να εξασφαλίσει μέσα από το διαδίκτυο μια παγκόσμια προβολή του μηνύματος ριζοσπαστικοποίησης του και της αποδοχής των ατόμων να πάρουν μέρος σε επιθέσεις στα πεδία των μαχών ή στον τόπο διαμονής τους. Ως εκ τούτου, τα μέσα κοινωνικής δικτύωσης μπορούν να θεωρηθούν ως ένα άριστο διαδικτυακό μέσο στελέχωσης των εξτρεμιστικών οργανώσεων.

Αναγνωρίζοντας τη σημασία του περιεχομένου των τιτιβισμάτων, οι Fadel & Öz (2020) ανέλυσαν τις αναρτήσεις που ακολουθούσαν μετά από επιθέσεις του ISIS με σκοπό να κατηγοριοποιήσουν τους χρήστες. Η μεθοδολογία που ακολούθησαν ανήκει στην εποπτευόμενη μάθηση και την ανάλυση συναισθήματος. Αξιοποίησαν λεξικό όρων για να δημιουργήσουν θετικές ή αρνητικές ετικέτες ενός συνόλου δεδομένων εκπαίδευσης και να επιτρέψει τον προσδιορισμό των υποστηρικτών του ISIS μέσα από το περιεχόμενο των τιτιβισμάτων τους.

Οι Ahmad et al. (2019), έχοντας τη γνώμη ότι συμμορίες και τρομοκρατικές οργανώσεις χρησιμοποιούν σημαντικά τα μέσα κοινωνικής δικτύωσης για να διαδώσουν την ιδεολογία τους και να προσελκύσουν νέα μέλη, ανέπτυξαν , μέσω ανάλυσης υποκειμενικής γνώμης, ένα σύστημα κατηγοριοποίησης των τιτιβισμάτων του Twitter σε ριζοσπαστικά και μη ριζοσπαστικά.

Στη μελέτη τους, με τίτλο «Blood and ink: the relationship between Islamic State propaganda and Western media», οι Courty et al. (2019) εξέτασαν την ακούσια συνενοχή των δυτικών μέσων ενημέρωσης στη διάδοση της προπαγάνδας του Ισλαμικού Κράτους. Παίρνοντας ως μελέτη περίπτωσης τις επιθέσεις του Νοεμβρίου 2015, ομαδοποίησαν τα βασικά μηνύματα του Ισλαμικού Κράτους στα δύο θέματα «φοβερός εχθρός» και «σύγκρουση πολιτισμών». Τα ευρήματα που αντλήθηκαν από τα αφηγήματα των εφημερίδων New York Times, The Times, Daily Mail και Le Figaro, και, σύμφωνα με τους ερευνητές, οδηγούν στο συμπέρασμα ότι αυτές οι πηγές ειδήσεων αντιγράφουν την προπαγάνδα του Ισλαμικού Κράτους άμεσα και έμμεσα σε διαφορετικό βαθμό.

Οι Borau & Wamba (2019), εστιάζουν στους λόγους που ωθούν τους άνδρες να ριζοσπαστικοποιηθούν και να ενταχθούν στον ISIS. Αυτό το αντικείμενο είναι ζωτικής σημασίας προκειμένου να αναπτυχθεί πλαίσιο αποτελεσματικής εκστρατείας αντιπροπαγάνδας μεταξύ των πιο ευάλωτων ατόμων. Χρησιμοποιώντας εξελικτική ψυχολογία προσδιορίζουν τους μηχανισμούς που χρησιμοποιεί ο ISIS στα μέσα κοινωνικής δικτύωσης για να προσηλυτίσει νέα μέλη. Οι ερευνητές υποστηρίζουν ότι ο ISIS, μέσα από αφηγήματα προπαγάνδας που σχετίζονται με τη βία, τον ανταγωνισμό, το φόβο, την εύρεση συντρόφων και την ηθική, παρακινεί τους άνδρες να ενταχθούν στο ISIS. Παράλληλα, καταδεικνύουν τη σημασία που θα είχε η ανάπτυξη στρατηγικών και εργαλείων για να αποτραπεί η χρήση των μέσων κοινωνικής δικτύωσης για τρομοκρατικούς σκοπούς (Borau & Wamba, 2019).

Η Windsor (2020), μέσα από τη μελέτη περίπτωσης της μουσουλμάνας Aqsa Mahmood που εγκατέλειψε το σπίτι της στη Σκωτία για να ενταχθεί στους μαχητές του ISIS στη Συρία, αναλύουν τα γλωσσικά μοτίβα των αναρτήσεων στο Tumblr για να αποκαλύψουν την τυπολογία της γλώσσας και τη διαδικασία της γυναικείας ριζοσπαστικοποίησης.

Ο Winter (2020) αμφισβητεί τον τρόπο με τον οποίο χρησιμοποιείται η λέξη «προπαγάνδα» στη σύγχρονη συζήτηση αναφορικά με τον πόλεμο και την τρομοκρατία. Παίρνοντας ως παράδειγμα, την περίπτωση του Ισλαμικού Κράτους, ισχυρίζεται ότι ο όρος προπαγάνδα σήμερα αναφέρεται σε ένα ολόκληρο σύνολο πληροφοριών που στοχεύουν σε διαφορετικούς σκοπούς και ως εκ τούτου, η παραδοσιακή ερμηνεία του όρου είναι

ανεπαρκής όταν πρόκειται να περιγράψει το πλήρες φάσμα των τακτικών και στρατηγικών προσεγγίσεων επικοινωνίας που χρησιμοποιούν οι τρομοκράτες σήμερα.

Οι Jain & Vaidya (2021) κάνουν μια εκτεταμένη αναφορά τρομοκρατικών ενεργειών ενώ επισημαίνουν την εσφαλμένη χρήση των μέσων κοινωνικής δικτύωσης λόγω της αξιοποίησης τους από τρομοκρατικές οργανώσεις με σκοπό τη διάδοση της προπαγάνδας τους και την προσέλκυση νέων μελών και την ανάγκη να στραφεί η επιστημονική κοινότητα σε μια συντονισμένη αντιμετώπιση του προβλήματος.

2.3 Βιβλιογραφική έρευνα αναφορικά με την αξιοποίηση της απόστασης Mahalanobis και της απόστασης Hellinger στην ανάλυση υποκειμενικής γνώμης

Μια σημαντική εργασία, αναφορικά με την ανάλυση κειμένου, έχει να κάνει με την κατηγοριοποίηση των κειμένων. Η ανάλυση αυτή παρουσιάζει ιδιαίτερο ενδιαφέρον ειδικά όταν το σύνολο των προς ανάλυση δεδομένων δεν είναι ισορροπημένο. Σημαντικοί παράγοντες που επηρεάζουν τα συμπεράσματα της ανάλυσης είναι τόσο ο τρόπος μέτρησης της απόστασης όσο και ο αλγόριθμος κατηγοριοποίησης. Λόγω των ερευνητικών ερωτημάτων της παρούσας θέσης, έχουμε επιλέξει την αξιοποίηση τόσο των μέτρων απόστασης Mahalanobis και Hellinger όσο και την αξιοποίηση των μεθόδων μη εποπτευόμενης κατηγοριοποίησης. Για το λόγο αυτό υλοποιήσαμε βιβλιογραφική έρευνα αναφορικά με την αξιοποίηση των παραπάνω αποστάσεων που αφορά το χρονικό διάστημα από το 2010 έως σήμερα. Στη συνέχεια, παρατίθενται τα ευρήματα.

Το 2010, οι Seroussi et al. παρουσίασαν μια συνεργατική προσέγγιση βασισμένη την μεθοδολογία του πλησιέστερου γείτονα (nearest neighbor) με σκοπό να βελτιώσουν την απόδοση των μοντέλων προσδιορισμού ομοιότητας χρηστών (users' similarity) με βάση τις αξιολογήσεις που αυτοί υποβάλουν. Οι εν λόγω ερευνητές αξιοποίησαν την απόσταση Hellinger ως το μέτρο ομοιότητας επισημαίνοντας ότι είναι σκόπιμο να αξιοποιείται αυτό το μέτρο απόστασης όταν διατίθεται μεγάλο δείγμα αξιολογήσεων χρηστών (Seroussi Y., 2010).

Το ίδιο έτος, οι Boyd-Graber & Resnik ανέπτυξαν ένα πολυγλωσσικό εποπτευόμενο μοντέλο με κατανομή Dirichlet (MLSLDA) και χρήση της απόστασης Hellinger. Σύμφωνα με τους ερευνητές, η προτεινόμενη μέθοδος επιτρέπει την ανακάλυψη συνδέσεων μεταξύ γλωσσών

και την αποκάλυψη υποβόσκων δομών σε παράλληλες συλλογές ενώ βοηθά στο σχηματισμό πολυγλωσσικών καταλόγων με συναισθηματικές συσχετίσεις (Boyd-Graber & Resnik, 2010)

Οι Gonzalez et al. (2013) αντιμετώπισαν το πρόβλημα της αυτόματης κατανομής σε ομάδες ενός συνόλου δεδομένων το οποίο δεν περιείχε ετικέτες. Η προσέγγιση τους βασίστηκε στην απόσταση Hellinger και τα αποτελέσματα συγκρίθηκαν με εκείνα που έδινε ένα μοντέλο κατηγοριοποίησης βασισμένο στην προσέγγιση Naïve. Σύμφωνα με τους ερευνητές, το μοντέλο κατηγοριοποίησης που αξιοποιούσε την απόσταση Hellinger υπερτερούσε του αντίστοιχου με τη μέθοδο Naïve, ειδικά στις περιπτώσεις που τα αναλυόμενα δεδομένα παρουσίαζαν ασυμμετρία (Gonzalez-Castro, et al., 2013).

Οι Krstovski & Smith (2013) εργάστηκαν πάνω στο θέμα της αποτελεσματικής αναζήτησης σε μια μεγάλη πολύγλωσση συλλογή με σκοπό την εύρεση εγγράφων που αποτελούν το ένα μετάφραση του άλλου. Αξιοποιώντας ιεραρχικά Bayesian μοντέλα σε πολύγλωσσες συλλογές, χρησιμοποίησαν διάφορες μετρήσεις απόκλισης, όπως Kullback-Leibler (KL), Jensen-Shannon (JS) και απόσταση Hellinger. Οι ερευνητές υποστήριξαν ότι μέσα από την αξιοποίηση της απόστασης Hellinger σημειώθηκε βελτίωση ταχύτητας 24,2 φορές σε σύγκριση την απόκλιση JS όλων των ζευγών, ενώ διατηρήθηκε η ίδια απόδοση.

Στην ίδια κατεύθυνση, οι Krstovski et al. (2013) ασχολήθηκαν με το θεμελιώδες ερώτημα της σύγκρισης εγγράφων μέσα από πιθανές αναπαραστάσεις τους μέσω latent topic distributions. Στην εργασία τους υιοθετούν την LDA και την απόσταση Hellinger έναντι της Ευκλείδειας απόστασης. Σύμφωνα με τους ερευνητές, η προσέγγιση τους επιτρέπει να εκμεταλλευτούν αποτελεσματικότερα μεθόδους όπως αυτή του πλησιέστερου γείτονα (NN), του locality-sensitive hashing (LSH) και του approximate search in k-d trees (Krstovski, et al., 2013)

Το ίδιο έτος οι Kim et al. εισήγαγαν την αναπαράσταση των ανθρώπινων συναισθημάτων μέσα από συνεχή μεταβλητή και πρότειναν ένα στατιστικό μοντέλο το οποίο συνδέει X έγγραφα με ένα διακριτό σύνολο συναισθημάτων Y . Χρησιμοποίησαν την κανονικοποίηση L1 και επέλεξαν ως μέτρο απόστασης την Hellinger λόγω του ότι έχει καλύτερη απόδοση από την Ευκλείδεια.

Το έτος 2014, οι Lebet & Collobert επέλεξαν την απόσταση Hellinger προκειμένου να πραγματοποιήσουν ανάλυση κύριων συνιστωσών (PCA) με σκοπό την αναπαράσταση ενός

λεξιλογίου σε χώρο μικρότερων διαστάσεων με ταυτόχρονη ελαχιστοποίηση του σφάλματος αναδημιουργίας των κειμένων.

Ο Bhargav (2014) ανέπτυξε ένα πρωτότυπο σύστημα ικανό να εξάγει χαρακτηριστικά ταινίας ως θέματα. Χρησιμοποίησε μια συλλογή ταινιών και με το μοντέλο εκπαίδευσης επιχείρησε να βρει παρόμοιες ταινίες χρησιμοποιώντας την απόσταση Hellinger. Σύμφωνα με τον ερευνητή, μια τέτοια προσέγγιση δίνει καλά αποτελέσματα ακόμα και σε μικρές συλλογές ταινιών.

Οι Wachsmuth et al. (2015) παρουσίασαν ένα μοντέλο εκπαίδευσης με στόχο την πρόβλεψη της συνολικής άποψης που αφορά μια ανασκόπηση. Αξιολόγησαν τη διαφορά μεταξύ των συνολικών κατανομών συναισθήματος χρησιμοποιώντας την απόσταση Hellinger. Επίσης, αξιολόγησαν τις σταθμισμένες αποστάσεις Hellinger για να καταγράψουν τις διαφορές που εμφανίζονται μεταξύ των παραλλαγών, οι οποίες, σύμφωνα με τους συγγραφείς, είναι μικρές.

Οι Amati et al. (2016) εξήγαγαν ποσοστά κατηγοριών συναισθήματος χρησιμοποιώντας την μέθοδο Divergence-Based Model (DVM), την Πολυωνυμική Naive Bayes (MNB) και τον αλγόριθμο SVM. Υιοθετήθηκε η απόσταση Mahalanobis και αξιοποιήθηκαν τόσο το τεστ Kolmogorov-Smirnov για να ελεγχθεί αν οι παρατηρούμενες και προσαρμοσμένες τιμές προέρχονται από την ίδια κατανομή για κάθε μέθοδο και ταξινομητή όσο και ο δείκτης συσχέτισης Pearson. Τα αποτελέσματα έδειξαν ότι μεταξύ των προσεγγίσεων που συγκρίθηκαν μόνο το SVM και το MNB έδιναν αποτελέσματα που ικανοποιούσαν το τεστ Kolmogorov-Smirnov τόσο για θετικές όσο και για αρνητικές κατηγορίες.

Το έτος 2016, οι Balasubramanian et al. παρουσίασαν ένα υβριδικό μοντέλο ταξινομητή προτύπων που αξιοποιούσε την προσέγγιση SVM σε δεδομένα που περιείχαν unigrams, bigrams, trigrams. Η απόσταση Mahalanobis επιλέχθηκε ως μέτρο απόστασης για να αποτυπώσει τη συνάφεια μεταξύ κειμένων που αφορούσαν θετικές ή αρνητικές κριτικές ταινίας. Οι ερευνητές υποστήριξαν ότι η προτεινόμενη προσέγγιση οδήγησε σε μια ταξινόμηση με ακρίβεια ταξινόμησης 98,8%.

Το ίδιο έτος, οι Xiong et al. ανέλυσαν δεδομένα κειμένου σε επίπεδο έννοιας. Τα κείμενα αφορούσαν ανασκόπηση πελατών. Συγκρίνοντας έναν αριθμό μεθόδων, οι οποίες περιλαμβάνουν τις μεθόδους αιχμής ως τις βασικές νευρωνικές μεθόδους και υιοθετώντας

ως μέτρο απόστασης τη συνάρτηση Mahalanobis, οι ερευνητές υποστήριξαν ότι η προσέγγισή τους παρέχει καλύτερη απόδοση σε σχέση με τις υπόλοιπες προσεγγίσεις που εξετάστηκαν.

Οι Al-Amin, et al. (2017) αξιολόγησαν έξι διαφορετικές προσεγγίσεις ανάλυσης υποκειμενικής γνώμης σε επίπεδο πρότασης προσδιορίζοντας τις επιδόσεις της καθεμίας προσέγγισης. Αξιοποιώντας ποικίλες μεθόδους όπως ομοιότητα συνημιτόνου, TF-IDF, το μοντέλο Naïve Bayes χρησιμοποίησαν unigrams και stemmer με ανάλυση κύριων συνιστωσών (PCA) βασισμένη στην απόσταση Hellinger.

Το ίδιο έτος οι Xu, Y. et al. (2017) αντιμετώπισαν το πρόβλημα όπου τα δεδομένα εκπαίδευσης και τα δεδομένα δοκιμής προέρχονται από διαφορετικές κατανομές. Προκειμένου να μειωθεί η απόκλιση κατανομής μεταξύ των δεδομένων αυτών, αντικατέστησαν την Ευκλείδεια απόσταση με την απόσταση Mahalanobis ώστε να καταγράψουν την εγγενή ομοιότητα ή ανομοιότητα μεταξύ των διαφόρων περιπτώσεων.

Οι Li et al. (2017) αντιμετώπισαν το πρόβλημα της μη εποπτευόμενης κατηγοριοποίησης και χρησιμοποίησαν την απόσταση Hellinger για να συγκρίνουν εάν οι κατανομές πηγής και στόχου είναι παρόμοιες. Υποστηρίξαν ότι η χρήση μιας μετρικής, όπως η απόσταση Hellinger, βελτιώνει την ακρίβεια του πλαισίου αναγνώρισης του μοντέλου κατηγοριοποίησης.

Οι Cong et al. (2018), στα πλαίσια έρευνας με σκοπό την ανίχνευση των ατόμων με κατάθλιψη μέσω αναρτήσεων από κοινωνικά δίκτυα, πρότειναν μια προσέγγιση βαθιάς μάθησης (deep learning). Το πρόβλημα στο οποίο εστίαζαν οι συγγραφείς αφορούσε την ανισορροπία δεδομένων που παρατηρείται στα σύνολα δεδομένων του πραγματικού κόσμου. Έτσι, αξιοποιήθηκε ένα σύνολο δεδομένων σχετικών με δημοσιεύσεις 9.000, περίπου, χρηστών που ισχυρίζονταν ότι είχαν διαγνωστεί με κατάθλιψη και τις δημοσιεύσεις από 107.000 χρήστες ελέγχου. Στη συνέχεια, επιχειρήθηκε η αντιστοίχιση των χρηστών ελέγχου με κάποιον από τους διαγνωσμένους χρήστες. Το αποτέλεσμα ήταν να αντιστοιχιστεί σε καθένα διαγνωσμένο χρήστη ένα σύνολο από δώδεκα χρήστες ελέγχου μέσω της αξιοποίησης της απόστασης Hellinger που προσδιοριζόταν μεταξύ των κατανομών πιθανοτήτων των κειμένων ανάρτησης του διαγνωσμένου χρήστη και του χρήστη ελέγχου.

Το έτος 2018, οι Dai & Man πρότειναν ένα πλαίσιο ανάλυσης συναισθήματος στα μέσα κοινωνικής δικτύωσης αποτελούμενο από έναν συναισθηματικό οπτικό περιγραφέα και μια

ολοκληρωμένη μέθοδο οπτικοκειμενικής ταξινόμησης. Με τον τρόπο αυτό, οι ερευνητές δημιούργησαν ένα δομημένο δάσος (structured forest) για τη δημιουργία ενός σάκου λέξεων (bag of words) που αφορούν συναισθήματα που θα επέτρεπε την παραπέρα ανάλυση συναισθήματος. Στη συνέχεια, χρησιμοποιείται ένα σύνολο χαρακτηριστικών μερών συναισθήματος για την ενσωμάτωση των οπτικών και κειμενικών περιγραφών σε στατιστικές πολλαπλότητες (manifolds). Τέλος, εφαρμόστηκε κατηγοριοποίηση συναισθημάτων πολλαπλής κλίμακας μέσω της μετρικής Mahalanobis.

Οι Abdul-Fattah et al. (2018) συγκέντρωσαν τα σχόλια ανασκόπησης των χρηστών τα οποία αφορούσαν τέσσερις κατηγορίες ήτοι «άριστα», «καλά», «κακά» και «δεν συνιστάται». Χρησιμοποίησαν την απόσταση Mahalanobis και ένα σετ εκπαίδευσης που έχει δύο ετικέτες, δηλαδή θετική και αρνητική. Υποστήριξαν ότι η προσέγγιση που υιοθέτησαν είχε ακρίβεια κατηγοριοποίησης μεγαλύτερη από 90%.

Οι Loyola-Gonzalez et al. (2019) χρησιμοποίησαν ανάλυση συναισθήματος σε δεδομένα κειμένου που προέρχονταν από τιτβίσματα (tweets) για να προσδιορίσουν την πολικότητα και το επίπεδο αντικειμενικότητας και ειρωνείας των τιτβισμάτων αυτών. Από την ανάλυση παράχθηκε ένα δέντρο απόφασης ενώ χρησιμοποιήθηκαν τρία διαφορετικά μέτρα απόστασης (απόσταση Hellinger, Bhattacharyya και, κέρδος πληροφοριών ή information gain).

Οι Wasid και Ali (2018) ασχολούνται με το ζήτημα που υπάρχει στα παραδοσιακά συστήματα συστάσεων (recommendation systems) που βασίζονται σε συνεργατικό φίλτράρισμα. Υποστηρίζουν ότι, αν και τα παραδοσιακά συστήματα συστάσεων υιοθετούνται ευρέως, στην πραγματικότητα δεν είναι σε θέση να παράγουν αποτελεσματικές συστάσεις σε περίπτωση πολυδιάστατων αντικειμένων. Προς αυτή την κατεύθυνση προτείνουν μια προσέγγιση ομαδοποίησης για την αποτελεσματική ενσωμάτωση αξιολογήσεων πολλαπλών κριτηρίων στα παραδοσιακά συστήματα συστάσεων. Η ενδοσυσταδική ομοιότητα των χρηστών αξιολογείται χρησιμοποιώντας την απόσταση Mahalanobis. Για να συγκρίνουν την προτεινόμενη προσέγγιση με την παραδοσιακή προσέγγιση συνεργατικού φίλτραρίσματος χρησιμοποίησαν το σύνολο δεδομένων Yahoo Movies. Σύμφωνα με τους συγγραφείς, η μέθοδος απόστασης Mahalanobis δημιούργησε πιο ακριβείς γείτονες για κάθε χρήστη μέσα στη συστάδα.

Το έτος 2020, οι Akiogamen et al. χρησιμοποίησαν δεδομένα που ελήφθησαν από διαδικτυακή συζήτηση γύρω από τα οπιούχα και τα χρησιμοποίησαν ως μελέτη περίπτωσης για να προσδιορίσουν τα υποβόσκοντα θέματα και να αποκαλύψουν κρυμμένες σημασιολογικές δομές στις αναρτήσεις. Για να μετρήσουν την ομοιότητα των κειμένων εντός του συνόλου των αναρτήσεων χρησιμοποίησαν απόσταση Hellinger, απόσταση Jensen-Sannon και την ομοιότητα συνημιτόνου. Υποστήριξαν ότι μέσα από την εξόρυξη κειμένου είναι εφικτός ο εντοπισμός και η ομαδοποίηση πληροφοριών σχετικών με τα διαφορετικά ενδιαφέροντα των χρηστών.

Οι Zelikman & Socker (2020) αντιμετώπισαν το πρόβλημα της καταγραφής της σημαντικότητας μιας λέξης χρησιμοποιώντας κατανομή διανυσμάτων συναφών εκφράσεων για να κανονικοποιηθούν οι αποστάσεις και τα βάρη στάθμισης. Η προσέγγιση που ακολούθησαν ανήκει στο χώρο των μοντέλων σάκου λέξεων ενώ αξιοποιήθηκε η απόσταση Mahalanobis. Σύμφωνα με τους συγγραφείς, η προτεινόμενη προσέγγιση επιτρέπει την ακριβή και μη εποπτευόμενη αναπαράσταση προτάσεων ή κειμένων με στόχο την κατηγοριοποίηση τους.

2.4 Βιβλιογραφική μελέτη σε σχέση με τον εντοπισμό ανωμαλιών σε κείμενο με μη εποπτευόμενη μάθηση

Η παρούσα ενότητα εστιάζει στην βιβλιογραφική έρευνα πεδίου αναφορικά με αλγορίθμους μη εποπτευόμενης μάθησης κατηγοριοποίησης για προσδιορισμό ανώμαλων δεδομένων. Στην κατεύθυνση αυτή, εστιάζουμε σε τρεις κατηγορίες αλγορίθμων και συγκεκριμένα:

- Στα συνελκτικά νευρωνικά δίκτυα (CNN),
- Στους αλγορίθμους που βασίζονται στο Support Vector Machine (SVM) και
- Στον αλγόριθμο Isolation Forest

Τα παραπάνω μεθοδολογικά εργαλεία, εποπτευόμενης ή μη εποπτευόμενης μάθησης, φαίνεται να έχουν την ικανότητα εξαγωγής περίπλοκων κρυφών χαρακτηριστικών από πολυδιάστατα δεδομένα με πολύπλοκη δομή. Όμως, όπως καταδεικνύουν πολλοί ερευνητές (Zhuang & Dai, 2006), η ανίχνευση ανώμαλων δεδομένων σε μη ισορροπημένα σύνολα δεδομένων μέσα από μη εποπτευόμενη μάθηση, αναδεικνύεται σε ένα ερευνητικό πρόβλημα που η επίλυση του θα εύρισκε πολλές εφαρμογές, όπως είναι η ασφάλεια υπολογιστικών συστημάτων, η ανάλυση συμπεριφοράς κ.λπ. Στη συνέχεια του κειμένου

παρατίθεται σχετική βιβλιογραφική έρευνα μελετών σχετικών με τον εντοπισμό ανωμαλιών σε δεδομένα κειμένου που αφορά το χρονικό διάστημα από το 2006 έως και σήμερα.

Οι Zuang & Dai (2006), διαπιστώνοντας ότι ένας ταξινομητής μίας τάξης (One class classifier) υπερτερεί αυτού των δύο τάξεων σε μη ισορροπημένα δεδομένα, πρότειναν ένα γενικό πλαίσιο που χρησιμοποιεί πρώτα τα δεδομένα μειοψηφίας (ανώμαλα δεδομένα) για εκπαίδευση και, στη συνέχεια, τη χρησιμοποίηση των κατηγοριών μειοψηφίας και πλειοψηφίας για την εκτίμηση της απόδοσης του προτεινόμενου ταξινομητή. Το προτεινόμενο μοντέλο δοκιμάστηκε σε δεδομένα κειμένου από το αποθετήριο μηχανικής μάθησης UCI και τους Reuters. Σύμφωνα με τους ερευνητές, τα αποτελέσματα δείχνουν ότι ο προτεινόμενος αλγόριθμος SVM μίας τάξης επιτυγχάνει πολύ καλύτερη απόδοση από τον τυπικό αλγόριθμο SVM μίας κατηγορίας καθώς και από άλλα σχήματα μη εποπτευόμενης μάθησης, όπως είναι το μοντέλο Naive Bayes και τα μοντέλα πλησιέστερου γείτονα μίας κατηγορίας.

Το έτος 2013, οι Ding & Fei πρότειναν έναν προσαρμοσμένο αλγόριθμο ανίχνευσης ανωμαλιών σε δεδομένα ροής (data stream) που βασίζεται στον αλγόριθμο iForestASD. Ο προτεινόμενος αλγόριθμος δοκιμάστηκε με τέσσερα σύνολα δεδομένων πραγματικού κόσμου που προέρχονται από το αποθετήριο UCI. Αν και οι συγγραφείς δεν δοκίμασαν τη συμπεριφορά του αλγορίθμου σε σχέση με άλλους αλγόριθμους ισχυρίζονται ότι ο προτεινόμενος αλγόριθμος είναι αποτελεσματικός στον εντοπισμό ανώμαλων δεδομένων ροής.

Το ίδιο έτος, οι Amer et al., προκειμένου να επιτύχουν τον εντοπισμό ανωμαλιών και να βελτιώσουν την απόδοση ενός μοντέλου SVM μιας κατηγορίας σε μη εποπτευόμενη μάθηση, εφάρμοσαν δύο τροποποιήσεις. Η βασική ιδέα και των δύο τροποποιήσεων που πρότειναν απορρέει από τη συλλογιστική ότι οι ακραίες τιμές θα πρέπει να συνεισφέρουν στη διαμόρφωση του μοντέλου εκπαίδευσης όπως και οι κανονικές περιπτώσεις. Το προτεινόμενο μοντέλο δοκιμάστηκε σε σύνολα δεδομένων από το αποθετήριο μηχανικής εκμάθησης UCI και έδειξαν ότι τα βελτιωμένα SVM μιας κατηγορίας είναι ανώτερα από τους τυπικούς αλγόριθμους ανίχνευσης ανωμαλιών χωρίς επίβλεψη.

Οι Wang et al. (2014), αναγνωρίζοντας ότι οι πλατφόρμες μέσω κοινωνικής δικτύωσης (Twitter κλπ) αποτελούν τεράστια πηγή πληροφοριών, επισημαίνουν ότι είναι σημαντικό να

επιτευχθεί ο εντοπισμός της ξαφνικής αλλαγής συναισθημάτων μέσα σε ένα μεγάλο κείμενο. Παράλληλα, αυτό αποτελεί και μια πρόκληση λόγω της ποικιλομορφίας και του μεγέθους των δεδομένων που εμφανίζονται στα μέσα κοινωνικής δικτύωσης. Στην κατεύθυνση επίλυσης του προβλήματος, πρότειναν μια μέθοδο βελτιωμένης ταξινόμησης συναισθημάτων με δυνατότητα ανίχνευσης ανωμαλιών που, σύμφωνα με τους ερευνητές, εμφάνισε ικανοποιητικά αποτελέσματα.

Και οι Erfani et al. (2016) εστίασαν στο πρόβλημα του εντοπισμού ανωμαλιών σε πολυδιάστατα δεδομένα. Όπως, τονίζουν, στις περιπτώσεις αυτές εμφανίζονται χαρακτηριστικά που ο συνυπολογισμός τους στην ανάλυση δεν επιτρέπει τον εντοπισμό των ανωμαλιών στα δεδομένα. Η άποψη των ερευνητών είναι ότι τα μοντέλα SVM είναι, συχνά αναποτελεσματικά στην ανάλυση δεδομένων μεγάλου όγκου και πολλαπλών χαρακτηριστικών. Οι ερευνητές πρότειναν ένα υβριδικό μοντέλο το οποίο αποτελείται από ένα μη εποπτευόμενο δίκτυο βαθιάς μάθησης (DBN) που εκπαιδεύεται να εξάγει γενικά υποκείμενα χαρακτηριστικά και ένα SVM μιας κατηγορίας που εκπαιδεύεται με βάση τα χαρακτηριστικά που τροφοδοτείται από το DBN. Οι συγγραφείς ισχυρίζονται ότι το προτεινόμενο μοντέλο είναι επεκτάσιμο και υπολογιστικά αποδοτικό λόγω του ότι επιτρέπει την αντικατάσταση του γραμμικού πυρήνα από μη γραμμικούς.

Οι Kumar & Ravi (2017) πρότειναν έναν ταξινομητή εγγράφων με βάση την ανάλυση κύριων συνιστωσών (PCA) και τον αλγόριθμο one-class SVM (ocSVM). Ο ρόλος της ανάλυσης κυρίων συνιστωσών είναι να μειωθούν οι διαστάσεις ενώ ο αλγόριθμος ocSVM εκτελεί ταξινόμηση. Στη συνέχεια, ο αλγόριθμος ocSVM εκπαιδεύτηκε στις εγγραφές της αρνητικής κλάσης και δοκιμάστηκε στις εγγραφές της θετικής τάξης. Τα δεδομένα που χρησιμοποιήθηκαν αφορούσαν δημοφιλή σύνολα δεδομένων όπως το αρχείο 20NG, το αρχείο Syskill και Webert, και το αρχείο Imperial Bank από το εργαλείο μοντελοποίησης του SPSS. Οι συγγραφείς ισχυρίζονται ότι ο συνδυασμός των μοντέλων παρουσίασε μεγάλη ακρίβεια κατηγοριοποίησης σε όλα τα σύνολα δεδομένων.

Οι Gorokhov et al. (2017) αντιμετώπισαν το πρόβλημα της ανίχνευσης ανωμαλιών σε δεδομένα κειμένου χρησιμοποιώντας συνελκτικά νευρωνικά δίκτυα (CNN). Ανέπτυξαν μια συγκεκριμένη αρχιτεκτονική του CNN που αποτελείται από ένα στρώμα CNN και ένα στρώμα υποδειγματοληψίας με ακτινική συνάρτηση βάσης ενεργοποίησης (activation function

βασισμένη στη συνάρτηση radial basis function kernel) και λογαριθμική συνάρτηση απώλειας (logarithmic loss function) στο τελικό στρώμα. Το προτεινόμενο μοντέλο δοκιμάστηκε στον εντοπισμό ανωμαλιών σε μια ροή εγγράφων κειμένου με το γνωστό σύνολο δεδομένων Enron. Οι ερευνητές ισχυρίζονται ότι η προτεινόμενη μέθοδος δίνει καλύτερα αποτελέσματα σε σύγκριση με τις παραδοσιακές μεθόδους ανίχνευσης ακραίων τιμών που βασίζονται σε SVM και NMF.

Οι Najafabadi et al. (2017) εστίασαν στον εντοπισμό επιθέσεων κατανεμημένης άρνησης υπηρεσιών (DDoS). Ακολουθώντας μια προσέγγιση εξόρυξης κειμένου κατέγραψαν με bigrams τα χαρακτηριστικά που αντιπροσώπευαν την ακολουθία αιτημάτων HTTP ενός χρήστη. Με αξιοποίηση του αλγόριθμου SVM μιας κλάσης, και πραγματικά αρχεία καταγραφής διακομιστή ιστού δοκιμάστηκαν τρεις διαφορετικές παραλλαγές επιθέσεων. Τα αποτελέσματά έδειξαν ότι η προτεινόμενη μέθοδος είναι σε θέση να ανιχνεύσει επιθέσεις DDoS στο επίπεδο εφαρμογής με πολύ καλά αποτελέσματα.

Οι Khreich et al. (2017) εστιάζουν στο πρόβλημα της ανίχνευσης ανωμαλιών συστήματος σε επίπεδο κεντρικού υπολογιστή και σε χρόνο εκτέλεσης. Όπως σημειώνουν οι συγγραφείς, οι υπάρχουσες τεχνικές εμφανίζουν μεγάλα ποσοστά ψευδών συναγερμών. Προκειμένου να μειωθεί το ποσοστό των ψευδών συναγερμών προτείνουν ένα νέο σχήμα το οποίο ξεκινά από την τμηματοποίηση των ιχνών των κλήσεων του συστήματος σε πολλαπλά n-grams τα οποία, στη συνέχεια, αντιστοιχίζονται σε διανύσματα αραιών χαρακτηριστικών σταθερού μήκους. Τέλος, τα διανύσματα αυτά χρησιμοποιούνται για την εκπαίδευση ανιχνευτών ocSVM. Οι παραπάνω ερευνητές ισχυρίζονται ότι το σύστημα ανίχνευσης ανωμαλιών που χρησιμοποιεί ocSVM, με πυρήνα Gauss υπερέρχει σε σχέση με τα αντίστοιχα μοντέλα Markov μια και επιτυγχάνει υψηλότερο επίπεδο ακρίβειας ανίχνευσης (με χαμηλότερο ποσοστό ψευδών συναγερμών).

Οι Kannan et al. (2017) παρουσίασαν έναν επαναληπτικό αλγόριθμο, ονομαζόμενο TONMF, που βασίζεται στη φιλοσοφία των αλγορίθμων block coordinate descent (BCD). Ο αλγόριθμος αυτός καλείται να αντιμετωπίσει το πρόβλημα της ανίχνευσης ακραίων τιμών σε δεδομένα, όπως το κείμενο, όπου οι τιμές των χαρακτηριστικών είναι συνήθως μη αρνητικές και οι περισσότερες τιμές είναι μηδενικές. Ισχυρίζονται ότι η προσέγγισή τους έχει σημαντικά

πλεονεκτήματα σε σχέση με τις παραδοσιακές μεθόδους για τον εντοπισμό ακραίων δεδομένων κειμένου.

Οι Xu, D. et al. (2017), έχοντας σκοπό τη βελτίωση της χαμηλής ακρίβειας και της κακής απόδοσης εκτέλεσης που εμφανίζεται σε αλγόριθμους ανίχνευσης ανωμαλιών δεδομένων, πρότειναν έναν αλγόριθμο που βασίζεται στον SA-iForest και στηρίζεται στην ιδέα της επιλεκτικής ολοκλήρωσης. Η προτεινόμενη μέθοδος συγκρίθηκε με τον παραδοσιακό αλγόριθμο Isolation Forest και τον αλγόριθμο LOF μέσα από ένα τυπικό σύνολο δεδομένων ως προς την ακρίβεια και την αποτελεσματικότητα των δεδομένων επιδεικνύοντας βελτίωση των επιδόσεων σε σχέση με τους άλλους αλγόριθμους.

Οι Schmitt & Spinoso (2018) πρότειναν το συνδυασμό συνελκτικών νευρωνικών δικτύων (CNN) και αλγορίθμων ομαδοποίησης με βάση την πυκνότητα. Ο αλγόριθμος ομαδοποίησης με βάση την πυκνότητα χρησιμοποιήθηκε για τον εντοπισμό και την αφαίρεση ακραίων τιμών από τα έγγραφα που θα ταξινομηθούν. Στη συνέχεια, εκπαίδευσαν το νευρωνικό δίκτυο με τις τιμές που απέμεναν πριν χρησιμοποιηθούν αυτά τα έγγραφα για την εκπαίδευση του συνελκτικού νευρωνικού δικτύου. Με βάση τα αποτελέσματα, οι ερευνητές ισχυρίζονται ότι η αφαίρεση ακραίων στοιχείων από έγγραφα μπορεί να βελτιώσει την ακρίβεια του μοντέλου και να μειώσει τον υπολογιστικό φόρτο της εκπαίδευσης.

Οι Karczmarek et al. (2020), ισχυριζόμενοι ότι πολλά από τα σχήματα ανίχνευσης ανωμαλιών εξακολουθούν να είναι ατελή και εστιάζουν σε συγκεκριμένο τύπο δεδομένων, αναλύουν την κλασική μέθοδο Isolation Forest και την επεκτείνουν. Παράλληλα, αξιοποιούν, τον αλγόριθμο k-Means, για την πρόβλεψη του αριθμού των διαιρέσεων σε κάθε κόμβο του δέντρου απόφασης. Οι ερευνητές ισχυρίζονται ότι η προτεινόμενη μέθοδος λειτουργεί αποτελεσματικά με δεδομένα που προέρχονται από διάφορους τομείς εφαρμογής.

Οι Heigl et al. (2021), υπογραμμίζοντας την έλλειψη μιας ευέλικτης και αποτελεσματικής λύσης που ικανοποιείται από το PCB-iForest, εισήγαγαν ένα νέο μεθοδολογικό πλαίσιο ονομαζόμενο PCB-iForest, για τον εντοπισμό ανωμαλιών σε δεδομένα ροής. Συγκεκριμένα, ενσωμάτωσαν στον αλγόριθμο PCB-iForest δύο παραλλαγές με σκοπό τη βελτίωση της βαθμολόγησης των χαρακτηριστικών ανάλογα με τη συμβολή τους στην ανωμαλία του δείγματος. Για την αξιολόγηση των αλγορίθμων χρησιμοποίησαν είκοσι τρία σύνολα δεδομένων από τον πραγματικό κόσμο και διαπίστωσαν ότι ο αλγόριθμος PCB-iForest

ξεπέρασε τους ανταγωνιστές στο 61% των περιπτώσεων που ελέγχθηκαν, εμφανίζοντας παράλληλα ελάττωση του υπολογιστικού φόρτου.

2.5 Εισαγωγή στο μεθοδολογικό πλαίσιο της Θέσης

Η παρούσα θέση εστιάζει στην ανάλυση κειμένων (documents) που αποτελούν μια συλλογή (corpus). Ένα σύνθετο πρόβλημα που καλείται να αντιμετωπίσει ο ερευνητής είναι ο προσδιορισμός των θεμάτων (topics) που υποκρύπτονται στα κείμενα αυτά.

Η εν λόγω ανάλυση μπορεί να αξιοποιηθεί πολύπλευρα και να παράσχει λύσεις αναφορικά με πολλά και ενδιαφέροντα προβλήματα, όπως (Liu, 2012):

- Ο προσδιορισμός του θέματος συζήτησης σε σχέση με τα πολιτικά και κοινωνικά προβλήματα στο twitter κατά τη διάρκεια ενός χρονικού διαστήματος.
- Η εξαγωγή των επιστημονικών θεμάτων που ερευνώνται μέσα σε ένα σύνολο δημοσιευμένων επιστημονικών εργασιών.
- Η εύρεση των προβλημάτων αλλά και των πλεονεκτημάτων που αναδεικνύουν χρήστες για προϊόντα και υπηρεσίες, μέσω των αναρτήσεων τους στα μέσα κοινωνικής δικτύωσης.

Ο προσδιορισμός του θέματος ενός κειμένου ή μιας συζήτησης αποτελεί ένα πολυσύνθετο πρόβλημα για τη μηχανική μάθηση που δεν επιδέχεται γενική λύση. Έτσι, η επιλογή της προσέγγισης (εποπτευόμενη/μη εποπτευόμενη μάθηση, αλγόριθμος κλπ), σε μεγάλο βαθμό επιλέγεται κατά περίπτωση εξαρτώμενη από τα διαθέσιμα δεδομένα.

Οι αλγόριθμοι που χρησιμοποιούνται για την αντιμετώπιση του εν λόγω προβλήματος διακρίνονται σε:

- Εποπτευόμενης μάθησης
- Μη εποπτευόμενης μάθησης
- Ημιοπτευόμενης μάθησης

Η παρούσα Θέση επικεντρώνεται στις προσεγγίσεις και τους αλγόριθμους που αφορούν την μη εποπτευόμενη μάθηση.

Γενικά, μπορεί να πει κανείς ότι οι αλγόριθμοι αυτοί απαιτούν κάποια είσοδο (input) και παράγουν μια έξοδο.

Στην περίπτωση που εξετάζουμε, η είσοδος των αλγορίθμων είναι κατά βάση:

- Μια συλλογή (corpus) από N κείμενα
- Το εκτιμώμενο πλήθος των υποκρυπτόμενων θεμάτων k
- Ο κατά προσέγγιση αριθμός των χαρακτηριστικών (features), ή αλλιώς λέξεων κλειδιά (keywords) f , που συμβάλλουν στο σχηματισμό του εκάστοτε θέματος.

Συχνά, το επιθυμητό εξαγόμενο (output) των αλγορίθμων αυτών αφορά:

- Τον προσδιορισμό των ενυπαρχόντων θεμάτων στα κείμενα αυτά καθώς και
- Το βαθμό κάλυψης του κάθε θέματος από ένα κείμενο της συλλογής.

Έστω ότι ο βαθμός κάλυψης συμβολίζεται ως π_{ij} , όπου ο δείκτης i αναφέρεται στο αναλυόμενο κείμενο της συλλογής και ο δείκτης j αφορά το θέμα. Ο βαθμός κάλυψης π_{ij} του θέματος j από ένα κείμενο i μπορεί να ερμηνευθεί ως η πιθανότητα το θέμα j να καλύπτεται από το κείμενο i . Υπό την προϋπόθεση ότι πέραν των θεωρούμενων k θεμάτων δεν υπάρχουν άλλα θέματα στη συλλογή, το άθροισμα των βαθμών κάλυψης των θεμάτων ανά κείμενο οφείλει να ισούται με 1 (100%). (Πίνακας 1)

ΠΙΝΑΚΑΣ 1 ΤΟ ΕΠΙΘΥΜΗΤΟ ΕΞΑΓΟΜΕΝΟ ΤΗΣ ΑΝΑΛΥΣΗΣ ΘΕΜΑΤΩΝ ΣΕ ΚΕΙΜΕΝΑ

	Θέμα 1	Θέμα 2	...	Θέμα K	Άθροισμα
Κείμενο 1	$\pi_{11} = 30\%$	1%	...	15%	1
Κείμενο 2	$\pi_{21} = 40\%$	15%	...	5%	1
Κείμενο N	$\pi_{N1} = 0\%$	30%	...	20%	1

2.6 Το μοντέλο Latent Dirichlet Allocation (LDA)

Όπως αναφέραμε παραπάνω, ένα σημαντικό πρόβλημα που εμφανίζεται στην ανάλυση κειμένου είναι αυτό που αφορά τον προσδιορισμό των θεμάτων (topics), τα οποία υποκρύπτονται σε μια συλλογή κειμένων (corpus). Η ανάλυση αυτή μπορεί να χρησιμοποιηθεί σε πολλές περιπτώσεις όπως είναι η κατηγοριοποίηση των κειμένων, η δημιουργία περιλήψεων, η αναζήτηση κειμένων που είναι συναφή ή ακόμα και ο προσδιορισμός πρωτοτυπίας (ChengXiang & Masung, 2016).

Ο συνηθέστερος τρόπος προσέγγισης της διαχείρισης του προβλήματος αυτού προέρχεται από το χώρο των μηχανών αναζήτησης. Σύμφωνα με αυτό το πλαίσιο ανάλυσης, τα κείμενα

που ανήκουν σε μια συλλογή αντικαθίστανται από διανύσματα πραγματικών αριθμών. Τυπική προσέγγιση αυτού του τύπου είναι το μοντέλο TF-IDF (Salton & McGill, 1983) το οποίο μετασχηματίζει κείμενα αυθαίρετου μήκους σε διανύσματα με αριθμητικές συνιστώσες, ακολουθώντας την παραδοχή του σάκου λέξεων.

Οι συνιστώσες είναι πραγματικοί αριθμοί και αντιπροσωπεύουν τη σχετική συχνότητα εμφάνισης λέξεων/όρων (terms), που εμφανίζονται εντός των κειμένων αυτών. Συγκεκριμένα, οι συνιστώσες αυτές προκύπτουν από το γινόμενο της κανονικοποιημένης συχνότητας εμφάνισης των όρων με το αντίστροφο της συχνότητας εμφάνισης των όρων στο corpus. Συνδυάζοντας όλα τα διανύσματα των κειμένων δημιουργούμε ένα πίνακα όρων/κειμένων, όπου κάθε γραμμή αντιπροσωπεύει ένα κείμενο (document) και κάθε στήλη αφορά έναν όρο. Οι τιμές που απαρτίζουν κάθε γραμμή για τους αντίστοιχους όρους, αποτελούν τη σχετική συχνότητα εμφάνισης ενός όρου στο κείμενο αυτό (Πίνακας 2). Αν και η μέθοδος TF-IDF παρουσιάζει κάποια πλεονεκτήματα δεν προσφέρει τόσο ικανοποιητική στατιστική δομική περιγραφή μεταξύ των κειμένων της συλλογής.

ΠΙΝΑΚΑΣ 2 ΣΧΕΤΙΚΕΣ ΣΥΧΝΟΤΗΤΕΣ ΕΜΦΑΝΙΣΗΣ ΟΡΩΝ ΑΝΑ ΚΕΙΜΕΝΟ (ΠΑΡΑΔΕΙΓΜΑ)

	Όρος 1 (Term_1)	Όρος 2 (Term_2)	Όρος S (Term_S)
Κείμενο 1	f_{11}	f_{12}	f_{1s}
Κείμενο 2	f_{21}	f_{22}	f_{2s}
Κείμενο N	f_{N1}	f_{N2}	f_{Ns}

Προκειμένου να θεραπευθούν τα προβλήματα που εμφάνιζε η μέθοδος TF-IDF προτάθηκαν διάφορες άλλες προσεγγίσεις που εστίαζαν στο πρόβλημα της μείωσης των διαστάσεων, όπως η μέθοδος Latent Semantic Index ή αλλιώς LSI (Deerwester, et al., 1990) και η pLSI (Hofmann, 1999). Η pLSI είναι μια πιθανοθεωρητική προσέγγιση της LSI όπου κάθε λέξη παράγεται από ένα θέμα και διαφορετικές λέξεις ενός κειμένου θεωρούνται ότι παράγονται από διαφορετικά θέματα. Τελικά, κάθε κείμενο αντιπροσωπεύεται από ένα κατάλογο. Ένα σοβαρό μειονέκτημα του μοντέλου pLSA είναι ότι δεν μπορεί να αξιοποιηθεί ως παραγωγικό μοντέλο για έγγραφα, με άλλα λόγια, δεν μπορεί να δώσει μια κατανομή για όλα τα πιθανά έγγραφα (ChengXiang & Masung, 2016). Για την άρση της αδυναμίας αυτής του μοντέλου pLSA δημιουργήθηκε το μοντέλο Latent Dirichlet Allocation (LDA) .

Η μέθοδος Latent Dirichlet Allocation (σύντομα LDA) προτάθηκε από τους Blei et al. (2003) και προσφέρει βελτιώσεις σε πολλές δυσλειτουργίες των προηγούμενων μεθόδων. Έχει χρησιμοποιηθεί σε πλείστες περιπτώσεις από ερευνητές στο χώρο της μηχανικής μάθησης και θεωρείται ιδιαίτερα αξιόπιστη μέθοδος. Να σημειώσουμε εδώ ότι όλες οι παραπάνω μέθοδοι βασίζονται στη υπόθεση του σάκου λέξεων, με βάση την οποία αμελείται η σειρά διάταξης των λέξεων και η μεταξύ τους σχέση.

Επειδή η εν λόγω μεθοδολογία αξιοποιείται σε σημαντικό βαθμό στην παρούσα θέση θεωρούμε σκόπιμο να γίνει μια περιγραφή της LDA. Η LDA είναι μια μεθοδολογία που αποσκοπεί στην εξεύρεση των θεμάτων που υποκρύπτονται σε ένα κείμενο και υιοθετεί την κατανομή Dirichlet (Blei, et al., 2003). Το μοντέλο Dirichlet αξιοποιείται για να καταγράψει το μορφότυπο των λέξεων που συναντώνται συχνά μαζί, καθώς και εκείνων που είναι συναφείς. Η LDA βασίζεται στις υποθέσεις ότι υπάρχουν διάφορα θέματα σε ένα κείμενο και ότι κάθε θέμα αποτελεί μίγμα λέξεων. Η μέθοδος αυτή:

- Χρησιμοποιεί κατανομές πιθανότητας.
- Προσδιορίζει την πυκνότητα πιθανότητας ανά θέμα για κάθε έγγραφο.
- Προσδιορίζει την πυκνότητα πιθανότητας των λέξεων ανά θέμα.
- Αντιστοιχίζει στα έγγραφα τα θέματα που ενυπάρχουν σε αυτά.

Βασικά στοιχεία της LDA είναι:

- Το πλήθος των εγγράφων της συλλογής.
- Το πλήθος των λέξεων ανά έγγραφο.
- Το υποκρυπτόμενο θέμα z σε μια λέξη.
- Η κατανομή θεμάτων θ και οι παράμετροι α και β .

Η LDA, στοχεύοντας να προσδιορίσει τη βέλτιστη λύση, ξεκινά από το επίπεδο του εγγράφου και με επαναληπτικό τρόπο προσδιορίζει τα θέματα που θα μπορούσαν να δημιουργηθούν από τα έγγραφα της συλλογής. Εντοπίζει τις λέξεις που θα δημιουργούσαν κάθε θέμα και καταλήγει στον πίνακα των θεμάτων των εγγράφων (Document-Topic matrix) και στον πίνακα των λέξεων ανά θέμα (Topic-Word matrix).

Συγκεκριμένα, στην LDA, η κατανομή κάλυψης θέματος ανά έγγραφο είναι μια πολυωνυμική κατανομή που υποτίθεται ότι αντλείται από κατανομή Dirichlet, με αποτέλεσμα τη

δημιουργία ενός διανύσματος πιθανοτήτων των θεμάτων. Με τον ίδιο τρόπο, όλες οι κατανομές λέξεων που αναπαριστούν τα υποβόσκοντα θέματα στη συλλογή των εγγράφων θεωρούνται ότι επιλέγονται μέσα από μια άλλη κατανομή Dirichlet. Έτσι, στη μέθοδο LDA εμφανίζονται παράμετροι που προσδιορίζουν αυτές τις δύο κατανομές Dirichlet και, κατ' επέκταση, ορίζουν και τη συμπεριφορά όλου του μοντέλου. Πιο αναλυτικά, η παράμετρος α καθορίζει τις κατανομές πιθανότητας των θεμάτων ανά έγγραφο και η παράμετρος β διαμορφώνει τις κατανομές των λέξεων ανά θέμα.

2.7 Η αξιοποίηση των λεξικών στην ανάλυση υποκειμενικής γνώμης

Κατά την ανάλυση υποκειμενικής γνώμης σε συλλογές κειμένων, σημαντικό ρόλο διαδραματίζουν οι λεγόμενες λέξεις-δείκτες ή αλλιώς, λέξεις-άποψης (opinion words) (Liu, 2015).

Όπως σημειώνει ο Liu (2010), οι λέξεις-άποψης:

- Εκφράζουν θετική άποψη (πχ. θαυμάσιος, εξαιρετικός, ονειρεμένος) ή αρνητική άποψη (πχ. φτωχός, δυστυχής) πάνω σε ένα θέμα.
- Συνδυάζονται μεταξύ τους προκειμένου να δημιουργήσουν εκφράσεις με περιεχόμενο:
 - Αντικειμενικό. Σύμφωνα με αυτή, η λέξη/φράση έχει πάντα το ίδιο εννοιολογικό περιεχόμενο ανεξάρτητα από τα συμφραζόμενα. Παράδειγμα η λέξη καταστροφή έχει πάντα αρνητικό περιεχόμενο.
 - Υποκειμενικό. Σύμφωνα με αυτή, η λέξη/φράση μπορεί να έχει μεταβαλλόμενο εννοιολογικό περιεχόμενο ανάλογα με τα υπόλοιπα συμφραζόμενα. Για παράδειγμα, η λέξη "αύξηση" μπορεί να έχει θετικό περιεχόμενο στη φράση «Το τελευταίο εξάμηνο σημειώθηκε αύξηση του ΑΕΠ της χώρας» ή αρνητικό περιεχόμενο αν εξετάζεται η φράση «οι τιμές των καταναλωτικών προϊόντων σημείωσαν μεγάλη αύξηση την τελευταία εβδομάδα».

Είναι φανερό ότι οι λέξεις άποψης είναι κεφαλαιώδους σημασίας στην ανάλυση υποκειμενικής γνώμης και η κωδικοποίηση τους θα συνέβαλε τα μέγιστα στην εξασφάλιση ποιοτικών αποτελεσμάτων ανάλυσης. Η συγκέντρωση των εν λόγω λέξεων και φράσεων σε ένα ενιαίο σύνολο, οδηγεί στη δημιουργία ενός Λεξικού συναισθήματος (sentiment lexicon) ή Λεξικού άποψης/ γνώμης (opinion lexicon). Έχουν αναπτυχθεί διάφορες προσεγγίσεις

δημιουργίας λεξικών συναισθήματος/ άποψης, οι πιο σημαντικές εκ των οποίων αναφέρονται παρακάτω.

Τα κυριότερα λεξικά άποψης με χρονολογική σειρά είναι:

- Το λεξικό **SentiWordNet** (Esuli & Sebastiani, 2006). Βασίζεται στο λεξικό WordNet. Αξιοποιεί τη διεπαφή των συνόλων συνωνύμων (synsets). Η διεπαφή αυτή σαρώνει το λεξικό WordNet και δημιουργεί ομάδες συνώνυμων λέξεων, δηλαδή λέξεων που εκφράζουν το ίδιο νόημα. Είναι φανερό ότι ενώ κάποιες λέξεις μπορεί να ανήκουν σε ένα μόνο σύνολο συνωνύμων άλλες μπορούν να ανήκουν σε περισσότερα από ένα σύνολα συνωνύμων. Σε κάθε λέξη αποδίδεται θετική, αρνητική και αντικειμενική επίδοση που ανήκει στο διάστημα [0, 1]. Το λεξικό καλύπτει πάνω από 150 χιλ. λέξεις. Λόγω, του γεγονότος ότι παρέχεται και θετική και αρνητική εκτίμηση για μια λέξη απαιτείται να είναι γνωστό το περιεχόμενο της λέξης μέσα στη φράση.
- Το λεξικό με όνομα **SO-CAL** (Taboada, et al., 2011). Αναπτύχθηκε με σκοπό να εξάγει άποψη από κείμενο. Χρησιμοποιεί ως δομική μονάδα τη λέξη. Οι λέξεις προσδιορίζονται με τον σημασιολογικό τους προσανατολισμό, όπως αυτός εκφράζεται από το συνδυασμό της πολικότητας και της έντασης. Συγκεκριμένα, κάθε λέξη αντιστοιχίζεται με έναν ακέραιο αριθμό που ανήκει στο διάστημα [-5, +5].
- Το λεξικό **Sentiment Treebank** (Socker, et al., 2013). Το λεξικό αυτό παρέχει εξαιρετικά προσδιορισμένες ετικέτες για 215.154 φράσεις. Οι ετικέτες των φράσεων παίρνουν τιμές στο πεντάβαθμο διάστημα [πολύ αρνητικό, αρνητικό, ουδέτερο, θετικό, πολύ θετικό].
- Το λεξικό **Vader** (Hutto & Gilbert, 2014). Το λεξικό αυτό χρησιμοποιείται σε προβλήματα ανάλυσης υποκειμενικής γνώμης που βασίζεται σε κανόνες και είναι ειδικά προσαρμοσμένο να χειρίζεται απόψεις που εκφράζονται σε μέσα κοινωνικής δικτύωσης ενώ συμπεριφέρεται καλά και σε κείμενα από άλλες πηγές προέλευσης.

Περιληπτικές πληροφορίες των κυριότερων λεξικών παρατίθενται στους πίνακες 3 και 4.

ΠΙΝΑΚΑΣ 3 ΤΑ ΛΕΞΙΚΑ ΕΞΟΡΥΞΗΣ ΓΝΩΜΗΣ ΕΝ ΣΥΝΤΟΜΙΑ

Λεξικό	Προσέγγιση	Λεξική Μονάδα	Ετικέτα	Σχόλια
SO-CAL	Χειρωνακτικά	Λεξικό δίκτυο	Ακέραιος [-5, +5]	Δυνατή η βελτίωση μέσω χρήσης γλωσσικών στοιχείων.

SentiWordNet	Αυτόματα	Λεξικό δίκτυο, Δίκτυο γνώμης	3 κλασματικές αξίες, Θετικό, Αρνητικό, Αντικειμενικό [0, 1]	Εντοπίζει διάφορα συναισθήματα, τα οποία έχουν διαφορετική πόλωση.
Sentiment Treebank	Χειρωνακτικά, Πληθοπορισμός	Φράση	5 ετικέτες από «πολύ αρνητικό» έως «πολύ θετικό»	Χρειάζεται προετοιμασία ανάλογα το έργο.
Macquaire Semantic Orientation Lexicon	Ημιοποπτευόμενη	Λέξη	Θετικό / Αρνητικό	Χρησιμοποιεί συνδέσμους προς λεξικό θησαυρό για να ανακαλύψει νέες λέξεις.

ΠΙΝΑΚΑΣ 4 ΤΑ ΛΕΞΙΚΑ ΣΥΝΑΙΣΘΗΜΑΤΩΝ ΕΝ ΣΥΝΤΟΜΙΑ

Λεξικό	Προσέγγιση	Ετικέτα	Σχόλια
LIWC	Χειρωνακτικά	Ιεραρχία κατηγοριών	Απόφαση ιεράρχησης κατηγοριών. Αλληλοεπιδρώντες κριτές.
ANEW	Χειρωνακτικά	Σθένος, εξέγερση, επικράτηση.	Λίστες ScanSAM.
EmoLex	Χειρωνακτικά	8 συναισθήματα, 2 κατηγορίες σθένους	Πληθοπορισμός. Επικεντρώνεται στον ποιοτικό έλεγχο.
WordNet Affect	Ημιοποπτευόμενη	Συναισθηματικές ετικέτες.	Σχολιασμός. Επέκταση με συσχετίσεις του WordNet.
Chinese Emotion Lexicon	Ημιοποπτευόμενη	5 συναισθήματα	Σχολιασμός. Επέκταση με μήτρες ομοιότητας.
NRC Hashtag Emotion Lexicon	Αυτόματα	8 συναισθήματα	Χρησιμοποιεί εποπτεία βάσει των hashtags.
SenticNet 4	Ημιοποπτευόμενη	Εκτεταμένη δομή.	Ημιοποπτευόμενη γραφική δομή, φτιαγμένη με τεχνικές όπως ιεραρχική ομαδοποίησης.

Όπως έχει διαφανεί ήδη σε προηγούμενα σημεία της Θέσης αυτής, ένα μεγάλο μέρος τεχνικών ανάλυσης συναισθήματος βασίζεται στην αξιοποίηση κάποιου Λεξικού. Η δημιουργία ενός Λεξικού είναι επίπονη και χρονοβόρα διαδικασία, που έχει σκοπό τη δημιουργία ενός συνόλου όρων με αντιστοιχίσεις σε συναισθήματα και ένταση συναισθημάτων. Η αξιοποίηση του λεξικού γίνεται μέσω αλγορίθμων προκειμένου να αντιμετωπίσουμε μεγάλο όγκο δεδομένων και να εξαχθεί πληροφορία αναφορικά με:

- Συγκεκριμένες έννοιες-στόχους (entity): Για παράδειγμα κάποιο προϊόν, πρόσωπο, ή εταιρία.
- Διαβαθμίσεις προτίμησης: Σύμφωνα με αυτές, επιδιώκεται να καταγραφεί αν ο εκφράζων την άποψη δηλώνει ότι είναι ευχαριστημένος, δυσαρεστημένος ή αδιάφορος ως προς την έννοια και σε ποιο βαθμό.

Μέσω αυτής της προσέγγισης, καθίσταται εφικτή η αυτοματοποιημένη ανάλυση ενός κειμένου με σκοπό τον προσδιορισμό των συναισθημάτων που αυτό εκφράζει.

Η πολυπλοκότητα της αντιμετώπισης του προβλήματος είναι μεγάλη και έγκειται στο γεγονός ότι:

- Μια έννοια ή μια συναισθηματική κατάσταση μπορεί να εκφράζεται με πολλές διαφορετικές λέξεις (Babbie, 2014).
- Μια λέξη μπορεί να εκφράζει διαφορετικά συναισθήματα. (Mohammad & Turney, 2013)
- Στην γραπτή ή την προφορική έκφραση μπορεί να εμφανίζεται η αποτίμηση μιας έννοιας με ειρωνική διάθεση και να ανατρέπει την διαβάθμιση προτίμησης.
- Η ένταση στη διαβάθμιση της προτίμησης είναι υποκειμενική. Για παράδειγμα, το «αρκετά καλά» αλλιώς βιώνεται σε ένα υποκείμενο Α και αλλιώς σε ένα υποκείμενο Β.
- Ο προσδιορισμός της άποψης του υποκειμένου εξαρτάται από τα πιστεύω του, τα συναισθήματα του και τον τρόπο με τον οποίο αυτό έχει διαμορφωθεί από την κοινωνία (Babbie, 2014), (Wiebe, 1994).

Τα μέγιστα στην κατεύθυνση της αυτοματοποίησης της ανάλυσης θα προσέφερε η δημιουργία και αξιοποίηση ενός Λεξικού όρων συναισθημάτων (Mohammad & Turney, 2013). Όπως έχουν αναφέρει οι τελευταίοι, δεν διατίθεται Λεξικό υψηλής ποιότητας και ευρύτητας κάλυψης. Αντίθετα, υπάρχουν περιορισμένα Λεξικά για ορισμένες γλώσσες όπως είναι για παράδειγμα, το WordNet Affect Lexicon (σύντομα, WAL) (Strapparava & Valitutti, 2004) και το Affective Norms for English Words (σύντομα, ANEW) (Bradley & Lang, 1999).

Οι Mohammad & Turney (2013) αποδεχόμενοι την προσέγγιση του Plutchnik αναφορικά με τα οχτώ πρωτογενή συναισθήματα του ανθρώπου:

- Χαράς
- Λύπης
- Φόβου
- Εμπιστοσύνης
- Αποστροφής
- Έκπληξης
- Προσμονής

(Plutchik, 1980) προσπάθησαν να δημιουργήσουν ένα Λεξικό που φέρει τον τίτλο EmoLex, βασίζεται στα παραπάνω συναισθήματα, αποτυπώνει δύο εντυπώσεις (sentiments) θετική και αρνητική, ενώ ταυτόχρονα περιλαμβάνει ακολουθίες μιας (unigram) ή πολλαπλών λέξεων (n-gram). Οι ακολουθίες αυτές δομούνται με τα συνηθέστερα ουσιαστικά, ρήματα, επίθετα και επιρρήματα σχετικά με τα παραπάνω συναισθήματα. Ο σκοπός του Λεξικού αυτού ήταν, σύμφωνα με τους εν λόγω ερευνητές, να δώσει απαντήσεις σε ερωτήσεις όπως:

1. Τα συναισθήματα εκφράζονται συχνότερα με ουσιαστικά, ρήματα, επίθετα ή επιρρήματα;
2. Πόσο συμφωνούν οι άνθρωποι με την αντιστοίχιση ενός συναισθήματος με μια συγκεκριμένη λέξη;
3. Υπάρχει συσχέτιση μεταξύ της πολικότητας μιας λέξης και του συναισθήματος που συνδέεται με αυτή;
4. Ποια συναισθήματα τείνουν να εμφανίζονται μαζί;

Προκειμένου να τελεσφορήσει το εγχείρημα τους αξιοποίησαν την συνεργατική πλατφόρμα Mechanical Turk της Amazon και αποδόμησαν το σύνολο της προσπάθειας σε μικρά ανεξάρτητα τμήματα που υλοποιήθηκαν συνεργατικά.

Το Λεξικό EmoLex αξιοποιείται σε διάφορες θέσεις της παρούσας διατριβής λόγω των πλεονεκτημάτων του σε σχέση με άλλα Λεξικά.

2.8 Γενικά περί μέτρησης της απόστασης στη μηχανική μάθηση

Η μέτρηση της απόστασης εμφανίζεται σε πολλά είδη ανάλυσης της μηχανικής μάθησης όπως είναι η συσταδοποίηση, η κατηγοριοποίηση μέσα από την εύρεση των πλησιέστερων γειτόνων, η ανάλυση κυρίων συνιστωσών κλπ. Βασικά, μέσω της μέτρησης της απόστασης που εμφανίζεται στα δεδομένα μας, διαμορφώνουμε ένα μέτρο ομοιότητας (similarity) μεταξύ τους. Η γενική άποψη είναι ότι όσο αυξάνεται η απόσταση μεταξύ δύο εγγραφών (records) των δεδομένων μας τόσο η ομοιότητα μεταξύ τους μειώνεται. Τις τελευταίες

δεκαετίες η έννοια της μέτρησης της απόστασης/ομοιότητας των εγγραφών, απασχολεί πολλούς επιστήμονες και βρίσκει εφαρμογές σε πολλά επιστημονικά πεδία όπως είναι η ιατρική, η πληροφορική, η ασφάλεια πληροφοριακών συστημάτων, η εξόρυξη δεδομένων μέσα από κοινωνικά δίκτυα, η αναγνώριση ομιλίας, η ανάκληση πληροφορίας (information retrieval), τα συμβουλευτικά συστήματα (recommender systems) και η περιοχή της ψηφιακής όρασης.

Όπως αναφέρθηκε παραπάνω, η εισαγωγή της μέτρησης της απόστασης μεταξύ δύο εγγραφών μας επιβάλλεται από την ανάγκη να ποσοτικοποιήσουμε την ομοιότητα μεταξύ των εγγραφών αυτών. Όμως, η μέτρηση της απόστασης είναι δυνατό να γίνει μέσα από ένα πολύ μεγάλο πλήθος προσεγγίσεων. Ως εκ τούτου, η επιλογή της κατάλληλης μετρικής απόστασης είναι μια διαδικασία που, σε μεγάλο βαθμό, εξαρτάται από το είδος και τη μορφή των προς ανάλυση δεδομένων (choice data driven). Ο λόγος που ισχύει το παραπάνω είναι ότι έχοντας ανάγκη να ανακαλύψουμε κρυμμένες σχέσεις ή πρότυπα μέσα στα αντικείμενα του συνόλου των δεδομένων μας, υιοθετούμε κάποια μορφή απόστασης που μπορεί να είναι σε θέση να αναδείξει τη σχέση αυτή (πχ. η απόσταση Mahalanobis) ενώ μια άλλη επιλογή να μην ήταν αποδοτική (πχ. η Ευκλείδεια απόσταση). Σε επίπεδο θεωρίας, μια μετρική για να μπορεί να θεωρηθεί ότι εκφράζει μετρική απόσταση θα πρέπει να ικανοποιεί ορισμένες ιδιότητες που περιέχονται στον ορισμό που ακολουθεί.

Ορισμός: Έστω ένα σύνολο $X \neq \emptyset$. Μια απεικόνιση D του καρτεσιανού γινομένου του συνόλου X στο σύνολο των πραγματικών (συμβολικά, $D: X \times X \rightarrow R$) ονομάζεται μετρική απόσταση όταν ικανοποιεί τις παρακάτω ιδιότητες:

1. $D(x, y) = 0 \Leftrightarrow x = y, \forall x, y \in X$
2. $D(x, y) = D(y, x), \forall x, y \in X$
3. $D(x, y) \leq D(x, z) + D(y, z), \forall x, y, z \in X$

Ανάλογα με τη μορφή των δεδομένων, συνήθως χρησιμοποιούνται τα παρακάτω είδη απόστασης:

- Ευκλείδεια απόσταση: Αποτελεί την παλαιότερη και πιο δοκιμασμένη μέθοδο υπολογισμού της απόστασης μεταξύ δύο σημείων στον n -διάστατο χώρο. Σύμφωνα με τη μετρική αυτή, η απόσταση ενός σημείου $\vec{p} = (p_1, p_2, \dots, p_n)$ από ένα σημείο $\vec{q} = (q_1, q_2, \dots, q_n)$ στον n -διάστατο χώρο υπολογίζεται με βάση τον τύπο:

$$D_E(p, q) = \sum_{i=1}^n (p_i - q_i)^2$$

Η απόσταση αυτή αξιοποιείται όταν οι συνιστώσες των διανυσμάτων \vec{p} και \vec{q} είναι ποσοτικές μεταβλητές.

- Απόσταση Manhattan: Η απόσταση αυτή είναι γνωστή και ως city block distance ή L_1 norm διανύσματος. Η τιμή της προσδιορίζεται από το άθροισμα των απολύτων τιμών των διαφορών μεταξύ των συνιστωσών των δύο διανυσμάτων. Σύμφωνα με τη μετρική αυτή, η απόσταση ενός σημείου $\vec{p} = (p_1, p_2, \dots, p_n)$ από ένα σημείο $\vec{q} = (q_1, q_2, \dots, q_n)$ στον n -διάστατο χώρο με βάση τον τύπο

$$D_{L1} = \sum_{i=1}^n |p_i - q_i|$$

όπου p_i και q_i είναι ποσοτικές μεταβλητές.

- Απόσταση Chebyshev: Ορίζεται ως η μέγιστη των απολύτων τιμών των διαφορών των συνιστωσών δύο διανυσμάτων. Συμβολικά,

$$D_{L1} = \max(|p_i - q_i|), i \in \{1, \dots, n\}$$

- Απόσταση Hamming: Χρησιμοποιείται σε κατηγορικές μεταβλητές. Αν οι μεταβλητές αυτές περιέχουν ίδια τιμή τότε η απόσταση Hamming παίρνει την τιμή 0, αλλιώς 1.
- Απόσταση (ή ομοιότητα) Συνημιτόνου: Η μετρική αυτή παίρνει τιμές από μηδέν έως μονάδα. Στην περίπτωση που η τιμή της απόστασης συνημιτόνου μεταξύ δύο διανυσμάτων είναι μηδέν τότε τα διανύσματα θεωρούνται ξένα (όποτε καλούνται ορθογώνια), ενώ όταν παίρνει την τιμή της μονάδας θεωρούνται όμοια. Η απόσταση συνημιτόνου σχετίζεται με τη γωνία που σχηματίζουν μεταξύ τους δύο διανύσματα \vec{p} και \vec{q} του n -διάστατου χώρου. Σε επίπεδο τύπου, ο υπολογισμός γίνεται με την παρακάτω σχέση:

$$S_{\cosine} = \frac{\vec{p} \cdot \vec{q}}{|\vec{p}| \cdot |\vec{q}|} = \frac{\sum_{i=1}^n p_i \cdot q_i}{\sqrt{\sum_{i=1}^n p_i^2} \cdot \sqrt{\sum_{i=1}^n q_i^2}}$$

- Απόσταση Jaccard ($D_{Jaccard}$). Σχετίζεται με την ομοιότητα Jaccard ($S_{Jaccard}$) μέσω του τύπου:

$$D_{Jaccard} = 1 - S_{Jaccard}$$

Η ομοιότητα Jaccard χρησιμοποιείται στην περίπτωση που θέλουμε να υπολογίσουμε την ομοιότητα μεταξύ δύο συνόλων A, B. Υπολογίζεται μέσω του παρακάτω τύπου:

$$S_{Jaccard} = \frac{|A \cap B|}{|A \cup B|}$$

- Απόσταση Mahalanobis. Περιγράφεται εκτενώς στην παράγραφο 2.8.1.
- Απόσταση Hellinger. Ικανοποιεί τις βασικές ιδιότητες που οφείλει να ικανοποιεί μια μετρική απόστασης, δηλαδή:
 - $H(x,y) \geq 0$
 - $H(x,y) = 0 \leftrightarrow x = y$
 - $H(x,y) = H(y,x)$
 - $H(x,z) \leq H(x,y) + H(y,z)$

2.8.1 Η απόσταση Mahalanobis

Μια συνηθισμένη τακτική για να αναπαραστήσουμε ένα κείμενο είναι αυτή της διανυσματικής απεικόνισης. Σύμφωνα με την προσέγγιση αυτή, κάθε κείμενο $D_j \in C$ μπορεί να αναπαρασταθεί ως ένα διάνυσμα στον n-διάστατο χώρο, όπου οι διαστάσεις αφορούν τις λέξεις που εμφανίζονται στο λεξικό. Η συνιστώσα $i \in \{1,2, \dots, n\}$ του διανύσματος που αναπαριστά το κείμενο D_j συνήθως εκφράζει τη σταθμισμένη εμφάνιση της λέξης $w_i \in W = \{w_1, w_2, \dots, w_n\}$ στο εν λόγω κείμενο. Με τον τρόπο αυτό, κάθε κείμενο απεικονίζεται από ένα σημείο του n-διάστατου χώρου και, πολύ συχνά, εμφανίζεται η ανάγκη να μετρηθεί με κάποιο τρόπο η απόσταση μεταξύ διάφορων τέτοιων σημείων.

Η Ευκλείδεια απόσταση είναι χρήσιμη σε περιπτώσεις που οι συνιστώσες των σημείων είναι ανεξάρτητες μεταξύ τους και έχουν ίσα βάρη στάθμισης. Δυστυχώς, όταν έχουμε ένα σύνολο οντοτήτων που περιγράφονται από πολλές μεταβλητές / χαρακτηριστικά, τότε είναι συχνό να μην ικανοποιούνται οι προηγούμενες δύο απαιτήσεις με αποτέλεσμα η χρήση της Ευκλείδειας απόστασης να μην οδηγεί σε ορθά αποτελέσματα και, ως εκ τούτου, να επιβάλλεται η αξιοποίηση μιας άλλης μετρικής, όπως είναι αυτή της απόστασης Mahalanobis (Mahalanobis, 1936). Η συγκεκριμένη προσέγγιση αφορά μια πολυμετάβλητη μετρική

σύμφωνα με την οποία υπολογίζεται η απόσταση ενός σημείου από τον μέσο της κατανομής του δείγματος, μέσω της αξιοποίησης των επιμέρους τυπικών αποκλίσεων των μεταβλητών.

Έστω ότι το διάνυσμα:

- $\vec{p} = (p_1, p_2, \dots, p_n)$ του οποίου το πέρας ορίζεται από το σημείο p στον πολυδιάστατο χώρο και το οποίο ανήκει σε ένα σύνολο N σημείων.
- $\vec{\mu} = (\mu_1, \mu_2, \dots, \mu_n)$ του οποίου το πέρας ορίζεται από το σημείο μ στον πολυδιάστατο χώρο και το οποίο έχει ως συνιστώσες τις μέσες τιμές των n μεταβλητών του δείγματος.
- $\vec{\sigma} = (\sigma_1, \sigma_2, \dots, \sigma_n)$ έχει συνιστώσες τις τυπικές αποκλίσεις των n μεταβλητών του δείγματος.

Η απόσταση Mahalanobis του σημείου p ορίζεται από τον τύπο:

$$D_m(p) = \sqrt{(\vec{p} - \vec{\mu})^T * S^{-1} * (\vec{p} - \vec{\mu})}$$

Όπου:

- Το διάνυσμα $(\vec{p} - \vec{\mu})$ έχει ως συνιστώσες τις επιμέρους διαφορές των συνιστωσών του θεωρούμενου σημείου p με τις αντίστοιχες του σημείου μ που εκφράζει τον μέσο όρο της κατανομής του δείγματος.
- Το διάνυσμα $(\vec{p} - \vec{\mu})^T$ αποτελεί το ανάστροφο διάνυσμα του $(\vec{p} - \vec{\mu})$.
- Ο πίνακας S^{-1} είναι ο αντίστροφος του πίνακα συμμεταβλητότητας S .

Ο πολλαπλασιασμός του υπόριζου με τον πίνακα S^{-1} έχει ως αποτέλεσμα:

- Οι συνιστώσες που εμφανίζουν μεγάλη συμμεταβλητότητα να πολλαπλασιάζονται με μικρό αριθμό.
- Οι συνιστώσες που εμφανίζουν μικρή συμμεταβλητότητα να πολλαπλασιάζονται με μεγάλο αριθμό.

Αξίζει να σημειωθεί ότι:

- Η τιμή της απόστασης Mahalanobis δεν εξαρτάται από τις:
 - μονάδες μέτρησης και
 - χρησιμοποιούμενες κλίμακες των διαφόρων μεταβλητών.
- Η Ευκλείδεια απόσταση αποτελεί μερική περίπτωση της απόστασης Mahalanobis.

- Η απόσταση Mahalanobis σχετίζεται με το μέτρο της Ανάλυσης Κυρίων Συνιστωσών (PCA). Συγκεκριμένα, το τετράγωνο της απόστασης Mahalanobis είναι ίσο με το άθροισμα των τετραγώνων όλων των μη μηδενικών δεικτών PCA (Brereton, 2015).
- Το άθροισμα των τετραγώνων των αποστάσεων Mahalanobis όλων των παρατηρήσεων ισούται με το γινόμενο του πλήθους των παρατηρήσεων επί το πλήθος των θεωρούμενων μεταβλητών (Brereton, 2015).

Η έννοια της απόστασης Mahalanobis αναδεικνύεται ιδιαίτερα σημαντική στο χώρο της μηχανικής μάθησης, όπου είναι σύνηθες να απαιτείται η ανάλυση οντοτήτων (entities-cases) που χαρακτηρίζονται από πολλές μεταβλητές (attributes) προκειμένου να γίνει:

- Προσδιορισμός ανώμαλων παρατηρήσεων (anomaly detection).
- Κατηγοριοποίηση παρατηρήσεων (classification) κλπ.

2.8.2 Η απόσταση Hellinger

Μια άλλη εκδοχή μέτρησης απόστασης αφορά την απόσταση Hellinger. Η προσέγγιση αυτή χρησιμοποιείται για να προσδιορίσει την απόσταση μεταξύ δυο κατανομών πιθανοτήτων.

Έστω ότι υπάρχουν δύο διακριτές κατανομές πιθανοτήτων:

$$P = \{p_1, p_2, \dots, p_n\} \text{ και } Q = \{q_1, q_2, \dots, q_n\} \text{ όπου } i \in \{1, 2, \dots, n\}$$

Όπως είναι φανερό, τα διανύσματα

$$\sqrt{P} = \{\sqrt{p_1}, \sqrt{p_2}, \dots, \sqrt{p_n}\} \text{ και } \sqrt{Q} = \{\sqrt{q_1}, \sqrt{q_2}, \dots, \sqrt{q_n}\}$$

$$\text{με } i \in \{1, 2, \dots, n\}$$

έχουν μοναδιαίο μήκος στο L_2 .

Το μέτρο της απόστασης Hellinger, στην περίπτωση αυτή, ορίζεται από τον ακόλουθο τύπο:

$$H(P, Q) = \frac{1}{\sqrt{2}} \|P - Q\|_{L_2}$$

Στην περίπτωση που έχουμε δύο συνεχείς κατανομές πιθανοτήτων P και Q που ορίζονται αντίστοιχα μέσω των συναρτήσεων πυκνότητας πιθανότητας $f(x)$ και $g(x)$, η απόσταση Hellinger, ορίζεται από τον ακόλουθο τύπο:

$$H(P, Q) = \frac{1}{\sqrt{2}} \sqrt{\int (\sqrt{f(x)} - \sqrt{g(x)})^2 dx}$$

Από τα παραπάνω απορρέει ότι η απόσταση Hellinger ικανοποιεί την παρακάτω ανισότητα:

$$0 \leq H(P, Q) \leq 1$$

Όσο η τιμή της απόστασης Hellinger τείνει στο μηδέν τόσο πιο μικρή είναι η απόσταση μεταξύ των δύο κατανομών. Η απόσταση Hellinger χρησιμοποιείται στη μηχανική μάθηση σε ποικίλες περιπτώσεις.

2.8.3 Εφαρμογή πάνω στη μέτρηση αποστάσεων

Για να γίνει κατανοητή η μέτρηση της απόστασης στην ανάλυση κειμένου παρατίθεται το εξής παράδειγμα:

Έστω ότι διαθέτουμε τη συλλογή από τα παρακάτω (προεπεξεργασμένα) κείμενα:

D1= ['bank', 'river', 'shore', 'water']

D2= ['river', 'water', 'flow', 'fast', 'tree']

D3= ['bank', 'water', 'fall', 'flow']

D4= ['bank', 'bank', 'water', 'rain', 'river']

D5= ['river', 'water', 'mud', 'tree']

D6= ['money', 'transaction', 'bank', 'finance']

D7= ['bank', 'borrow', 'money']

D8= ['bank', 'finance']

D9= ['finance', 'money', 'sell', 'bank']

D10= ['borrow', 'sell']

D11= ['bank', 'loan', 'sell']

Μέσω της LDA αναζητούμε ποια είναι τα πιθανά δύο θέματα που υποκρύπτονται στη συλλογή αυτή. Ο πίνακας 5 που ακολουθεί περιέχει τα προτεινόμενα δύο θέματα από την LDA καθώς και την πιθανότητα εμφάνισης της κάθε λέξης της συλλογής σε αυτά.

ΠΙΝΑΚΑΣ 5 ΠΙΝΑΚΑΣ ΥΠΟΛΟΓΙΣΜΟΥ ΤΗΣ ΑΠΟΣΤΑΣΗΣ HELLINGER ΜΕΤΑΞΥ ΔΥΟ ΚΑΤΑΝΟΜΩΝ ΠΙΘΑΝΟΤΗΤΩΝ

Λέξη	Topic1	Topic2	$\Delta(\text{Topic1}, \text{Topic2})^2$
bank	0,284	0,296	0,000144
water	0,172	0,065	0,011449
finance	0,000	0,220	0,0484
river	0,128	0,067	0,003721
flow	0,076	0,000	0,005776
borrow	0,067	0,050	0,000289
sell	0,063	0,060	9E-06
tree	0,082	0,046	0,001296
money	0,041	0,100	0,003481
fast	0,043	0,000	0,001849
transaction	0,000	0,056	0,003136
rain	0,044	0,000	0,001936
mud	0,000	0,040	0,0016
Απόσταση Hellinger μεταξύ των κατανομών των δύο θεμάτων			0,4077

3 Η ανάλυση συναισθήματος στο χώρο της πολιτικής

3.1 Γενικά για τον υβριδικό πόλεμο

Η ανάπτυξη των ΤΠΕ που σημειώνεται τις τελευταίες δεκαετίες δεν μπορούσε παρά να επηρεάζει σοβαρά τον τρόπο με τον οποίο ασκείται η πολιτική και ο πόλεμος σε ολόκληρο τον πλανήτη (Janczewski & Colarik, 2005). Ειδικότερα, η ανάπτυξη της τεχνολογίας προκάλεσε το μετασχηματισμό του τρόπου με τον οποίο εφαρμόζεται ο υβριδικός πόλεμος με αποτέλεσμα να προσφέρονται νέα εργαλεία για την εφαρμογή παλαιών πρακτικών, όπως είναι η διεξαγωγή προπαγάνδας ή η υποκλοπή ζωτικών δεδομένων (Janczewski & Colarik, 2005), (Cullen & Reichborn-Kjennerud, 2017). Είναι γνωστό ότι ο υβριδικός πόλεμος (Cullen & Reichborn-Kjennerud, 2017):

- Αξιοποιεί τη συγχρονισμένη χρήση πολλαπλών εργαλείων ισχύος σε σχέση με τον χώρο, το χρόνο και τον σκοπό.
- Έχει ως απώτερο στόχο να καταφέρει ο επιτιθέμενος πολλαπλά κτυπήματα στον αμυνόμενο των οποίων τα αποτελέσματα συνεργούν μεταξύ τους.
- Διαθέτει το πλεονέκτημα που διαθέτουν τα εργαλεία αυτά, δηλαδή επιτρέπουν στον επιτιθέμενο να αυξομειώνει κατά βούληση εύκολα την ένταση των ενεργειών του.

Όπως σημειώνουν οι Cullen & Reichborn-Kjennerud (2017), η δυνατότητα αυξομείωσης της έντασης των ενεργειών επιτρέπει στον επιτιθέμενο να:

- Εξασφαλίζει οικονομία πόρων.
- Διαφεύγει της προσοχής του στόχου του.
- Εκμεταλλεύεται αποτελεσματικά τα ευάλωτα σημεία του στόχου του.
- Προκαλεί τόσο γραμμικά όσο και μη γραμμικά αποτελέσματα από τις υβριδικές επιθέσεις.
- Εστιάζει σε πιθανά σημεία κατάρρευσης του αντιπάλου του και να εξαπολύει επιθέσεις που επιβουλεύονται όλο το φάσμα των ζωτικών λειτουργιών του στόχου.

Η μη γραμμικότητα των επιπτώσεων έχει να κάνει με την αδυναμία του αμυνόμενου να κατανοήσει τη σχέση μεταξύ των ενεργειών δεύτερου και τρίτου επιπέδου του επιτιθέμενου και τα προσδοκώμενα από αυτόν αποτελέσματα επί του αμυνόμενου ή του στόχου.

Σύμφωνα με την άποψη ερευνητών του πεδίου του υβριδικού πολέμου (Cullen & Reichborn-Kjennerud, 2017), (Janczewski & Colarik, 2005), (Janczewski & Colaric, 2008), τα συνήθη εργαλεία ισχύος που αξιοποιούνται στον υβριδικό πόλεμο μπορούν να κατηγοριοποιηθούν σε:

- Στρατιωτικά.
- Πολιτικά.
- Οικονομικά.
- Κοινωνικά.
- Πληροφοριακά.

Σχετικά με την ισχύ των παραπάνω εργαλείων, οι παραπάνω ερευνητές εκτιμούν ότι η υψηλότερη ένταση σε ισχύ προσφέρεται πλέον από τα εργαλεία που αξιοποιούν τις τεχνολογίες ICT. Τα εν λόγω εργαλεία, συνήθως, αποσκοπούν στην τρώση των πληροφοριακών συστημάτων ώστε να μην είναι λειτουργικά, την απόσπαση ή καταστροφή δεδομένων, ή τη διάχυση προπαγανδιστικού υλικού με σκοπό τον προσηλυτισμό, ή τον εκφοβισμό αμάχων (Janczewski & Colaric, 2008).

Είναι προφανές ότι ο υβριδικός πόλεμος διαφοροποιείται από τον συνήθη (Cullen & Reichborn-Kjennerud, 2017):

- Χρησιμοποιεί ευρύτερο πλήθος από εργαλεία και τεχνικές.
- Στοχεύει σε πολλαπλά ευάλωτα σημεία της κοινωνίας με καινοφανείς τρόπους προκαλώντας αποτελέσματα που δεν μπορούν να προβλεφθούν από τον στόχο.
- Γίνεται αντιληπτός πολύ αργότερα από τον συμβατικό και, μάλιστα, όταν ο σκοπός έχει πλέον επιτευχθεί.
- Δρα σε διεθνές επίπεδο με μεγαλύτερη κάλυψη, νομιμοφάνεια και εφευρετικότητα.

Το Ισλαμικό Κράτος του Ιράκ και του Αλ-Σαμ (Islamic State of Iraq and al-Sham, εν συντομία ISIS) αναδείχθηκε ως ένας από τους ισχυρούς παίκτες στην παγκόσμια σκακιέρα του υβριδικού πολέμου τα τελευταία χρόνια (Bloom & Daymon, 2018), (Ferrara, 2017), (Ozeren, et al., 2018). Αν και οι σκοποί και οι πρακτικές του ISIS έμοιαζαν εν γένει με αυτές της Al-Qaida (Nesser, et al., 2016), ο ISIS φάνηκε να ξεπερνά σε αποτελεσματικότητα την Al-Qaida (Carafella, et al., 2019), (Ozeren, et al., 2018) (Winter, 2015) και να αξιοποιεί στο έπακρο τις νέες τεχνολογίες (Nesser, et al., 2016), (Bloom & Daymon, 2018) καταφέροντας σημαντικά πλήγματα στους αντιπάλους του σε πολιτικό, κοινωνικό, οικονομικό και στρατιωτικό επίπεδο. Επιπρόσθετα, έχει διαπιστωθεί ότι το πλήθος των μαχητών που στρατολόγησε ο ISIS μέσω του διαδικτύου ήταν μεγαλύτερο παρά ποτέ (Nesser, et al., 2016). Αυτός είναι και ο λόγος που, μας ώθησε να αναλύσουμε την προπαγάνδα του ISIS μέσω της ανάλυσης υποκειμενικής γνώμης και της τεχνητής νοημοσύνης.

3.2 Γενικά για την προπαγάνδα

Ο R. Casey στο πρώτο κεφάλαιο του βιβλίου του με τίτλο «EM2. What is propaganda» αναφέρει ότι «η προπαγάνδα είναι μια έννοια που έχει πολλές εκφάνσεις και δεν ορίζεται εύκολα» (Casey, 1944).

Αν ανατρέχαμε σε έγκυρα λεξικά προκειμένου να προσδιορίσουμε το περιεχόμενο της έννοιας αυτής θα παρατηρούσαμε ότι, σύμφωνα με το λεξικό:

- **Cambridge**, η προπαγάνδα (propaganda) ορίζεται ως πληροφορία, ιδέες, απόψεις ή εικόνες που, συχνά, προωθούν ή δημοσιεύουν, μια οπτική ενός ζητήματος με σκοπό να επηρεάσουν την άποψη των ανθρώπων.
- **Meriam Webster**, η προπαγάνδα (propaganda) ορίζεται ως η διάδοση ιδεών, πληροφοριών ή φημών, με σκοπό την υποβοήθηση ή τη βλάβη ενός οργανισμού, ενός ατόμου ή μιας αντίθετης άποψης.

Η προπαγάνδα ορίζεται ως:

- Η διαχείριση συλλογικών στάσεων ανθρώπων μέσω του χειρισμού σημαντικών συμβόλων (Lasswell, 1927).
- Η διαρκής και σταθερή προσπάθεια να δημιουργήσει ή να μορφοποιήσει κάποιος περιστατικά σε τρόπον ώστε να επηρεάσει τις σχέσεις του κοινού αναφορικά με μια οργάνωση, μια ιδέα ή μια επιχείρηση (Bernays, 1928).
- Η σχεδιασμένη και στοχευμένη έκφραση άποψης ή τέλεση ενεργειών με στόχο τον επηρεασμό της γνώμης, ή των πράξεων άλλων ανθρώπων, ή ομάδων σε σχέση με προκαθορισμένους σκοπούς (Miller, 1939).
- Μια διαδικασία, μέσω της οποίας ο προπαγανδιστής αξιοποιεί διάφορες τεχνικές και επιχειρεί να διασφαλίσει την επιθυμητή συμπεριφορά του προπαγανδιζόμενου (Henderson, 1943).
- Η προσπάθεια να επηρεαστούν άτομα και να ελεγχθεί η συμπεριφορά τους σε σχέση με τα αποτελέσματα που θα προκληθούν σε ορισμένη στιγμή για την κοινωνία (Doob, 1948).
- Μεροληπτική επικοινωνία που χρησιμοποιείται συχνά από πολλούς (Dovring & Lasswell, 1959).
- Η στοχευμένη προσπάθεια κάποιου ατόμου ή ομάδας ατόμων να ελέγξουν ή να μεταβάλουν τη στάση άλλων ομάδων ανθρώπων με τη χρήση εργαλείων επικοινωνίας και με απώτερο στόχο να εξασφαλίσουν την προς όφελος του προπαγανδιστή αντίδραση των προπαγανδιζομένων (Qualter, 1962).
- Ένα σύνολο μεθόδων που εφαρμόζονται από μια οργανωμένη ομάδα η οποία έχει σκοπό να πετύχει μέσα από μεθόδους διαχείρισης την ενεργητική ή παθητική συμμετοχή μιας ομάδας ανθρώπων, την ψυχολογική διαχείριση τους και τη συμμετοχή τους σε έναν οργανισμό (Ellul, 1965).
- Η συστηματική προσπάθεια να μορφοποιηθούν αντιλήψεις, να χειραγωγηθούν οι γνωστικές λειτουργίες και η άμεση αντίδραση του κοινού-στόχος προς όφελος του προπαγανδιστή (Jowett & O'Donnell, 2012).
- Η ενσυνείδητη προσπάθεια να επηρεαστούν οι πεποιθήσεις μιας ομάδας ατόμων, μέσω μη λογικών και συνήθως ανήθικων τεχνικών, με σκοπό να οδηγηθούν αυτοί σε προκαθορισμένες στάσεις και επιλογές (Smith, 1989).

- Επικοινωνία που εμπεριέχει ένα μήνυμα, μια ιδέα ή μια ιδεολογία και η οποία έχει σχεδιαστεί κύρια για να εξυπηρετεί τα προσωπικά ενδιαφέροντα του ατόμου που υλοποιεί την επικοινωνία (Taylor, 1990).
- Επικοινωνία που έχει επιλεγμένη μορφή και περιεχόμενο ώστε να επιτύχει σε κάποιο ακροατήριο-στόχο ώστε να αποδεχτεί αυτό στάσεις και αντιλήψεις που έχουν προκαθοριστεί από τον προπαγανδιστή (Catey, 1997).
- Μηνύματα που προωθούνται σε μάζες ανθρώπων και έχουν επινοηθεί έτσι ώστε να δράσουν στρατηγικά προς την κατεύθυνση της δράσης των επηρεαζόμενων ατόμων προς όφελος της πηγής του μηνύματος (Parry-Giles, 2002).

Όπως διαπιστώνει κανείς, η έννοια της προπαγάνδας τείνει να έχει αρνητικό περιεχόμενο αφού, συνήθως, συνδέεται με την εσκεμμένη ελλιπή πληροφόρηση των ανθρώπων για ιδιοτελείς σκοπούς.

Η προπαγάνδα κατηγοριοποιείται με βάση το αν είναι γνωστή η πηγή της (Lieberman, 2017) σε:

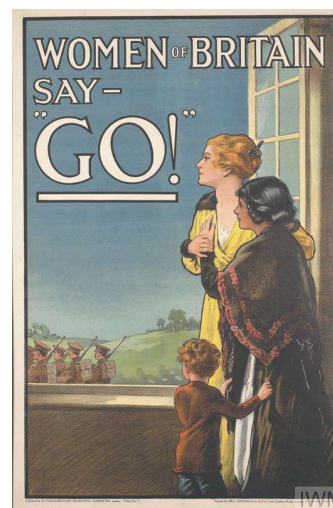
- **Λευκή προπαγάνδα:** Αφορά την προπαγάνδα που είναι φανερή η πηγή προέλευσης της και το περιεχόμενο αληθές.
- **Μαύρη προπαγάνδα:** Είναι η προπαγάνδα που δεν είναι γνωστή η πηγή προέλευσης της και, συνήθως, αποτελείται εξ ολοκλήρου από αναληθή μηνύματα.
- **Γκρίζα προπαγάνδα:** Αποτελεί μορφή προπαγάνδας με περιεχόμενο που εν μέρει είναι αληθές και εν μέρει ψευδές και η πηγή προέλευσης της δεν είναι απαραίτητα γνωστή.

Η προπαγάνδα έχει χρησιμοποιηθεί από αρχαιοτάτων ετών ως μέσο προώθησης των πολιτικών και κοινωνικών απόψεων ηγετών (Lieberman, 2017). Ας δούμε μερικά παραδείγματα:

- Στην αρχή του 1^{ου} Παγκόσμιου Πολέμου, οι Βρετανοί δεν διέθεταν έναν σοβαρό μηχανισμό πολεμικής προπαγάνδας. Διάφορες διευθύνσεις προωθούσαν τη δική τους προπαγάνδα χωρίς να υπάρχει ενιαία γραμμή μέχρις ότου ανέλαβε ο Charles Masterman (Sanders, 1975). Το 1918 ιδρύεται το Υπουργείο Πληροφοριών οπότε και γίνεται εκτεταμένη αξιοποίηση της προπαγάνδας. Τα μέσα που χρησιμοποιήθηκαν ήταν, κατά βάση, αφίσες (Πίνακας 6), λογοτεχνικά έργα και ταινίες. Ο απώτερος

στόχος της προπαγάνδας ήταν διττός. Αφενός στόχευε να πλήξει την υπόληψη των Γερμανών σε παγκόσμιο επίπεδο και αφετέρου να εξασφαλίσει την αποδοχή των θυσιών που ακολουθούν μια πολεμική αναμέτρηση από τους Βρετανούς (Sanders, 1975).

ΠΙΝΑΚΑΣ 6 ΒΡΕΤΑΝΙΚΕΣ
ΑΦΙΣΕΣ ΤΟΥ 1^{ΟΥ}
ΠΑΓΚΟΣΜΙΟΥ ΠΟΛΕΜΟΥ
ΜΕ ΣΚΟΠΟ ΝΑ
ΚΑΛΛΙΕΡΓΗΣΟΥΝ
ΚΑΤΑΛΗΛΟ ΚΛΙΜΑ ΣΤΟΥΣ
ΠΟΛΙΤΕΣ ΩΣΤΕ ΝΑ
ΚΑΤΑΤΑΓΟΥΝ ΟΙ ΑΝΤΡΕΣ
ΣΤΑ ΟΠΛΑ



Κατά τον 2^ο Παγκόσμιο Πόλεμο, οι εμπλεκόμενοι επιδόθηκαν σε μάχη «χαρακωμάτων» αναφορικά με την προπαγάνδα. Πιο συγκεκριμένα:

- Οι Ναζί στην Γερμανία, τόσο κατά τη διάρκεια του μεσοπολέμου όσο και κατά τη διάρκεια του 2^{ου} Παγκοσμίου Πολέμου, απέδιδαν μεγάλη σημασία στην προπαγάνδα. Δια στόματος του Αδόλφου Χίτλερ είχε διατυπωθεί η άποψη ότι η Γερμανία ηττήθηκε στον 1^ο Παγκόσμιο Πόλεμο επειδή η βρετανική προπαγάνδα υπερτερούσε της γερμανικής. Στο έκτο κεφάλαιο του βιβλίου «Ο Αγών μου» ο Χίτλερ ισχυρίζεται ότι «Ο σκοπός της προπαγάνδας δεν είναι να προσφέρει επιχειρήματα σε μια διαμάχη δικαιωμάτων αλλά, ξεκάθαρα, να δώσει έμφαση στο δικαίωμα που εμείς επιθυμούμε. Η προπαγάνδα δεν πρέπει να ερευνά για την εξεύρεση της αντικειμενικής αλήθειας αλλά να υποστηρίζει τη δική μας άποψη» (Hitler, 1925). Στο ίδιο κεφάλαιο, ο Hitler τονίζει ότι «Η προπαγάνδα οφείλει να απευθύνεται σε ευρείες μάζες ανθρώπων ... με μορφή προσαρμοσμένη στο επίπεδο ευφυΐας τους. ... Η τέχνη της προπαγάνδας είναι να ενεργοποιεί τη φαντασία του κοινού και τα συναισθήματα του και να αιχμαλωτίζει την προσοχή ενός κοινού το οποίο έχει έρθει σε κατάλληλη

ψυχολογική κατάσταση». Δεν είναι, άλλωστε, η πρώτη φορά που διανοητές ή ερευνητές διατυπώνουν παρόμοιες απόψεις. Ο G. Le Bon (1895) είχε ισχυριστεί ότι το άτομο όταν απωλέσει την ιδιωτικότητα του και ενοποιηθεί με το πλήθος, υφίσταται την επιρροή των συναισθημάτων που δεσπόζουν στο πλήθος. Το κόμμα των Ναζί, στην προσπάθειά του να ομογενοποιήσει τις απόψεις και να διαδώσει την προπαγάνδα του, διένεμε κατά τη διάρκεια του μεσοπολέμου δωρεάν ραδιόφωνα στους πολίτες. Τα ραδιόφωνα αυτά είχαν μια ιδιαιτερότητα, μπορούσαν να συντονιστούν μόνο στη συχνότητα του ραδιοσταθμού του κόμματος των Ναζί (Adena, et al., 2015), παράλληλα, ο Franz von Pappen¹ προσπάθησε να μειώσει την πολυφωνία στο ραδιόφωνο (Doherty, 2000). Σε επίπεδο παραδοσιακών μέσων μαζικής ενημέρωσης, κυκλοφορούσε μια ημερήσια εφημερίδα και ένα έντυπο που εξέφραζαν τις απόψεις των Ναζί. Εκτός από τα παραπάνω, χρησιμοποιούντο ταινίες, αφίσες (πίνακας 7), λογοτεχνικά έργα κλπ. Ένα μεγάλο μέρος της προπαγάνδας των Ναζί αποσκοπούσε στο να προκαλέσει μίσος στον μέσο Γερμανό πολίτη προς τους Εβραίους, τους οποίους παρουσίαζε ως αιτία όλων των προβλημάτων που αντιμετώπιζε η Γερμανική κοινωνία και να δομήσει την εικόνα του Γερμανού Υπεράνθρωπου. Ένα άλλο σημαντικό μέρος της προπαγάνδας είχε ως στόχο το εξωτερικό και προσπαθούσε να δημιουργήσει φόβο στους αντιπάλους ή να σπείρει την αμφιβολία μεταξύ των συμμάχων για τις αγαθές προθέσεις που είχε ο ένας για τον άλλο (Casey, 1944).

ΠΙΝΑΚΑΣ 7 ΓΕΡΜΑΝΙΚΕΣ ΠΡΟΠΑΓΑΝΔΙΣΤΙΚΕΣ ΑΦΙΣΕΣ ΤΟΥ 2ΟΥ ΠΑΓΚΟΣΜΙΟΥ ΠΟΛΕΜΟΥ



¹ Franz von Papen: Γερμανός διπλωμάτης και στρατιωτικός. Μεταξύ άλλων, ανέλαβε καθήκοντα Αντικαγκελάριου τον Ιανουάριο του 1933 και είχε σημαντικό ρόλο στην ανατολή του Αδόλφου Χίτλερ στην εξουσία. Ήταν ένα από τα μέλη που προσήχθησαν στη γνωστή Δίκη της Νυρεμβέργης, όπου όμως απαλλάχθηκε των κατηγοριών.

- Στις ΗΠΑ, ιδρύθηκε το 1942 το Πολεμικό Γραφείο Πληροφοριών, μετά από απόφαση του προέδρου Ρούσβελτ. Η μονάδα αυτή υποστηριζόταν στο έργο της από διάφορες άλλες δομές προκειμένου να επιτύχει τη συλλογή πληροφοριών, αλλά και τη διάχυση της αμερικάνικης προπαγάνδας. Τα εργαλεία που αξιοποιούνταν ήταν διάφορα, όπως πόστερς (εικόνες του πίνακα 8), φυλλάδια που ρίχνονταν στις περιοχές που ήλεγχε ο εχθρός, έντυπα και κόμικς, ταινίες κλπ (Winkler, 1978).

ΠΙΝΑΚΑΣ 8 ΠΡΟΠΑΓΑΝΔΙΣΤΙΚΕΣ ΑΦΙΣΣΕΣ ΤΟΥ 2ΟΥ ΠΑΓΚΟΣΜΙΟΥ ΠΟΛΕΜΟΥ ΑΠΟ ΤΙΣ ΗΠΑ



- Οι Βρετανοί αξιοποίησαν την εμπειρία που είχαν αποκτήσει κατά τη διάρκεια του 1^{ου} ΠΠ. Επανιδρύουν το υπουργείο Πληροφοριών (McLaine, 1979), (Welch, 2016), που ήταν αρμόδιο για τα θέματα προπαγάνδας, όπως και στον 1^ο Παγκόσμιο Πόλεμο. Φροντίζουν να διαχέουν προπαγανδιστικό υλικό μέσω πολλαπλών διαύλων όπως τα πόστερς (πίνακας 9), ο ημερήσιος τύπος, οι κινηματογραφικές ταινίες, οι ραδιοφωνικές εκπομπές κα. Οι στόχοι ήταν πολλαπλοί, όπως:
 - Να στηρίξουν και να επηρεάσουν τους βρετανούς (Tate-Goins, 2011) (Aldgate & Richards, 2007), (Welch, 2016).
 - Να διασπείρουν ηττοπάθεια στις τάξεις των στρατευμάτων του Άξονα αλλά και να ενισχύσουν την διάθεση για αντίσταση των λαών που ήταν υπό γερμανική κατοχή. (Welch, 2016) .

ΠΙΝΑΚΑΣ 9 ΒΡΕΤΑΝΙΚΕΣ ΠΡΟΠΑΓΑΝΔΙΣΤΙΚΕΣ ΑΦΙΣΣΕΣ ΑΠΟ ΤΟ 2Ο ΠΑΓΚΟΣΜΙΟ ΠΟΛΕΜΟ



Κατά τη διάρκεια του ψυχρού πολέμου μεταξύ των ΗΠΑ και της ΕΣΣΔ, τόσο οι Σοβιετικοί όσο και οι ΗΠΑ επιδόθηκαν σε μια εκτεταμένη προπαγανδιστική εκστρατεία που στόχευε σε ακροατήρια τόσο στο εσωτερικό των χωρών τους όσο και στο εξωτερικό, με απώτερο σκοπό να παρουσιάσουν τον αντίπαλο τους ως την υπέρτατη απειλή για την ανθρωπότητα (Whitton, 1951). Στη μάχη για τη διάδοση της προπαγάνδας είχαν επιστρατευτεί όλα τα κανάλια επικοινωνίας από ραδιοφωνικές εκπομπές (Φωνή της Αμερικής, Ράδιο Μόσχα κλπ) έως εφημερίδες, κινηματογράφος και εικονογραφημένα περιοδικά (Cull, et al., 2017), (Kenez, 1985). Παραδείγματα στον Πίνακα 10.

ΠΙΝΑΚΑΣ 10 ΠΡΟΠΑΓΑΝΔΙΣΤΙΚΕΣ ΑΦΙΣΣΕΣ ΑΠΟ ΤΗΝ ΠΕΡΙΟΔΟ ΤΟΥ ΨΥΧΡΟΥ ΠΟΛΕΜΟΥ



Κατά τη διάρκεια τόσο του πολέμου στην Κορέα, στο Βιετνάμ όσο και πρόσφατα στο Ιράκ και το Αφγανιστάν, οι εμπλεκόμενοι, και κυρίως οι ΗΠΑ, είχαν εναποθέσει μεγάλο μέρος της στρατηγικής τους προσπάθειας στο να προπαγανδίσουν για το δίκαιο του αγώνα τους (Casey, 2008), (Berger, 1959), (Chandler, 1981), (Altheide & Grimes, 2005), (Blair, 1987). Βασικός πυλώνας στη διάχυση της προπαγάνδας αυτής ήταν φυσικά, τα μέσα μαζικής ενημέρωσης (Edwards & Cromwell, 2003). Όπως αναφέρει η Lieberman, ο Πρόεδρος των

ΗΠΑ έχει την εξουσία να διαχέει παγκόσμια, με βάση την εκτίμηση του, προκειμένου να προωθηθούν τα συμφέροντα και οι επιδιώξεις των ΗΠΑ (Lieberman, 2017). Παραδείγματα προπαγάνδας της περιόδου στον Πίνακα 11.

ΠΙΝΑΚΑΣ 11 ΠΡΟΠΑΓΑΝΔΙΣΤΙΚΕΣ ΑΦΙΣΣΕΣ ΑΠΟ ΤΟΥΣ ΠΟΛΕΜΟΥΣ ΣΤΗΝ ΚΟΡΕΑ ΚΑΙ ΤΟ ΒΙΕΤΝΑΜ



Σύμφωνα με τους Jowett & O'Donnell η ανάλυση της προπαγάνδας είναι μια σύνθετη και πολυεπίπεδη διαδικασία που μπορεί να αναλυθεί σε δέκα επίπεδα που αφορούν τον προσδιορισμό (Jowett & O'Donnell, 2012):

1. Της ιδεολογίας και των σκοπών της προπαγανδιστικής εκστρατείας.
2. Του γενικού πλαισίου μέσα στο οποίο λαμβάνει χώρα η προπαγάνδα.
3. Του προπαγανδιστή.
4. Της δομής του οργανισμού που ασκεί την προπαγάνδα.
5. Του ακροατήριου-στόχου.
6. Των τεχνικών αξιοποίησης των μέσων.
7. Των ειδικών τεχνικών που αξιοποιούνται με σκοπό την μεγιστοποίηση των αποτελεσμάτων.
8. Των αντιδράσεων του ακροατηρίου στις διάφορες τεχνικές.
9. Της αντιπροπαγάνδας.
10. Των αποτελεσμάτων και της αξιολόγησης τους.

3.3 Εισαγωγή στο φαινόμενο του ISIS

Για έναν ιδιωτικό ή πολιτικό οργανισμό έχει κρίσιμη σημασία η επιτυχημένη διάχυση της προπαγάνδας αλλά και της ιδεολογίας του στο ακροατήριο-στόχο που έχει. Η επιτυχής ή μη έκβαση του εγχειρήματος αυτού κρίνει σε μεγάλο βαθμό όχι μόνο την επιβίωση αλλά και την

εδραίωση και την παραπέρα ισχυροποίηση του οργανισμού αυτού. Έχει παρατηρηθεί ότι κατά τη διάρκεια των τελευταίων δεκαετιών, η διάχυση της ιδεολογίας καθώς και της προπαγάνδας ενός φορέα ή κράτους, επιτυγχάνεται με εκτεταμένη αξιοποίηση των μεθόδων και εργαλείων που σχετίζονται με τις ΤΠΕ. Ένας σημαντικός λόγος που καθορίζει την επιλογή αυτή, όπως επισημαίνεται από ερευνητές, είναι το γεγονός ότι αυτή η προσέγγιση προσφέρει το ιδανικό περιβάλλον για μια ευρεία, ταχύτατη και εύκολη διάχυση των ιδεών (Oates, 2003). Την αντίληψη αυτή υιοθετούν και ερευνητές σε θέματα στρατιωτικών και τρομοκρατικών οργανώσεων, υποστηρίζοντας ότι οι εν λόγω προσεγγίσεις διάχυσης της ιδεολογίας στο ευρύ κοινό αποδεικνύονται εξ ίσου πολύτιμες και για τις παραστρατιωτικές και τρομοκρατικές οργανώσεις (Bloom & Daymon, 2018), (Ingram, 2017), (Prucha, 2016), (Svete, 2008), (Jefferson, 2007), (Weimann, 2006), (Archetti, 2015). Μάλιστα, ο Weisman (2006) αναφέρει ότι μεταξύ των ετών 1997 και 2006 οι διαδικτυακοί τόποι που ήταν αφιερωμένοι σε τρομοκρατικές οργανώσεις αυξήθηκαν από 12 σε πάνω από 7000. Το γεγονός αυτό από μόνο του, καταδεικνύει τη σημασία που αποδίδουν οι τρομοκρατικές οργανώσεις στην προβολή τους μέσα από το διαδίκτυο.

Μεταξύ των τρομοκρατικών και παραστρατιωτικών οργανώσεων των τελευταίων δεκαετιών την σημαντικότερη θέση κατέχει ο ISIS. Πολλοί ερευνητές έχουν επιχειρήσει να μελετήσουν τη δομή και τον τρόπο λειτουργίας του ISIS με σκοπό να εξηγηθεί ο τρόπος με τον οποίο επετεύχθη η μεγάλη διεξόδωση των ιδεών της jihad όχι μόνο στον ισλαμικό κόσμο αλλά και σε ανθρώπους που είναι ξένοι προς το Ισλάμ (Winter, 2015), (Gambhir, 2016), (Ferrara, 2017), (Bloom & Daymon, 2018), (Carafella, et al., 2019).

Μελετώντας την πορεία που ακολούθησε ο ISIS, αντιλαμβάνεται κανείς ότι η οργάνωση αυτή, εργάστηκε από νωρίς σε πολυεπίπεδο βαθμό, προκειμένου να μην επιτρέψει μια στρατηγική ήττα του Χαλιφάτου, αναπτύσσοντας τις δυνάμεις του με τέτοιο τρόπο ώστε να είναι ικανό να διεξάγει υβριδικό πόλεμο και να διατηρεί κάτω από αντίξοες συνθήκες το ετοιμοπόλεμο των μονάδων του (Fernandez, 2015), (Carafella, et al., 2019). Στην κατεύθυνση αυτή δημιούργησε κεντρικές μονάδες σε διάφορες περιοχές του πλανήτη όπως στην Κεντρική Αφρική, τις Φιλιππίνες και την Ινδία. Αυτή η διάχυση του ISIS ερμηνεύθηκε από κάποιους ειδικούς στην τρομοκρατία ως δείγμα μιας απέλπιδας προσπάθειας της οργάνωσης να επιβιώσει (Shabad, 2016) ενώ άλλοι ερευνητές ισχυρίζονται ότι η παγκόσμια

εκστρατεία του ISIS είναι απόλυτα μελετημένη και δεν έχει ανακοπεί (Hoffman, 2019), (Carafella, et al., 2019).

Το στρατηγικό πλαίσιο δράσης του ISIS επικεντρωνόταν σε δυο άξονες, τον τοπικό μέσω της λειτουργίας των ελεγχόμενων περιοχών, και τον καθολικό μέσω της προώθησης της προπαγάνδας και άλλων κινήσεων (McFate, 2015).

Μετά από μελέτες, διάφοροι ερευνητές έχουν συγκλίνει στην άποψη ότι ο ISIS δαπάνησε ένα σημαντικό πλήθος ανθρώπινων και υλικών πόρων προκειμένου να είναι σε θέση να αξιοποιεί στο έπακρο τις ΤΠΕ στα πολιτικό-στρατιωτικά του σχέδια (Bloom & Daymon, 2018), (Ferrara, 2017), (Gambhir, 2016), (Ozeren, et al., 2018). Οι λόγοι που οδήγησαν τον ISIS στις επιλογές αυτές, όπως επισημαίνουν οι ερευνητές, ήταν πολλοί και σημαντικοί. Οι κυριότεροι είναι:

1. η εξασφάλιση της αποδοχής των ιδεών τους από ένα ευρύ ακροατήριο μέσα από το οποίο θα υπήρχε σταθερή ροή υποστηρικτών και ενεργών μελών (Mansour, 2018), (Ferrara, 2017), (Ozeren, et al., 2018)
2. η τρομοκράτηση των εχθρών τους (Bloom & Daymon, 2018), (Ferrara, 2017) (Cohen-Almagor, 2012), (Impara, 2018). Παράλληλα, οι Carafella et al. (2019) τονίζουν ότι οι επιλογές αυτές εξασφάλισαν στον ISIS αναχώματα αντίστασης. Συγκεκριμένα, οι εν λόγω εμπειρογνώμονες σε θέματα τρομοκρατίας γράφουν: “αυτές οι στρατηγικές επιλογές έχουν εξασφαλίσει στον ISIS να μην έχει ηττηθεί, παρά την απώλεια κυριαρχίας που έχει υποστεί στα εδάφη της Συρίας, γεγονός που τον διαφοροποιεί από την Al Qaeda στο Ιράκ το 2011” (Carafella, et al., 2019).

Οι ειδικοί της αντιτρομοκρατίας συμφωνούν ότι ο ISIS είχε μεγάλες προσδοκίες από την επιτυχή εφαρμογή του Πληροφοριακού πολέμου (Information Warfare) και για το λόγο αυτό ανέδειξε αυτόν ως ακρογωνιαίο λίθο του στρατηγικού του σχεδιασμού (Winter, 2015), (Gambhir, 2016), (Ingram, 2017), (Bloom & Daymon, 2018). Ειδικότερα:

- Ο Winter (2015) έχει την άποψη ότι «ο ISIS έχει υιοθετήσει την ιδέα της αξιοποίησης των διεθνών μέσων σε τρόπο που καμία τρομοκρατική ομάδα δεν είχε επιτύχει προηγουμένα και εξαπέλυσε, σε παγκόσμιο επίπεδο, μια τεράστια εκστρατεία διαμόρφωσης του προφίλ του και του ειδικού brand name, που φιλοδοξούσε να συνδέεται με την οργάνωση (branding strategy). Η στρατηγική εκπονήθηκε μέσα από την προσεκτικά

σχεδιασμένη διαμόρφωση και προώθηση της προπαγάνδας. Η διάδοση της προπαγάνδας υλοποιήθηκε μέσα από ένα ευρύ δίκτυο διανομένων ικανό να ανταποκρίνεται σε μεγάλες προκλήσεις». Τις ίδιες απόψεις έχει και η Gambhir (2016) που ισχυρίζεται ότι “Ο πόλεμος πληροφοριών που εξαπέλυσε ο ISIS υπηρέτησε πολλαπλούς στόχους, όπως τον έλεγχο της κατεχόμενης από αυτόν επικράτειας, τον εξαναγκασμό των πληθυσμών και την προσέλκυση εθελοντών”.

- Ο Ingram διατυπώνει ότι “είναι σημαντικό να αναγνωρίσει κανείς ότι οι προσπάθειες της διαδικτυακής προπαγάνδας του ISIS δεν εντάσσονταν απλώς σε μια στρατηγική προπαγάνδας, αλλά σε μια ευρύτερη πολιτικό-στρατιωτική εκστρατεία”.

Εστιάζοντας ο Winter (2015) στο ακροατήριο της προπαγάνδας του ISIS, το κατηγοριοποιεί στις ακόλουθες υποκατηγορίες:

- a) Υποστηρικτές και εμπλεκόμενα άτομα,
- b) Εχθροί και πιθανοί αντίπαλοι,
- c) Άτομα που δεν έχουν εμπλακεί,
- d) Φορείς διάδοσης, προσηλυτιστές και μέλη των καταλόγων του Ισλαμικού Κράτους.

Η επιτυχία που σημείωσε η προπαγανδιστική εκστρατεία του ISIS θα μπορούσε να θεωρηθεί ότι επιβεβαιώνει τις απόψεις του Ellul, σύμφωνα με τον οποίο, η επιτυχημένη προπαγάνδα οφείλει να στοχεύει αφενός στην ενεργοποίηση και αφετέρου στην πραγματοποίηση της σύνθεσης και της κρυστάλλωσης των προϋπαρχουσών ιδεών σε εστιασμένες ομάδες (Ellul, 1965). Ο ίδιος αναφέρει σε άλλο σημείο «..Ο προπαγανδιστής αναπτύσσει και προσαρμόζει τις τεχνικές του πάνω στη γνώση, τις επιθυμίες, τις ανάγκες, τον ψυχισμό και τη συμπεριφορά του ατόμου στο οποίο απευθύνεται».

Πράγματι, έχει επισημανθεί ότι η προπαγάνδα του ISIS στόχευε, με εξειδικευμένο τρόπο, σε πολλαπλά ακροατήρια ενώ ο ISIS είχε αρχίσει να εξαπολύει μια σειρά από τρομοκρατικά χτυπήματα σε διάφορες περιοχές του πλανήτη και ειδικά στην Ευρώπη (Cafarella & Zhou, 2017), (Nesser, et al., 2016). Όπως θα δειχθεί στη συνέχεια, η προπαγάνδα του ISIS ήταν προσεκτικά εστιασμένη στο ψυχισμό και τις ιδιαιτερότητες των επιμέρους ακροατηρίων που τον ενδιέφεραν.

Οι ΗΠΑ, σε μια απόπειρα πολυεπίπεδης αντεπίθεσης στην προπαγάνδα του ISIS προσπάθησαν να θέσουν σε αμφισβήτηση το ιδεολογικό και δογματικό πλαίσιο του ISIS.

Έτσι, όπως αναφέρουν οι Carafella et al. (2019), αξιοποίησαν κληρικούς και εκπαιδευτικά ιδρύματα από τη Σαουδική Αραβία και την Αίγυπτο, καθώς και πλατφόρμες ενημέρωσης, όπως η Global Coalition (2014). Με τον τρόπο αυτό ήθελαν αφενός να υποβαθμίσουν την προπαγάνδα που διέχεε ο ISIS μέσω των πλατφορμών κοινωνικής δικτύωσης (Twitter, Facebook κλπ) και αφετέρου να ανακόψουν την επιτυχημένη χρήση των ΤΠΕ από τον ISIS.

Οι παραπάνω ενέργειες αντιπροπαγάνδας των ΗΠΑ δεν φάνηκε ότι μείωσαν την απήχηση που είχαν τα μηνύματα του ISIS, ο οποίος συνέχισε να επενδύει πόρους στην παγκόσμια καμπάνια του, προκειμένου να φιλοτεχνήσει το προφίλ του και να εξυφάνει ένα αφήγημα που να δείχνει ότι διατηρεί τόσο την δυναμική του όσο και τη διατήρηση του αριθμού των υποστηρικτών του.

Στην κατεύθυνση αυτή, καλούσε τους Μουσουλμάνους απανταχού της Γης να συμμετάσχουν σε έναν ιερό πόλεμο (Fernandez, 2015) είτε μεταβαίνοντας στα πεδία των μαχών της Συρίας είτε ενεργώντας ως μοναχικοί λύκοι στους τόπους διαμονής τους (για παράδειγμα, στην όγδοη έκδοση του ηλεκτρονικού περιοδικού του ISIS με τίτλο Rumīyah, καλούντο όλοι οι συμπαθούντες τον ISIS σε έναν πόλεμο χωρίς σύνορα με στόχο όλους τους οικονομικούς πόρους της Δύσης).

Όπως θα αναλυθεί στη συνέχεια, η εκστρατεία αυτή πραγματοποιήθηκε μέσω:

- Ηλεκτρονικών περιοδικών που το περιεχόμενο τους συντασσόταν από μια επιλεγμένη ομάδα αρθρογράφων.
- Tweets και μικρών κειμένων που διαχέονταν στα μέσα κοινωνικής δικτύωσης.
- Πολυμεσικών αρχείων (φωτογραφίες, βίντεο, κλπ).

Με αυτόν τον τρόπο ο ISIS δημιούργησε πολιτική πόλωση στις δυτικές κοινωνίες και πέτυχε να εξαγει τις «μάχες» από τις περιοχές που ήλεγχε, μέσα στις πόλεις της Δυτικής Ευρώπης και να αναστατώσει τις κοινωνίες των χωρών αυτών.

Για τους παραπάνω λόγους, οι ερευνητές του ISIS έχουν προσπαθήσει να αποκωδικοποιήσουν τα μηνύματα που εκπέμπονταν από τον ISIS και εξέφραζαν την προπαγάνδα του Ισλαμικού Κράτους. Η άποψη των περισσότερων ειδικών είναι ότι η προπαγάνδα του ISIS βασίζεται σε ένα σύνολο από τέσσερα έως έξι αφηγήματα (Fernandez, 2015), (Winter, 2015), (Impara, 2018), (Bloom & Daymon, 2018). Τα αφηγήματα αυτά έχουν

ως στόχο να πυροδοτήσουν και να ενισχύσουν το πρώιμο, κατά Ellul, προπαγανδιστικό υπόβαθρο των ατόμων (Ellul, 1965), σε σχέση με τις παρακάτω συνιστώσες:

- ✓ Βαρβαρότητα (Brutality)
- ✓ Έλεος (Mercy)
- ✓ Πόλεμος (War)
- ✓ Θυματοποίηση (Victimhood)
- ✓ Συμμετοχή σε ομάδα (Belonging)
- ✓ Ουτοπισμός (Utopianism)

Πιστεύουμε ότι η ενδελεχής έρευνα αναφορικά με τους παράγοντες και τις παραμέτρους που διασφάλισαν την επιτυχή διάχυση της προπαγάνδας του ISIS αποτελεί ένα ενδιαφέρον πρόβλημα έρευνας, το οποίο θα μπορούσε να έχει εφαρμογή τόσο στην πολιτική, όσο και στην διοίκηση επιχειρήσεων και το μάρκετινγκ. Για το λόγο αυτό προχωρήσαμε στην μελέτη των κειμένων της προπαγάνδας του ISIS μέσω της μηχανικής μάθησης και της ανάλυσης υποκειμενικής γνώμης.

Οι ερευνητικές αναζητήσεις που καθόρισαν την έρευνα μας είναι οι παρακάτω:

1. **RQ1.** Είναι δυνατόν, μέσω κάποιας προσέγγισης, να μην περιοριστούμε σε μια ανάλυση πολικότητας της προπαγάνδας του ISIS και να αποκτήσουμε μέσα από δομημένο τρόπο μια βαθύτερη γνώση για τα συναισθήματα που οι δημιουργοί της προπαγάνδας του ISIS είχαν σκοπό να προκαλέσουν στα υποκείμενα που ανήκαν στο ακροατήριο-στόχο του;
2. **RQ2.** Μπορούμε να βελτιώσουμε και σε ποιο βαθμό την κατανόηση μας, συγκριτικά με τις παλαιότερες μελέτες που περιορίζονταν στην ανάλυση πολικότητας και μόνο;
3. **RQ3.** Θα μπορούσαμε να χρησιμοποιήσουμε την συνολική συλλογή των κειμένων της προπαγάνδας (corpus) προκειμένου να προσδιορίσουμε τα υποβόσκοντα πρότυπα στα οποία ανήκαν τα υποκείμενα του ακροατηρίου-στόχου;
4. **RQ4.** Εμφανίζονται διαφορές, ανάλογα με την έκδοση του περιοδικού προπαγάνδας (περιοδικό) που χρησιμοποιούσε ο ISIS, (Αγγλόφωνα, Γαλλόφωνα, Ρωσόφωνα, Τουρκόφωνα) και αν ναι, πως αυτές συσχετίζονται με τα διαφορετικά ακροατήρια; Έχουν σχέση οι τυχόν παρατηρούμενες διαφορές με διαφορές στα προκαλούμενα συναισθήματα ή τα υποβόσκοντα πρότυπα;

5. **RQ5.** Διακρίνονται κάποια διαρκή θέματα (topics) συζήτησης τα οποία επιλέγονται από τους δημιουργούς της προπαγάνδας του ISIS;
6. **RQ6.** Παρατηρούνται ομοιότητες/διαφορές μεταξύ των συναισθηματικών προτύπων που εντοπίζονται, ανάμεσα σε άρθρα γνωστών ειδησεογραφικών πρακτορείων και αυτά του ISIS;
7. **RQ7.** Μπορούμε να χρησιμοποιήσουμε μοντέλα μηχανικής μάθησης, ώστε να εντοπίσουμε ανωμαλίες σε μια συλλογή κειμένων (corpus), βασιζόμενοι μόνο στο συναίσθημα;
8. **RQ8.** Ποια μοντέλα μηχανικής μάθησης είναι πιο αποτελεσματικά στον εντοπισμό ανώμαλων κειμένων με βάση το συναίσθημα;

Λόγω του γεγονότος ότι τα δεδομένα μας έχουν αδόμητη μορφή (κείμενα) επιλέξαμε να απαντήσουμε στα παραπάνω ερευνητικά ερωτήματα μέσα από το μεθοδολογικό πλαίσιο της ανάλυσης κειμένου με μη εποπτευόμενη μάθηση και την εξόρυξη άποψης (opinion mining). Μέσω των επιλεχθεισών τεχνικών αναλύθηκαν τα κείμενα που περιέχονταν σε όλα τα τεύχη των γνωστών περιοδικών του ISIS, που έχουν εκδοθεί και αναρτηθεί ηλεκτρονικά στο διαδίκτυο με σκοπό να δοθούν απαντήσεις στα παραπάνω ερωτήματα.

Η δομή του υπόλοιπου κεφαλαίου έχει ως εξής:

- Η επόμενη παράγραφος με τίτλο “Κατανοώντας την προπαγάνδα του ISIS” παρέχει μια ανασκόπηση από σχετικές έρευνες και εκθέσεις. Στην παράγραφο αυτή εστιάζουμε στο επιχειρησιακό πλάνο του ISIS, αναφορικά με τον Πληροφοριακό πόλεμο και ειδικότερα με την προπαγάνδα. Παράλληλα, παρέχεται μια εκτενής περιγραφή των διάφορων περιοδικών που είχε εκδώσει ο ISIS προκειμένου να διαχύσει την προπαγάνδα του σε διάφορα ακροατήρια-στόχους. Όπως θα διαπιστώσει ο αναγνώστης, καθένα από τα περιοδικά κάλυψε μια συγκεκριμένη περίοδο, γεωγραφία και κοινό, μέσω των διαφόρων εκδόσεων του.
- Η παράγραφος 3.5 περιγράφει εκτενώς το μεθοδολογικό πλαίσιο του αξιοποιήθηκε κατά την έρευνα. Ειδικότερα, στην παράγραφο αυτή αναλύονται θέματα που άπτονται της:
 - Συλλογής των δεδομένων προς ανάλυση.
 - Προετοιμασίας των δεδομένων και της προεπεξεργασίας.
 - Στατιστικής ανάλυσης.
 - Ανάλυσης κειμένου.

- Η παράγραφος 3.6 περιέχει τα αποτελέσματα της έρευνας.
- Η παράγραφος 3.7 εστιάζει στα συμπεράσματα της έρευνας που υλοποιήθηκε.

3.4 Κατανοώντας την προπαγάνδα του ISIS

Όπως έχει αναφερθεί και σε προηγούμενα σημεία της Θέσης, ο ISIS προχώρησε σε μια προπαγανδιστική εκστρατεία που χαρακτηρίστηκε από την μεγάλη της ένταση, την ευρύτητα της διεξαγωγής της, την πληθώρα των ακροατηρίων στα οποία στόχευε αλλά και των μέσων που αξιοποίησε. Μέσα από κείμενα, ηχητικά αρχεία, φωτογραφίες και βίντεο που προωθούνταν μέσω του Διαδικτύου, ο ISIS πέτυχε τη διάχυση των πληροφοριών αυτών σε παγκόσμιο επίπεδο (Bloom & Daymon, 2018) (Gambhir, 2016). Το αποτέλεσμα αυτό δείχνει την υψηλή ποιότητα της διάρθρωσης του ISIS, καθώς και το «αξιόμαχο» της πτέρυγας του οργανισμού που είχε την ευθύνη των Επικοινωνιών. Όπως σημειώνουν ερευνητές, η πτέρυγα αυτή διέθετε αποτελεσματική διοίκηση, άψογη στρατιωτική εκπαίδευση και μεγάλη εξοικείωση με την αξιοποίηση των Μέσων Επικοινωνίας και των ICT (Gambhir, 2016), (Carafella, et al., 2019). Συγκεκριμένα, στο Gambhir (2016) περιέχεται μια εξαιρετική αποτύπωση για την πτέρυγα του ISIS που ήταν επιφορτισμένη με την διαχείριση των MME και των ΤΠΕ. Στη σελίδα 21 της έκθεσης αυτής αναφέρεται ότι «όλες οι κύριες μονάδες των MME λειτουργούν κάτω από την καθοδήγηση και την αρχηγεία του Κεντρικού Διβανίου των MME (Central Media Diwan)».



ΕΙΚΟΝΑ 15 ΜΑΧΗΤΗΣ ΤΟΥ ISIS. ΦΑΝΕΡΗ Η ΑΝΤΙΓΡΑΦΗ ΤΗΣ ΙΔΕΑΣ ΑΠΟ ΤΗΝ ΠΡΩΤΗ ΕΙΚΟΝΑ ΤΟΥ ΠΙΝΑΚΑ 8

Με τον τρόπο αυτό, ο ISIS κατάφερε να δημιουργήσει μια σταθερή και αναγνωρίσιμη μάρκα (Brand) που είχε ασύγκριτα μεγαλύτερη επίδραση από άλλες παρόμοιες οργανώσεις όπως η Al-Qaeda (Gambhir, 2016). Παράλληλα, επέδειξε ικανότητα να δημιουργεί υποστηρικτές σε παγκόσμιο επίπεδο και να στρατολογεί μεγάλο πλήθος αλλοδαπών Τζιχαντιστών σε σύγκριση με τον αριθμό των μαχητών που στρατολογούσε από τα κατεχόμενα εδάφη της Συρίας και του Ιράκ (McCabe, 2016). Το έτος 2015, ο ακαδημαϊκός ερευνητής Aymenn al-Tamimi διέρρευσε ένα πλήθος εγγράφων του ISIS στο πρακτορείο ειδήσεων Guardian. Τα

έγγραφα αυτά περιείχαν σημαντικές πληροφορίες για τη δομή του ISIS (Al-Tamimi, 2015) (Malik, 2015) και αποτελούν το σύνολο των δεδομένων που αναλύονται στο παρόν κεφάλαιο.

Συγκεκριμένα, η έρευνα μας εστιάζει στις εκδόσεις του Al-Hayat Media Center. Το συγκεκριμένο κέντρο εξέδιδε τα περιοδικά:

- ✓ Dabiq,
- ✓ Rumiyah,
- ✓ Istok,
- ✓ Dar al Islam
- ✓ Konstantiniyye

Από αυτά, το Dabiq και το Rumiyah αποτελούσαν τις πλέον σταθερές, σε βάθος χρόνου, εκδόσεις. Η μεθοδολογία με την οποία τα περιοδικά αυτά προωθούσαν την προπαγάνδα του ISIS για την καθιέρωση της επωνυμίας του έχει εξετασθεί από διάφορους ερευνητές (Kinney, et al., 2018), (Impara, 2018), (Gambhir, 2016), (Winter, 2015), (Ingram, 2017).



ΕΙΚΟΝΑ 16 ΕΞΩΦΥΛΛΟ ΤΟΥ DABIQ, ΜΕ ΑΦΟΡΜΗ ΤΡΟΜΟΚΡΑΤΙΚΗ ΕΠΙΘΕΣΗ ΣΤΟ ΠΑΡΙΣΙ

Ο παρακάτω πίνακας περιέχει βασικές πληροφορίες για τις εκδόσεις αυτές που αφορούν:

- τη γλώσσα που χρησιμοποιούσε το κάθε περιοδικό,
- την ημερομηνία έναρξης και λήξης των εκδόσεων καθώς και
- το πλήθος των εκδόσεων.

Επειδή τα περιοδικά ακολουθούσαν το Ισλαμικό ημερολόγιο (μήνας, έτος) παρατίθεται σε παρένθεση η κατά προσέγγιση ημερομηνία έναρξης και λήξης, με βάση το αντίστοιχο Γρηγοριανό ημερολόγιο (Richards, 2012).

ΠΙΝΑΚΑΣ 12 ΣΤΟΙΧΕΙΑ ΤΩΝ ΕΚΔΟΣΕΩΝ ΤΟΥ AL-HAYAT MEDIA CENTER

Τίτλος Περιοδικού	Γλώσσα	Αριθμός εκδόσεων	Πρώτη έκδοση	Τελευταία γνωστή έκδοση
Dabiq	Αγγλικά	15	Ramadan 1435 (Ιούλιος 2014)	Shawwal 1437 (Ιούλιος 2016)
Rumiyah²	Αγγλικά	13	Dhul-Hijjaj 1437 (Σεπτέμβριος 2016)	Dhul-Hijjaj 1438 (Σεπτέμβριος 2017)
Dar al Islam	Γαλλικά	10	Rabi al-Awwal 1436 (Δεκέμβριος 2014 / Ιανουάριος 2015)	Dhul Qidah 1437 (Αύγουστος/ Σεπτέμβριος 2016)
Istok	Ρωσικά	4	Rajab 1436 (Απρίλιος/ Μάιος 2015)	Rajab 1437 (Απρίλιος/ Μάιος 2016)
Konstantiniyye	Τούρκικα	7	Shaban 1436 (Μάιος/ Ιούνιος 2015)	Dhu al-Qidah 1437 (Αύγουστος/ Σεπτέμβριος 2016)

3.5 Μεθοδολογικό πλαίσιο

3.5.1 Συλλογή Δεδομένων

Τα τεύχη των περιοδικών ανευρέθηκαν σε μορφή PDF από τον δημόσιο ιστότοπο jihodology.net (Kinney, et al., 2018). Λόγω του γεγονότος ότι τα περιοδικά διανεμήθηκαν κυρίως μέσω του Deep Web³, φροντίσαμε να διασταυρώσουμε το περιεχόμενο των περιοδικών, με τα αρχεία των εν λόγω εκδόσεων των περιοδικών που ανακτήσαμε από διαφορετική πηγή (archive.org). Μέσω της διασταύρωσης αυτής εξασφαλίστηκε η εγκυρότητα του περιεχομένου των προς ανάλυση περιοδικών.

3.5.2 Προετοιμασία δεδομένων

Το μορφότυπο των αρχείων μετατράπηκε από αυτό του φορητού εγγράφου (PDF) σε εκείνο του αρχείου κειμένου (text file) με την αξιοποίηση λογισμικού αναγνώρισης χαρακτήρων (OCR) (Kinney, et al., 2018). Τα κείμενα που δεν ήταν σε Αγγλική γλώσσα και προέρχονταν από τα περιοδικά Dar al Islam, Istok και Konstantiniyye μεταφράστηκαν στα Αγγλικά πριν από την ανάλυση. Έχει αποδειχθεί ότι η παρέμβαση αυτή δεν επιδρά αρνητικά στη διαδικασία της ανάλυσης (Chatzakou, et al., 2017) (Balahur & Turchi, 2012) (Martin-Valdivia, et al., 2013). Η μετάφραση των εν λόγω κειμένων υλοποιήθηκε μέσω της Python με ταυτόχρονη αξιοποίηση της Google Translate API. Η επιλογή μας αυτή καθορίστηκε από το γεγονός ότι η

² Dabiq's successor (Kinney, et al., 2018) (Gambhir, 2016).

³ Το βαθύτερο μέρος του internet (γνωστό και ως αόρατο, ή κρυμμένο web) το οποίο δεν ευρετηριάζεται από μία συνηθισμένη μηχανή αναζήτησης και συνήθως απαιτεί επιπλέον εργαλεία πλοήγησης.

δεδομένη υπηρεσία μετάφρασης κειμένου έχει χρησιμοποιηθεί επιτυχώς από άλλους ερευνητές σε μελέτες με αντικείμενο την ανάλυση συναισθήματος (Chatzakou, et al., 2017) (Balahur & Turchi, 2012).

3.5.3 Προεπεξεργασία

Στο στάδιο της προεπεξεργασίας χρησιμοποιήθηκε η Python 3.0 καθώς και η βιβλιοθήκη NLTK 3.3 προκειμένου να υλοποιηθούν οι εργασίες που αφορούσαν την επεξεργασία της φυσικής γλώσσας (natural language processing, σύντομα NLP). Το κείμενο από κάθε έκδοση των περιοδικών πήρε τη μορφή λεκτικών μονάδων (tokenization) και αφαιρέθηκαν οι κοινές λέξεις (stopwords) όπως είναι άρθρα, προθέσεις, κλπ. Ακόμα αφαιρέθηκαν τα σημεία στίξης (όπως τελείες, ερωτηματικά) καθώς και οι αριθμοί. Τέλος, οι λέξεις που απέμειναν υπέστησαν λημματικοποίηση (lemmatization) προκειμένου οι διάφορες φόρμες μιας λέξης να ομαδοποιηθούν και να αναχθούν στο λήμμα τους, με βάση τη συντακτική τους θέση (Chatzakou, et al., 2017).

3.5.4 Η εύρεση του θέματος μέσω της Latent Dirichlet Allocation (LDA-Topic Modelling)

Για να υλοποιηθεί η εξεύρεση των υποβοσκοτών θεμάτων στα αναλυόμενα κείμενα χρησιμοποιήθηκε η βιβλιοθήκη Gensim (Rehurek & Sojka, 2010). Συγκεκριμένα, τα δεδομένα που παράχθηκαν στο στάδιο της προεπεξεργασίας αναλύθηκαν με τον αλγόριθμο της LDA (Blei, et al., 2003). Όπως έχει ήδη αναφερθεί σε προηγούμενο σημείο της Θέσης, ο αλγόριθμος αυτός αποτελεί ένα γενετικό πιθανοθεωρητικό μοντέλο που μπορεί να αξιοποιηθεί σε κείμενα (text corpora) (Blei, et al., 2003). Το μοντέλο αυτό έχει επιλεγεί τόσο λόγω της επάρκειας που έχει επιδείξει να διαχειρίζεται την εύρεση θεμάτων (topics) σε προσεγγίσεις σάκου λέξεων, όσο και λόγω του ότι έχει αξιοποιηθεί με επιτυχία σε προηγούμενες συναφείς έρευνες (Kinney, et al., 2018).

3.5.5 Το Λεξικό

Προκειμένου να προσδιορίσουμε και να εξαγάγουμε συγκεκριμένα συναισθήματα υλοποιούμε ανάλυση βασίζεται στη χρήση Λεξικού. Συγκεκριμένα, αξιοποιούμε το Λεξικό EmoLex. Όπως έχει αναφερθεί και στο δεύτερο κεφάλαιο, πρόκειται για ένα λεξικό που έχει δημιουργηθεί με την προσέγγιση crowdsourcing και συσχετίζει λέξεις με πρωτογενή συναισθήματα (Mohammad & Turney, 2010), (Mohammad & Turney, 2013). Το εν λόγω

λεξικό έχει χρησιμοποιηθεί με επιτυχία σε προηγούμενες μελέτες (Chatzakou, et al., 2017), (Zuorba, et al., 2017), (Kinney, et al., 2018) και προσφέρει μεγάλη κάλυψη σε λέξεις. Ένα πρόσθετο πλεονέκτημα του εν λόγω λεξικού είναι ότι ενσωματώνει μέσα του λεξικά που έχουν δημιουργηθεί πριν από αυτό (Mohammad & Turney, 2013) (Meiselman, 2016). Κάθε λέξη που εμπεριέχεται στο λεξικό EmoLex αντιστοιχίζεται σε δέκα εκτιμήσεις (ratings). Οι οκτώ εξ αυτών αντιστοιχούν στα οκτώ πρωταρχικά ανθρώπινα συναισθήματα, σύμφωνα με τον τροχό των συναισθημάτων (Plutchik, 1980). Το μοντέλο αυτό ακολουθείται λόγω του ότι είναι το πλέον αποδεκτό (Meiselman, 2016). Οι υπόλοιπες δύο εκτιμήσεις αναπαριστούν την θετική και αρνητική πολικότητα της θεωρούμενης λέξης (Mohammad & Turney, 2010), (Mohammad & Turney, 2013). Με τον τρόπο αυτό επιχειρούμε να προσδιορίσουμε όχι μόνο την πολικότητα αλλά και τα συναισθήματα που υποβόσκουν σε κάποιο κείμενο. Οφείλουμε να επισημάνουμε ότι αν η ανάλυση περιοριζόταν μόνο στην εκτίμηση της πολικότητας των κειμένων τότε η περιγραφική ικανότητα του μοντέλου ανάλυσης θα ήταν εξαιρετικά περιορισμένη. Στην περίπτωση μας μάλιστα, θα ήταν αδύνατο να απαντηθούν τα ερευνητικά ερωτήματα που μας απασχολούν. Σε κάθε περίπτωση, η πολικότητα των κειμένων έχει αποτυπωθεί και θα συνεκτιμηθεί με τα υπόλοιπα αποτελέσματα, προκειμένου να αναδειχθεί η υπεροχή της ανάλυσης μέσω των συναισθημάτων σε σχέση με την ανάλυση μέσω μόνο της πολικότητας.

3.5.6 Κανονικοποίηση

Μετά από τον προσδιορισμό των λέξεων-Θεμάτων (topic words) από την LDA ακολουθεί η εξαγωγή εκτιμήσεων των συναισθημάτων (emotion ratings). Η διαδικασία αυτή πραγματοποιείται με την εύρεση και αντιστοίχιση κάθε λέξης με τις λέξεις που περιέχονται στο EmoLex. Για κάθε λέξη που γίνεται αντιστοίχιση στο λεξικό παράγονται δύο διανύσματα. Συγκεκριμένα:

- Το πρώτο διάνυσμα έχει οκτώ συνιστώσες. Κάθε μια εξ αυτών αντιστοιχεί στην ένταση ενός εκ των οκτώ πρωταρχικών συναισθημάτων που ανήκουν στο σύνολο $\Sigma = \{joy, trust, fear, surprise, sadness, disgust, anger, anticipation\}$ και εγείρονται μέσω της λέξης.
- Το δεύτερο διάνυσμα έχει δύο συνιστώσες που αφορούν στην θετική και αρνητική πολικότητα της λέξης.

Με σκοπό να δημιουργήσουμε για κάθε κείμενο στη συλλογή (corpus):

- Ένα συνολικό διάνυσμα που θα εκφράζει τις εντάσεις των πρωταρχικών συναισθημάτων,
- Ένα διάνυσμα που θα αναπαριστά τη συνολική πολικότητα (Θετική/Αρνητική), χρησιμοποιούμε την διανυσματική άθροιση ενώ στη συνέχεια τα παραγόμενα διανύσματα κανονικοποιούνται, για λόγους που έχουν να κάνουν με την αποτελεσματικότερη λειτουργία των αλγορίθμων ανάλυσης.

Αναλυτικά, η διαδικασία έχει ως εξής:

Έστω ότι:

- Το corpus είναι το σύνολο $C = \{D_1, D_2, \dots, D_m\}$ όπου D_i το έγγραφο με δείκτη i και το οποίο ανήκει στο θεωρούμενο corpus και
- Το σύνολο των λέξεων που υπάρχουν στο corpus είναι

$$W = \{w_1, w_2, \dots, w_n\}: \forall w_i \in W \exists \vec{e}_i \in R^8 \text{ και } \vec{p}_i \in R^2 \text{ με } i=1,2,\dots,n.$$

Αλγόριθμος

Στο corpus των εγγράφων αλλά και σε επίπεδο λέξεων-θεμάτων της LDA, εκτελείται ο εξής αλγόριθμος:

Βήμα 1^ο: Δημιουργία των διανυσμάτων συναισθήματος και πολικότητας σε επίπεδο εγγράφου.

$$\forall D_j \in C \text{ και } \forall i: w_i \in D_j \cap W \\ \Rightarrow \{ \vec{E}_j = \sum \vec{e}_i \text{ και } \vec{P}_j = \sum \vec{p}_i \}$$

Βήμα 2^ο: Κανονικοποίηση των διανυσμάτων συναισθήματος και πολικότητας.

$$\forall D_j \in C \Rightarrow \left\{ \vec{E}_j \leftarrow \frac{1}{\sum_{k=1}^8 E_{jk}} * \vec{E}_j \text{ και } \vec{P}_j \leftarrow \frac{1}{\sum_{k=1}^2 P_{jk}} * \vec{P}_j \right\}$$

Προκειμένου να γίνει αυτό κατανοητό παρατίθεται το επόμενο παράδειγμα. Έστω ότι έχουμε ένα έγγραφο D που απαρτίζεται από δύο λέξεις Word 1 και Word 2. Έστω ότι οι εντάσεις των βασικών συναισθημάτων στις λέξεις αυτές αντιστοιχούν στις συνιστώσες των διανυσμάτων που εμφανίζονται στην πρώτη και δεύτερη γραμμή των δεδομένων του πίνακα που ακολουθεί. Το διάνυσμα των εντάσεων των βασικών συναισθημάτων για το έγγραφο D δημιουργείται ως εξής:

1. Δημιουργούνται τα επιμέρους αθροίσματα των αντίστοιχων συνιστωσών όλων των λέξεων (3^η γραμμή του πίνακα 13).
2. Δημιουργείται το άθροισμα των επιμέρους αθροισμάτων (4^η γραμμή του πίνακα 13).
3. Διαιρούνται τα μερικά αθροίσματα της τρίτης γραμμής με το άθροισμα όλων των επιμέρους μερικών αθροισμάτων και αντικαθίστανται ως οι νέες τιμές στις συνιστώσες του διανύσματος έντασης συναισθημάτων του εγγράφου D (5^η γραμμή του πίνακα 13).

Ανάλογα βήματα ακολουθούνται, όπως περιγράφει και ο αλγόριθμος, για τη δημιουργία του διανύσματος πολικότητας του εγγράφου D. Σχετικός πίνακας 13.

ΠΙΝΑΚΑΣ 13 ΚΑΝΟΝΙΚΟΠΟΙΗΣΗ ΔΙΑΝΥΣΜΑΤΩΝ ΤΩΝ ΛΕΞΕΩΝ ΤΟΥ ΛΕΞΙΚΟΥ EmoLex

Διάνυσμα	Συνιστώσες							
	1 (joy)	2 (trust)	3 (fear)	4 (surprise)	5 (sadness)	6 (disgust)	7 (anger)	8 (anticip)
Word 1	1	0	0	0	0,67	0	1	0,56
Word 2	0,29	0	0	0,14	0	0,43	0	0,14
Βήμα 1: Άθροισμα ανά συναίσθημα	1,29	0	0	0,14	0,67	0,43	1	0,7
Βήμα 2: Συνολικό άθροισμα	4,23	4,23	4,23	4,23	4,23	4,23	4,23	4,23
Βήμα 3: Δημιουργία Κανον/μένων συνιστωσών	0,31	0	0	0,03	0,16	0,1	0,24	0,17

3.6 Αποτελέσματα

3.6.1 Βασική ανάλυση αποτελεσμάτων

Κατ' αρχήν προχωρήσαμε σε διερευνητική (exploratory) ανάλυση των αποτελεσμάτων μεταξύ των περιοδικών.

Αρχικά, υπολογίσαμε το πλήθος των λέξεων ανά έγγραφο μετά την προεπεξεργασία (word count) καθώς και την κάλυψη (coverage) η οποία υπολογίζεται από τον τύπο που ακολουθεί:

$$\text{Coverage εγγράφου} = \frac{\text{Πλήθος των λέξεων του εγγράφου που ανήκουν και στο EmoLex}}{\text{Πλήθος των λέξεων του εγγράφου, μετά τη διαδικασία προεπεξεργασίας}}$$

Βρέθηκε ότι η κάλυψη που επιτυγχανόταν μέσω του EmoLex ανερχόταν σε ποσοστό, κατά μέσο όρο, 65%. Η κάλυψη αυτή άγγιζε το μέσο όρο κάλυψης 67% για τις εκδόσεις του αγγλόφωνου περιοδικού Dabiq ενώ, τόσο για τις εκδόσεις του Rumiyaḥ, όσο και για το ρωσσόφωνο Istok, ο μέσος όρος κάλυψης των εκδόσεων από το EmoLex διαμορφωνόταν στο 70% (λεπτομερής πίνακας στο παράρτημα Α, πίνακας πλήθους λέξεων και κάλυψης).

Κατά την ανάλυση του πλήθους των λέξεων ανά περιοδικό, παρατηρήθηκαν δύο αξιοσημείωτα χαρακτηριστικά που αφορούσαν τα περιοδικά Dabiq και Rumiyaḥ και τα οποία στόχευαν στο αγγλόφωνο ακροατήριο του ISIS.

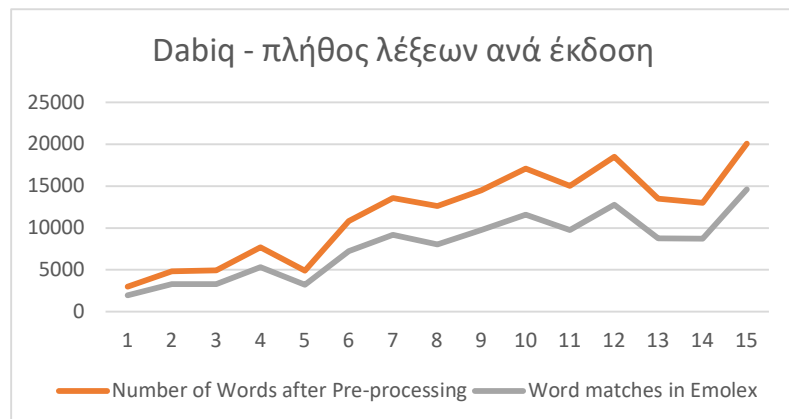
Συγκεκριμένα:

1. Αναφορικά με το Dabiq, παρατηρήθηκε αύξηση του πλήθους των λέξεων ανά έκδοση σε όλη την περίοδο που ήταν σε κυκλοφορία (Εικόνα 17). Η τελευταία έκδοση του Dabiq εμφάνισε και το μεγαλύτερο πλήθος λέξεων σε σχέση με τις εκδόσεις του αλλά και τις εκδόσεις όλων των υπολοίπων περιοδικών. Το υψηλό αυτό άγγιξε τις 20.000 λέξεις. Την περίοδο εκείνη ακολούθησε η αεροπορική επιδρομή των ΗΠΑ στη Συρία (Αύγουστος 2016) που είχε σαν αποτέλεσμα το θάνατο του Abu Muhammad Al-Adnani. Ο Al-Adnani ήταν κορυφαίο στέλεχος του ISIS, βασικός ομιλητής και εκφραστής της προπαγάνδας της οργάνωσης αλλά και επιφορτισμένος με την ευθύνη της διεξαγωγής των εξωτερικών επιχειρήσεων της οργάνωσης. Ο θάνατος του Al-Adnani έκανε τους ειδικούς της αντιτρομοκρατίας αναφορικά με τον ISIS να αναρωτιούνται αν:

- a. οι μονάδες του ISIS που είχαν ως αποστολή την αξιοποίηση των Μέσων Επικοινωνιών θα συνέχιζαν να είναι τόσο λειτουργικές.
- b. οι ενέργειες της οργάνωσης μέσω του διαδικτύου θα διατηρούσαν την αποτελεσματικότητά τους.

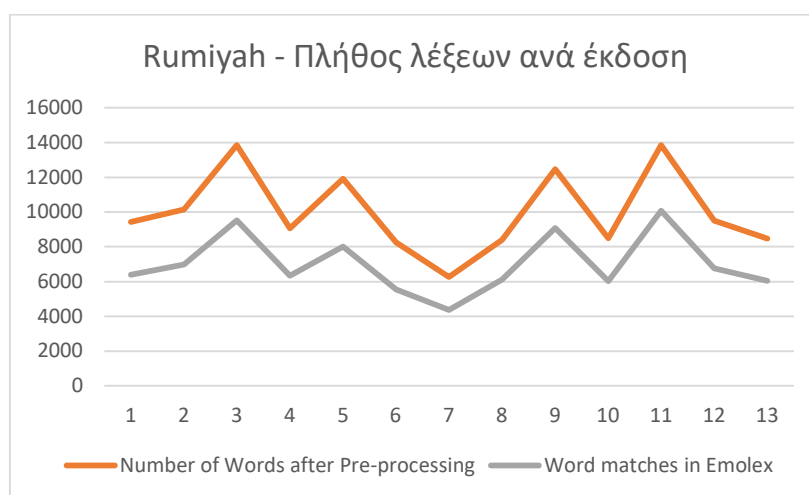
Πράγματι, έξι ημέρες αργότερα, διακόπτεται η κυκλοφορία του Dabiq και το διαδέχεται το περιοδικό Rumiyaḥ (Gambhir, 2016).

ΕΙΚΟΝΑ 17 ΜΕΤΑΒΟΛΗ ΤΟΥ ΠΛΗΘΟΥΣ ΛΕΞΕΩΝ ΑΝΑ ΕΚΔΟΣΗ ΤΟΥ DABIQ ΚΑΙ Η ΚΑΛΥΨΗ ΑΠΟ ΤΟ EMOLEX



2. Το διάγραμμα που περιέχεται στην Εικόνα 18, αφορά το πλήθος των λέξεων ανά έκδοση για το το Rumiyaḥ που απετέλεσε το διάδοχο του Dabiḡ. Είναι φανερό ότι:

- a. το Rumiyaḥ εμφάνιζε έντονες διακυμάνσεις στο πλήθος των λέξεων ανά έκδοση και
- b. δεν κατάφερε να αγγίξει ποτέ το ρεκόρ των λέξεων που είχε κατακτήσει το περιοδικό Dabiḡ. Το μεγαλύτερο πλήθος λέξεων ανά έκδοση για το Rumiyaḥ σημειώθηκε στην 11^η έκδοση ήταν 13839 λέξεις και φυσικά, ως αριθμός υπολείπεται σημαντικά του αριθμού των 20.000 λέξεων που είχε θέσει ως μέγιστο το Dabiḡ όταν έπαψε να κυκλοφορεί.



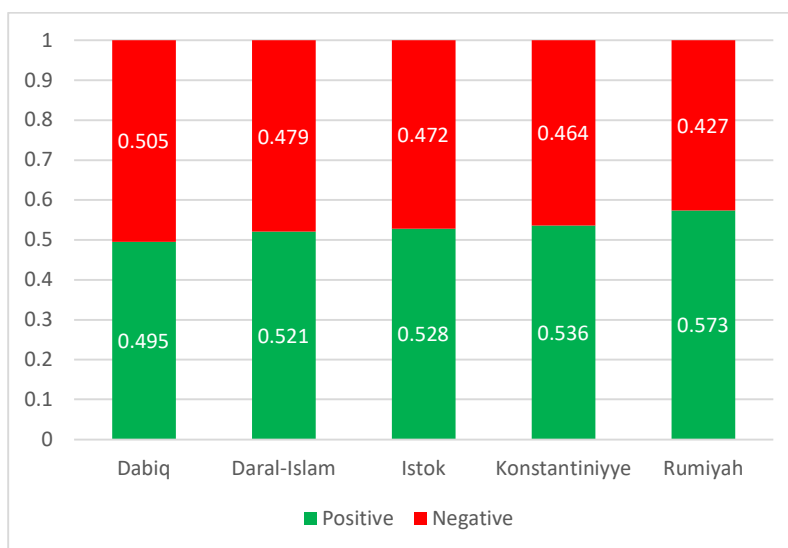
ΕΙΚΟΝΑ 18 ΠΛΗΘΟΣ ΛΕΞΕΩΝ ΑΝΑ ΕΚΔΟΣΗ ΤΟΥ RUMIYAḤ ΚΑΙ ΚΑΛΥΨΗ ΑΠΟ ΤΟ ΕΜΟΛΕΧ

Τα παραπάνω μπορεί να οφείλονται σε διάφορους παράγοντες όπως η έλλειψη συντονισμού των μονάδων αξιοποίησης των Μέσων Επικοινωνίας που παρατηρήθηκε μετά από το θάνατο του Al-Adhnanī ή σε μια επιλογή της ηγεσίας της οργάνωσης να προχωρήσει σε αλλαγή προτεραιοτήτων αναφορικά με τα μέτωπα-αιχμές και τους αξιοποιούμενους πόρους. Ένα άλλο σημαντικό χαρακτηριστικό που απορρέει από την Εικόνα 17 είναι ότι μετά την 4^η έκδοση του Rumiyaḥ παρατηρείται μια ελαφρά αύξηση του αριθμού των λέξεων ανά έκδοση που, όμως ακολουθείται από μια περίοδο όπου εμφανίζονται τα ελάχιστα πλήθη των λέξεων ανά έκδοση (έκδοση 6^η, 7^η και 8^η). Το γεγονός αυτό ίσως να σχετίζεται με το γεγονός ότι εκείνη την περίοδο (Δεκέμβριος 2016) πραγματοποιούνταν σοβαρές μάχες στη Συρία (Gambhir, 2016).

Τα διαγράμματα που αφορούν όλα τα περιοδικά εμπεριέχονται στο παράρτημα Α και Β.

3.6.2 Ανάλυση της πολικότητας σε επίπεδο περιοδικού

Όπως έχει ήδη αναφερθεί, ένας από τους στόχους της έρευνας που διεξαγάγαμε, ήταν ο προσδιορισμός της πολικότητας σε επίπεδο περιοδικού, όπως αυτή διαμορφώνεται από τον συνυπολογισμό των επιμέρους πολικότητων των εκδόσεων. Τα διαγράμματα που ακολουθούν αναδεικνύουν ένα χαρακτηριστικό που, πιθανά, δεν ήταν αναμενόμενο. Όπως παρατηρεί ο αναγνώστης, ενώ κάποιος θα περίμενε στα περιοδικά αυτά να υπερισχύει η αρνητική πολικότητα έναντι της αντίστοιχης θετικής, η θετική και η αρνητική πολικότητα εμφανίζουν περίπου ίσες επιδόσεις (Εικόνα 19). Με εξαίρεση το Dabiq, σε όλα τα υπόλοιπα περιοδικά παρατηρείται ότι η τιμή της μέσης θετικής πολικότητας υπερισχύει ελαφρά της αντίστοιχης αρνητικής. Για το λόγο αυτό διατηρούμε και τις δύο πολικότητες και δεν έχουμε υιοθετήσει τη συνήθη πρακτική, σύμφωνα με την οποία υπολογίζεται η ολική πολικότητα μέσω της διαφοράς των δύο πολικότητων. Το εύρημα αυτό μας οδηγεί στην πεποίθηση ότι στα κείμενα αυτά σκόπιμα καλλιεργούνται αλληλοδιαδόχως συναισθήματα που είχαν και θετική και αρνητική πολικότητα. Για παράδειγμα, το αφήγημα του Ουτοπιανισμού μέσα στο έγγραφο συνδέεται με θετική πολικότητα, ενώ το αφήγημα της Βαρβαρότητας συνδέεται με αρνητική πολικότητα. (Τα στοιχεία της παραπάνω ανάλυσης βρίσκονται στα παραρτήματα Ε και ΣΤ).



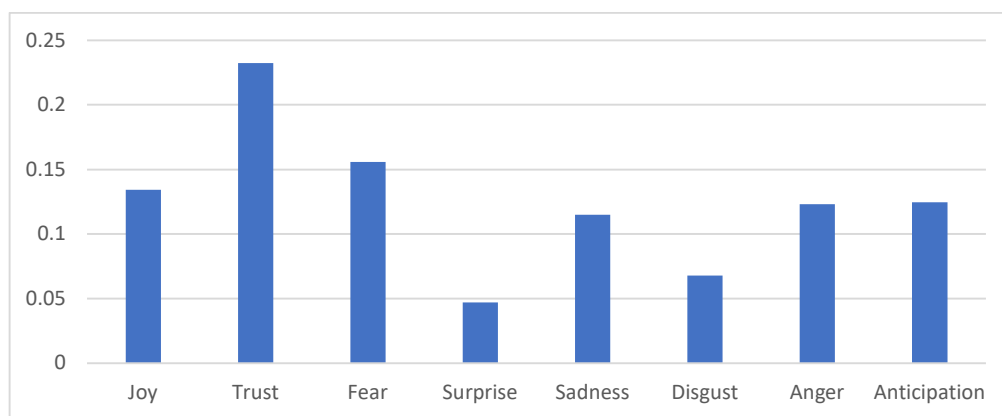
ΕΙΚΟΝΑ 19 ΜΕΣΗ ΠΟΛΙΚΟΤΗΤΑ ΑΝΑ ΠΕΡΙΟΔΙΚΟ ΥΠΟΛΟΓΙΣΜΕΝΗ ΣΕ ΟΛΕΣ ΤΙΣ ΕΚΔΟΣΕΙΣ ΤΟΥ.

3.6.3 Ανάλυση συναισθήματος σε επίπεδο έκδοσης περιοδικού

Κύριαρχο ρόλο στην ανάλυση μας αποτελούσε η αποτύπωση των συναισθημάτων που υπόβοσκαν στις διάφορες εκδόσεις των περιοδικών του ISIS. Η αποτύπωση αυτή θα

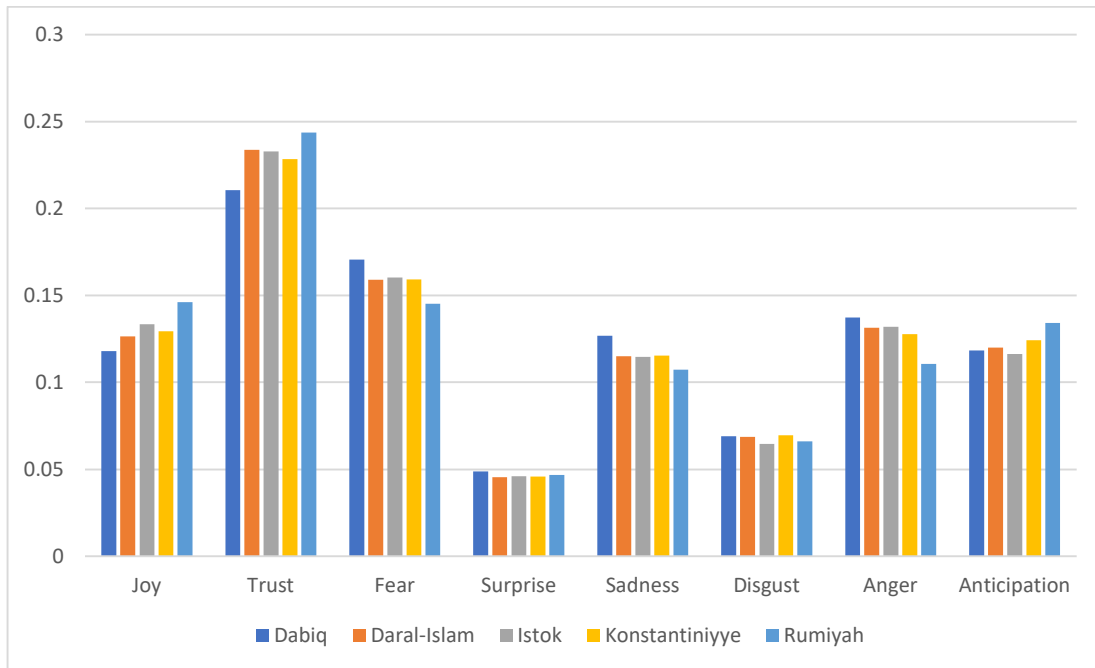
μπορούσε να μας αποκαλύψει αν τα περιοδικά μπορούσαν να κατηγοριοποιηθούν σε επιμέρους κατηγορίες ανάλογα με τα συναισθήματα που περιέκλειαν στα κείμενα της προπαγάνδας.

Βλέποντας την γενική εικόνα, όπως αυτή καταγράφεται από τον μέσο όρο της έντασης ανά συναίσθημα για όλες τις εκδόσεις των περιοδικών (corpus), καταλήγει κανείς σε ένα συνολικό συναισθηματικό πρότυπο που αφορά όλη τη συλλογή, όπως αυτό αναπαρίσταται στην Εικόνα 20. Αναλύοντας προσεκτικά το διάγραμμα της εικόνας αυτής, παρατηρείται κάτι εξαιρετικά ενδιαφέρον. Το συναίσθημα Εμπιστοσύνη (Trust) λαμβάνει μέση τιμή έντασης 0,232 και φαίνεται να κυριαρχεί των άλλων επτά βασικών συναισθημάτων σε όλη τη συλλογή. Το συναίσθημα Φόβος (Fear) καταλαμβάνει τη δεύτερη θέση στη συλλογή με μέση τιμή έντασης 0,156. Την τρίτη θέση καταλαμβάνει το συναίσθημα Χαρά (Joy) με μέση ένταση 0,134 ενώ την τέταρτη θέση μοιράζονται τα συναισθήματα Προσμονή (Anticipation) και Οργή (Anger) των οποίων οι τιμές της μέσης έντασης είναι αντίστοιχα 0,124 και 0,123. Ακολουθούν τα συναισθήματα Θλίψη (Sadness), Απέχθεια (Disgust) με τελευταίο το συναίσθημα Έκπληξη (Surprise). Όπως καταλαβαίνει κανείς από τη διενεργηθείσα ανάλυση, τα δύο τελευταία συναισθήματα είναι λιγότερο παρόντα στη συλλογή. Από τα παραπάνω συνάγεται ότι η προσέγγιση της ανάλυσης μέσω της οποίας αποτυπώνεται το φάσμα των συναισθημάτων που επικρατούν στα κείμενα της συλλογής, υπερτερεί της προσέγγισης που αποτυπώνει μόνο την πολικότητα των κειμένων, μιας και αναδεικνύει σημαντικότερα χαρακτηριστικά των αναλυόμενων κειμένων. Το γεγονός αυτό δεν προβάλλεται όσο θα έπρεπε τόσο σε ερευνητικές μελέτες όσο και εμπορικές εφαρμογές.



ΕΙΚΟΝΑ 20 ΜΕΣΟΣ ΑΝΑ ΣΥΝΑΙΣΘΗΜΑ ΥΠΟΛΟΓΙΣΜΕΝΟΣ ΣΕ ΟΛΗ ΤΗ ΣΥΛΛΟΓΗ

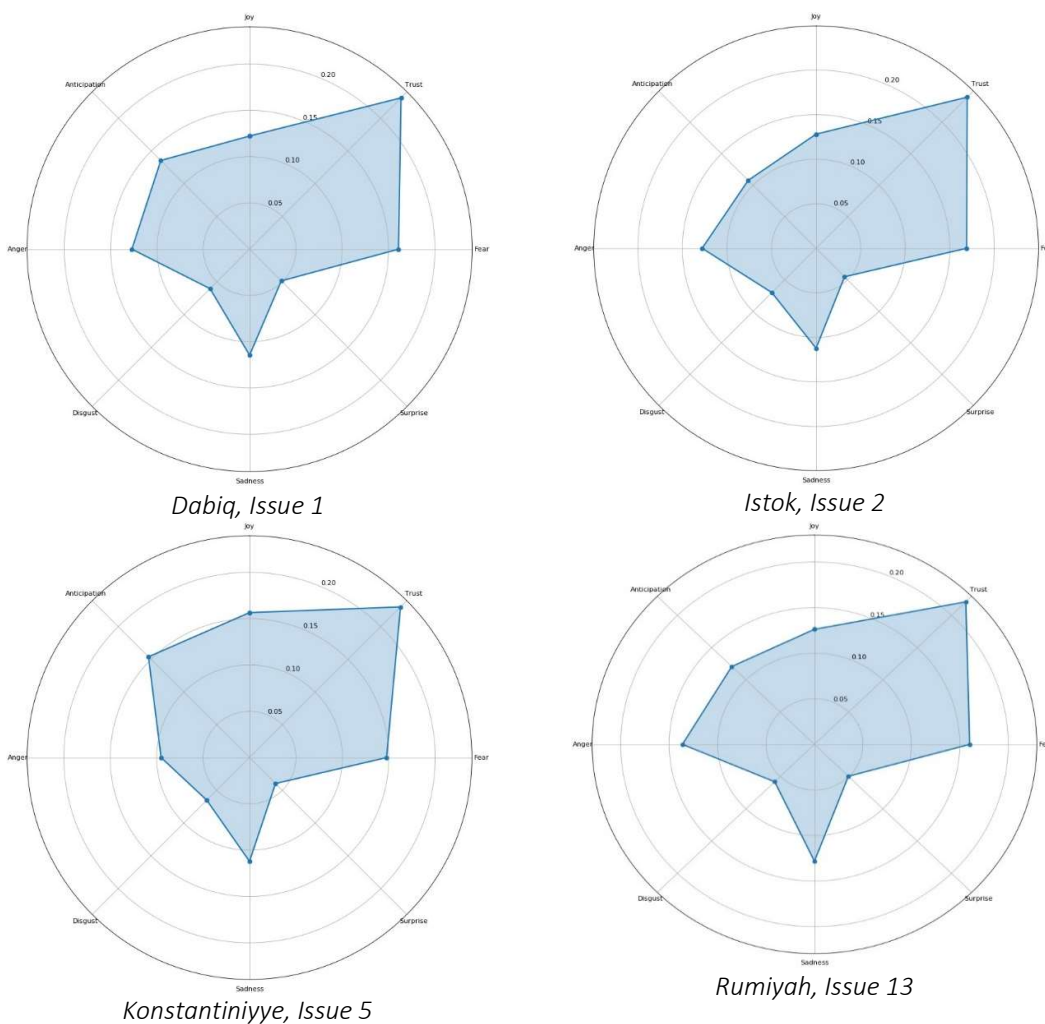
Το επόμενο βήμα της ανάλυσης μας είχε στόχο να εστιάσει στην καταγραφή του προτύπου (pattern) σε επίπεδο περιοδικού. Όπως αποτυπώνεται στο διάγραμμα που εμφανίζεται στην Εικόνα 21, η ένταση του εκάστοτε συναισθήματος σε κάθε περιοδικό παίρνει παραπλήσιες τιμές, με τις διαφορές τιμών να μην ξεπερνούν το 5%. Το γεγονός αυτό μας οδηγεί στο συμπέρασμα ότι υπάρχει ένα γενικό μορφότυπο συναισθημάτων μεταξύ των περιοδικών.



ΕΙΚΟΝΑ 21 ΜΕΣΗ ΕΝΤΑΣΗ ΑΝΑ ΣΥΝΑΙΣΘΗΜΑ ΣΕ ΕΠΙΠΕΔΟ ΠΕΡΙΟΔΙΚΟΥ

Στη συνέχεια, εστίασαμε σε κάθε έκδοση περιοδικού. Η απεικόνιση των μέσων τιμών έντασης των οκτώ συναισθημάτων ανά έκδοση περιοδικού πραγματοποιήθηκε μέσω της αξιοποίησης των αραχνοειδών διαγραμμάτων. Στις εικόνες που περιέχονται στον πίνακα 14, απεικονίζονται τα αραχνοειδή διαγράμματα τεσσάρων εκδόσεων, των περιοδικών Dabiq, Istok, Konstantiniyye και Rumiya. Είναι φανερό από τις εικόνες αυτές ότι, πέρα από τις ελαφρές διαφορές μεταξύ των μορφών στα αραχνοειδή διαγράμματα, οι εν λόγω εκδόσεις των περιοδικών μοιράζονται ένα κοινό μορφότυπο. Πλήρη αποτύπωση όλων των εκδόσεων των περιοδικών σε μορφή αραχνοειδών διαγραμμάτων περιέχεται στο παράρτημα ΣΤ.

ΠΙΝΑΚΑΣ 14 ΠΙΝΑΚΑΣ ΑΡΑΧΝΟΕΙΔΩΝ ΔΙΑΓΡΑΜΜΑΤΩΝ ΤΕΣΣΑΡΩΝ ΔΙΑΦΟΡΕΤΙΚΩΝ ΕΚΔΟΣΕΩΝ ΤΩΝ ΠΕΡΙΟΔΙΚΩΝ



3.6.4 Αποτελέσματα από την ανάλυση μέσω της LDA

Αναφορικά με την ανάλυση της συλλογής των κειμένων μέσω της LDA, ορίσαμε το πλήθος των θεμάτων (topics) σε 12 και το πλήθος των χαρακτηριστικών (features) σε 1600 (Kinney, et al., 2018). Στη συνέχεια εκπαιδεύσαμε το μοντέλο μέσω δέκα επαναλήψεων (Kohavi, 1995), (Bengio & Grandvalet, 2004). Ο αλγόριθμος παρήγαγε σε κάθε επανάληψη ως έξοδο συνεπή αποτελέσματα. Συνολικά, προσδιορίστηκαν μέσω της LDA εξήντα-οκτώ μοναδικές λέξεις. Αυτές παρατίθενται σε τυχαία σειρά στη συνέχεια:

'way', 'god', 'place', 'woman', 'right', 'imam', 'fear', 'allahin', 'east', 'group', 'among', 'order', 'good', 'fact', 'crusader', 'come', 'except', 'islamic', 'see', 'without', 'land', 'claim', 'saddle', 'mujahidin', 'messenger', 'city', 'khilafah', 'even', 'muslim', 'people', 'go', 'know', 'give', 'take', 'islam', 'cent', 'lord', 'brother', 'law', 'call', 'army', 'soldier', 'like', 'become', 'jihad', 'many',

'believer', 'battle', 'prophet', 'report', 'plus', 'time', 'religion', 'born', 'make', 'fight', 'force', 'indeed', 'honor', 'towards', 'kill', 'allah', 'state', 'sure', 'son', 'day', 'war', 'enemy'

Είναι προφανές ότι οι προσδιορισθείσες λέξεις διαδραμάτιζαν κυρίαρχο ρόλο στο προσδιορισμό της φήμης (brand) του ISIS. Εδώ αξίζει να θυμηθεί ο αναγνώστης ότι ο Winter (2015), σε μια πρώτη προσπάθεια αποδόμησης της φήμης του ISIS, είχε ισχυριστεί ότι: “Το αφήγημα του ISIS έχει τις ακόλουθες έξι συνιστώσες:

1. Βαρβαρότητα (Brutality)
2. Έλεος (Mercy)
3. Θυματοποίηση (Victimhood)
4. Πόλεμος (War)
5. Ανήκειν (Belonging)
6. Ουτοπισμός (Utopianism)

Οι παραπάνω λέξεις που προσδιορίστηκαν μέσω της LDA τεκμηριώνουν απόλυτα στις απόψεις του C. Winter. Όπως μπορεί να παρατηρήσει κανείς, υπάρχουν λέξεις που αναφέρονται σε περισσότερες από μια συνιστώσες του Winter. Για παράδειγμα:

- οι λέξεις “war”, “jihad⁴”, “army”, “battle”, “kill”, “fight”, “mujahidin⁵” and “soldier” αναφέρονται κυρίως στις συνιστώσες Πόλεμος (**War**) και Βαρβαρότητα (**Brutality**) ενώ
- λέξεις όπως “brother” και “son” συνεισφέρουν στο Ανήκειν (**Belonging**) και
- λέξεις όπως “state”, “land”, “Islamic⁶”, “khilafah⁷” στον Ουτοπισμό (**Utopianism**).

Η παραπάνω επιβεβαίωση της υπόθεσης αυτής, όπως είναι φανερό, πηγάζει τόσο από το περιεχόμενο των κειμένων των περιοδικών όσο και από τη μελέτη του Winter (2015) και από βίντεο που δημιουργήθηκαν από το al-Hayat Media Center, όπου απεικονίζονται μαχητές του ISIS να απολαμβάνουν ειρηνικές στιγμές μεταξύ συμμαχητών.

⁴ To strive/exert/fight (meaning) from جهاد (jihād)

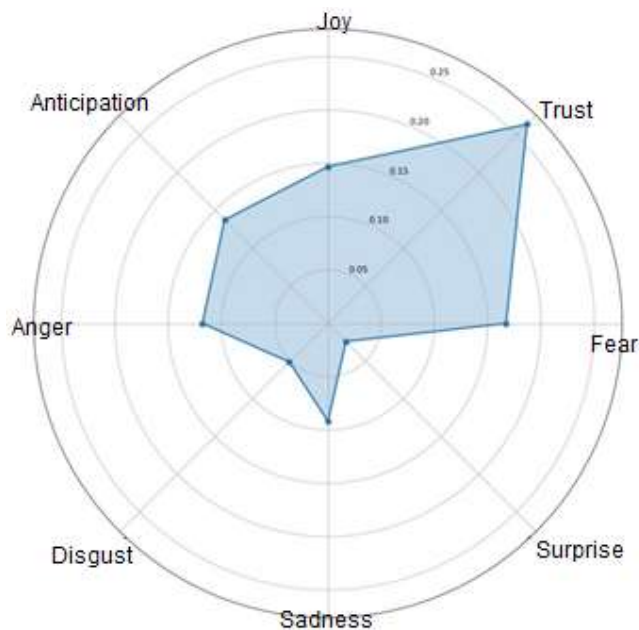
⁵ Someone that has engaged in Jihad (meaning) from مجاهدين (mujahidin)

⁶ Frequently appears as the bigram “Islamic State”.

⁷ Caliphate (meaning) from خلافة (khilāfah)

3.6.5 **Ανάλυση συναισθήματος στις λέξεις-θέματα (topic words) από την LDA**
 Την εξαγωγή των λέξεων-θεμάτων (topic words) ακολούθησε ο προσδιορισμός της έντασης των οκτώ βασικών συναισθημάτων με τα οποία συνδέονται αυτές. Το σημαντικό αποτέλεσμα του πειράματος που πραγματοποιήσαμε ήταν ότι τα μορφότυπα συναισθημάτων που προσδιορίζονταν σε κάθε επανάληψη της LDA, ταυτίζονταν και επιβεβαίωναν τα μορφότυπα των συναισθημάτων των περιοδικών. Επιπρόσθετα, το εύρημα αυτό έχει μεγάλη σημασία λόγω του ότι προσφέρει ένα μεθοδολογικό πλαίσιο για να δημιουργήσει κανείς μια φιλτραρισμένη (distilled) εκδοχή των συναισθηματικών μορφότυπων που εμπεριέχονται στη συλλογή των κειμένων (corpus) της προπαγάνδας. Όπως αναφέρθηκε στην παράγραφο 3.6.3, εμφανίζονται ελαφρές αποκλίσεις μεταξύ των μορφότυπων των περιοδικών, όπως αυτές απεικονίζονται και μέσω των αραχνοειδών διαγραμμάτων. Το γεγονός αυτό είναι αναμενόμενο επειδή η προσέγγιση εκείνη λάμβανε υπόψη της, χωρίς διάκριση, όλες τις λεξικογραφικές μονάδες (token words) ανά έκδοση περιοδικού. Αντίθετα, όπως καταδεικνύουμε, η συναισθηματική ανάλυση μέσω των θεματικών λέξεων που προσδιορίζει η LDA, περιορίζει την εμφάνιση του εν λόγω θορύβου, παράγει πιο ακριβή αποτελέσματα και προκρίνεται ως μια πιο συνεπής (robust) προσέγγιση.

Η εικόνα 22 απεικονίζει σε αραχνοειδές διάγραμμα το συναισθηματικό μορφότυπο των 68 μοναδικών λέξεων-θεμάτων που προσδιορίστηκαν μέσω της LDA. Πενήντα-τέσσερες εξ αυτών, σε σύνολο εξήντα-οκτώ, ανήκαν στο λεξικό EmoLex. Κατά συνέπεια, το εν λόγω λεξικό εμφάνισε κάλυψη 79%.



ΕΙΚΟΝΑ 22 ΑΡΑΧΝΟΕΙΔΕΣ ΔΙΑΓΡΑΜΜΑ ΓΙΑ ΟΛΕΣ ΤΙΣ ΛΕΞΕΙΣ-ΘΕΜΑΤΑ ΤΗΣ LDA

3.7 Συμπεράσματα

Ανακεφαλαιώνοντας, κατά τη διάρκεια της έρευνας που υλοποιήσαμε, προχωρήσαμε στη συλλογή όλων των γνωστών περιοδικών προώθησης της προπαγάνδας του ISIS. Στο υλικό

αυτό έγινε προεπεξεργασία, μετάφραση και επεξεργασία μέσω τεχνικών φυσικής γλώσσας (NLP). Το περιεχόμενο των κειμένων αναλύθηκε με σκοπό την ανάδειξη εντάσεων συναισθημάτων και πολικότητας. Τα αποτελέσματα οδήγησαν στον προσδιορισμό συναισθηματικών μορφοτύπων στο σύνολο των περιοδικών, ενώ κατέδειξαν ότι η ανάλυση των κειμένων μέσω αποτύπωσης των εντάσεων των συναισθημάτων υπερτερεί σαφώς αυτής της πολικότητας. Ένα σημαντικό συμπέρασμα που εξήχθη από την έρευνα αφορούσε τα κυρίαρχα συναισθήματα που προσπαθεί να αφυπνίσει η προπαγάνδα του ISIS. Σύμφωνα με τα αποτελέσματα μας, τα τρία κυρίαρχα συναισθήματα που προσπαθεί να αφυπνίσει η προπαγάνδα του ISIS στον αναγνώστη των περιοδικών είναι, με βάση την εξετασθείσα συλλογή, και με σειρά μειούμενης έντασης, η Εμπιστοσύνη (Trust), ο Φόβος (Fear), η Χαρά (Joy). Αξιοποιώντας την μέθοδο προσδιορισμού θεμάτων LDA, προσδιορίστηκαν οι λέξεις-θέματα που ενυπάρχουν στη συλλογή. Οι λέξεις αυτές αντιπροσώπευαν πλήρως τις έξι συνιστώσες της φήμας του ISIS, όπως αυτές έχουν αποτυπωθεί από τον Winter (Winter, 2015). Η ανάλυση των συναισθημάτων των λέξεων-θεμάτων παρείχε έναν εναλλακτικό και ακριβέστερο πλαίσιο εύρεσης των συναισθηματικών μορφοτύπων της συλλογής. Το κρίσιμο πλεονέκτημα αυτού του μεθοδολογικού πλαισίου είναι ότι παράγει αποτελέσματα απαλλαγμένα από το θόρυβο που προκαλεί στην αναπαράσταση των συναισθηματικών μορφοτύπων, μέσω όλων των λεξικογραφικών μονάδων.

Συγκεκριμένα, οι απαντήσεις που προκύπτουν από τα έως τώρα αποτελέσματα, έχουν ως εξής:

RQ1. Αποδεικνύουμε ότι η μέθοδός μας προσφέρει περισσότερη λεπτομέρεια και πληροφορία, ειδικά σε σύγκριση με την ανάλυση πολικότητας, εξάγοντας και αποτυπώνοντας επιτυχώς τα συναισθήματα που οι δημιουργοί της προπαγάνδας του ISIS είχαν σκοπό να προκαλέσουν στα υποκείμενα που ανήκαν στο κοινό-στόχος.

RQ2. Καταφέρνουμε να δούμε με ακρίβεια τα συναισθήματα και τους συνδυασμούς τους, γεγονός που βελτιώνει τον βαθμό κατανόησής μας για την προπαγάνδα. Τα ευρήματα αυτά είναι κλάσεις ανώτερα συγκριτικά με παλαιότερες μελέτες που περιορίζονταν στην ανάλυση πολικότητας.

RQ3. Αναδεικνύουμε την ύπαρξη υποβοσκόντων συναισθηματικών και ιδεολογικών προτύπων στα οποία ανήκαν τα υποκείμενα του ακροατηρίου-στόχου.

RQ4. Παρατηρείται συγκεκριμένο συναισθηματικό πλαίσιο, το οποίο ακολουθείται από τους συντάκτες ανεξάρτητα από το ακροατήριο.

RQ5. Εξορύξαμε επιτυχώς τα διαρκή θέματα (topics) συζήτησης τα οποία επιλέγονται από τους δημιουργούς της προπαγάνδας του ISIS.

Πιστεύουμε ότι τα συμπεράσματα της εν λόγω έρευνας αποτελούν σημαντική συνεισφορά στην βιβλιογραφία που είναι σχετική με τον Πληροφοριακό Πόλεμο, καθώς το μεθοδολογικό πλαίσιο αναδεικνύει τον τρόπο με τον οποίο υπάρχουντα και αξιόπιστα εργαλεία και τεχνικές μπορούν να εφαρμοστούν σε τέτοιες συνθήκες προσφέροντας νέες προσεγγίσεις επίλυσης των υπαρχόντων προβλημάτων.

4 Ανίχνευση ανωμαλιών με τη χρήση μη εποπτευόμενης Μηχανικής Μάθησης και βάση το συναίσθημα

4.1 Εισαγωγή

Στην προηγούμενη ενότητα έγινε εκτενής ανάλυση των περιοδικών του ISIS και των συναισθημάτων που αυτά προσπαθούν να εγείρουν στον αναγνώστη. Τα σημαντικότερα ευρήματα ήταν:

1. Επιβεβαίωση των θεμάτων (themes) του brand των ISIS, σύμφωνα και με προηγούμενες έρευνες,
2. Η Εμπιστοσύνη ήταν το πιο έντονο πρωτογενές συναίσθημα εντός της προπαγάνδας,
3. Επιτυχής εξαγωγή του συναισθηματικού προτύπου, κοινό ανά τις εκδόσεις και ανεξάρτητο της γλώσσας ή του κοινού-στόχος,
4. Περαιτέρω βελτιστοποίηση του συναισθηματικού προτύπου, μέσω του συνδυασμού με LDA.

Με βάση τα σημεία αυτά, συμπεραίνουμε ότι όντως, υπάρχει πλούτος πληροφορίας σε επίπεδο συναισθήματος, ειδικά σε ένα κείμενο φορτισμένο, όπως αυτά των ISIS. Ταυτόχρονα, η μέθοδος εξαγωγής των συναισθημάτων που πραγματοποιήσαμε ήταν ακριβής, ειδικά εάν λάβουμε υπόψη:

1. Προηγούμενη έρευνα περί των θεματικών συζήτησης του ISIS,
2. Την διασταυρωμένη επικύρωση μεταξύ των συναισθημάτων που προέκυψαν από τα κείμενα συνολικά, σε σύγκριση με αυτά που εξάχθηκαν από τα αποτελέσματα της LDA.

Στην παρούσα ενότητα επεκτείνουμε το σύνολο των δεδομένων πάνω στα οποία γίνεται η ανάλυση. Τα περιοδικά του ISIS διαιρούνται ανά άρθρο και προστίθενται επιπλέον άρθρα από γνωστά ειδησεογραφικά πρακτορεία δυτικών χωρών. Κάθε άρθρο μετατρέπεται σε διάνυσμα οχτώ διαστάσεων, με κάθε διάσταση να αντιπροσωπεύει ένα από τα οχτώ πρωτογενή συναισθήματα. Στη συνέχεια γίνεται χρήση διαφόρων μεθόδων μη εποπτευόμενης μηχανικής μάθησης, με σκοπό τον εντοπισμό ανωμαλιών, όπως αυτές προκύπτουν βάσει του συναισθηματικού προφίλ του εκάστοτε άρθρου. Με άλλα λόγια, επιχειρείται δυαδική κατηγοριοποίηση μεταξύ άρθρων των οποίων το συναισθηματικό προφίλ θεωρείται κανονικό, ή ακραίο.

Ερευνητικές ερωτήσεις που επιχειρούμε να απαντήσουμε είναι:

1. Υπάρχουν διαφορές στα συναισθηματικά πρότυπα που προβάλλουν οι ISIS, και σε αυτά που προβάλλουν τα δυτικά ειδησιογραφικά πρακτορεία;
2. Εάν υπάρχουν διαφορές, ποιες είναι αυτές; Ποια είναι τα κύρια συναισθήματα στα άρθρα των ISIS και ποια στα άρθρα της δύσης;
3. Υπάρχουν μήπως ομοιότητες; Αν ναι, ποιες είναι αυτές; Υποδεικνύουν πιθανά κοινές επικοινωνιακές τακτικές;
4. Είναι δυνατό να εντοπίσουμε ανώμαλα άρθρα βάσει των συναισθηματικών διανυσμάτων τους, με τη χρήση μοντέλων μηχανικής μάθησης, ειδικά όταν το μοντέλο δεν έχει ξαναδεί ανώμαλα δεδομένα;

Θεωρούμε ότι οι ερωτήσεις αυτές είναι κρίσιμες και οι απαντήσεις τους θα συμβάλουν τα μέγιστα στη βιβλιογραφία του τομέα.

4.2 Συλλογή Δεδομένων

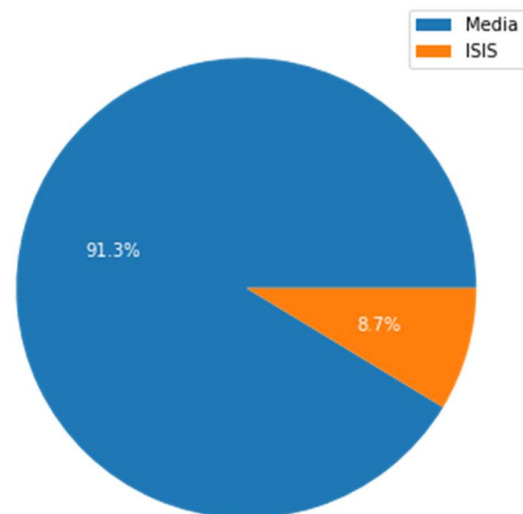
Έχει γίνει ήδη αναφορά στον τρόπο εύρεσης και συλλογής των περιοδικών του ISIS. Στο μέρος αυτό της έρευνας θα λάβουμε υπόψη μόνο τα αγγλόφωνα περιοδικά Dabiq και Rumiyah. Για τον εμπλουτισμό των δεδομένων μας, χρησιμοποιήσαμε το “News Articles Dataset” το οποίο συλλέχθηκε και αναρτήθηκε από τον Tianyu Dai και είναι διαθέσιμο διαδικτυακά από το Harvard Dataverse (Dai, 2017). Το δείγμα αυτό αποτελείται από 3.824 άρθρα και ενδείκνυται για έρευνες με αντικείμενο το “media bias”. Αυτό το κάνει καλό και ενδιαφέρον συμπλήρωμα στο υπάρχον σύνολο δεδομένων που εμπεριέχουν προπαγάνδα. Κάθε ειδησεογραφικό άρθρο χαρακτηρίζεται από την ημερομηνία δημοσίευσης, τον τίτλο και τον υπότιτλο, το

κύριο κείμενο του άρθρου και την πηγή. Πρακτορεία που περιλαμβάνονται είναι τα ABC News, CNN News, The Huffington Post, BBC News, DW News, TASS News, Al Jazeera News, China Daily και RTE News. Όλα έχουν συλλεγεί μέσω των RSS feeds του κάθε ιστοτόπου, για μία περίοδο τριών μηνών (Δεκέμβριος 2016 έως Μάρτιος 2017). Είναι σημαντικό να σημειωθεί ότι η περίοδος αυτή καλύπτει σπουδαία πολιτικά γεγονότα στις ΗΠΑ.

4.3 Προετοιμασία Δεδομένων

Αφού κάθε παρατήρηση (κείμενο) στο News Articles Dataverse είναι σε επίπεδο άρθρου, έπρεπε να φέρουμε τα κείμενα των ISIS στο ίδιο επίπεδο. Αν και αυτό δεν είναι απαραίτητο, μιας και η κύρια μέθοδός μας αδιαφορεί για το μήκος του κειμένου, εξυπηρετεί ως προς τον διαχωρισμό των θεμάτων που πραγματεύεται το εκάστοτε άρθρο. Αυτό μας επιτρέπει να έχουμε πιο ξεκάθαρες συναισθηματικές κανονικότητες, ανάλογα το θέμα συζήτησης. Το αποτέλεσμα αυτής της διαίρεσης ανά άρθρο μας δίνει 363 παρατηρήσεις από τους ISIS, συν τις 3.824 του Dataverse, για συνολικά

4.187 άρθρα (ποσοστά στην εικόνα 23). Από εδώ και στο εξής, θα αναφερόμαστε στα άρθρα των ISIS ως “ISIS” και σε αυτά των δυτικών ειδησεογραφικών που περιέχονται στο Dataverse ως “Media”, χάριν ευκολίας. Η εικόνα 23 προσφέρει μία πρώτη γενική εικόνα της δομής των δεδομένων.



ΕΙΚΟΝΑ 23 ΠΟΣΟΣΤΟ ΑΡΘΡΩΝ MEDIA ENANTI ISIS

Παρατηρούμε ότι τα άρθρα του ISIS αποτελούν λιγότερο από 10% του συνόλου (Εικόνα 22). Παρόλο που αυτό θα ήταν πρόβλημα σε άλλες περιπτώσεις που απαιτούν ισορροπημένο dataset (π.χ. sentiment analysis), στην παρούσα περίπτωση της ανίχνευσης ανωμαλιών, αυτό είναι αποδεκτό και πολλές φορές, επιδιωκόμενο και ρεαλιστικό (Zuang & Dai, 2006; Thudumu, et al., 2020; Seliya, et al., 2021; Kumar & Ravi, 2017; Schmitt & Spinoso, 2018). Σημειώνεται ότι άρθρα του ISIS τα οποία περιέχουν αποκλειστικά και μόνο φωτομοντάζ, ή infographics, δεν έχουν συμπεριληφθεί καθώς δεν υπάρχει κείμενο. Τέλος, προστίθεται πραγματικό $\gamma = -1$ για τα άρθρα του ISIS και πραγματικό $\gamma = +1$ για τα άρθρα των Media, υπό

την αφελή υπόθεση ότι τα προπαγανδιστικά άρθρα του ISIS αποτελούν τις ανώμαλες παρατηρήσεις.

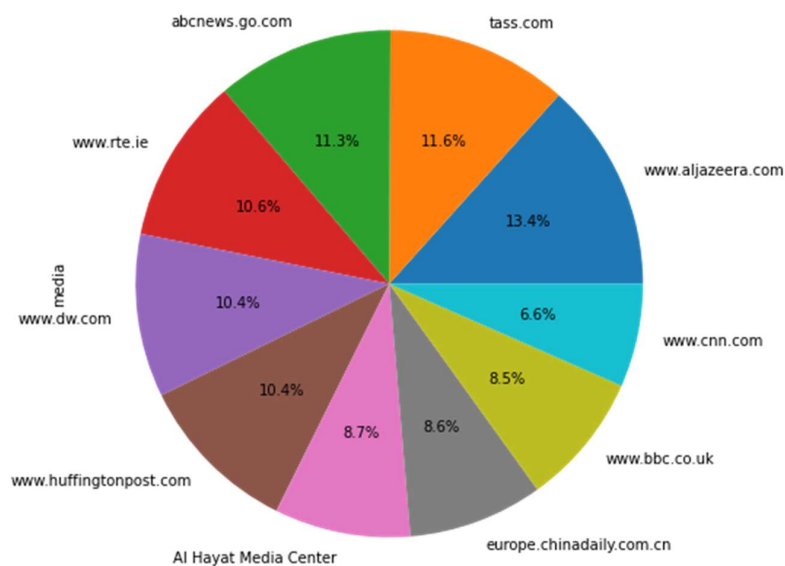
Με εξαίρεση τα άρθρα του ISIS τα οποία ήδη έχουν το Al-Hayat Media Center ως το πρακτορείο, εξάγουμε το κύριο όνομα ιστοτόπου (domain name) από το σύνδεσμο πηγής (source url). Αυτό μας δίνει μια αναλυτικότερη εικόνα της διάρθρωσης των δεδομένων (Πίνακας 15).

ΠΙΝΑΚΑΣ 15 ΆΡΘΡΑ ΑΝΑ ΠΗΓΗ

Πρακτορείο	Άρθρα
www.aljazeera.com	559
tass.com	485
abcnews.go.com	474
www.rte.ie	443
www.dw.com	436
www.huffingtonpost.com	436
europe.chinadaily.com.cn	360
www.bbc.co.uk	355
www.cnn.com	276
Al-Hayat Media Center	363
Σύνολο	4.187

Το επόμενο διάγραμμα πίτας περιγράφει τη δομή του dataset με ποσοστά (Εικόνα 24). Όπως φαίνεται, περιέχει ένα αντιπροσωπευτικό δείγμα άρθρων από το κάθε πρακτορείο.

ΕΙΚΟΝΑ 24 ΠΟΣΟΣΤΟ ΑΡΘΡΩΝ ΑΝΑ ΠΗΓΗ



4.4 Προεπεξεργασία

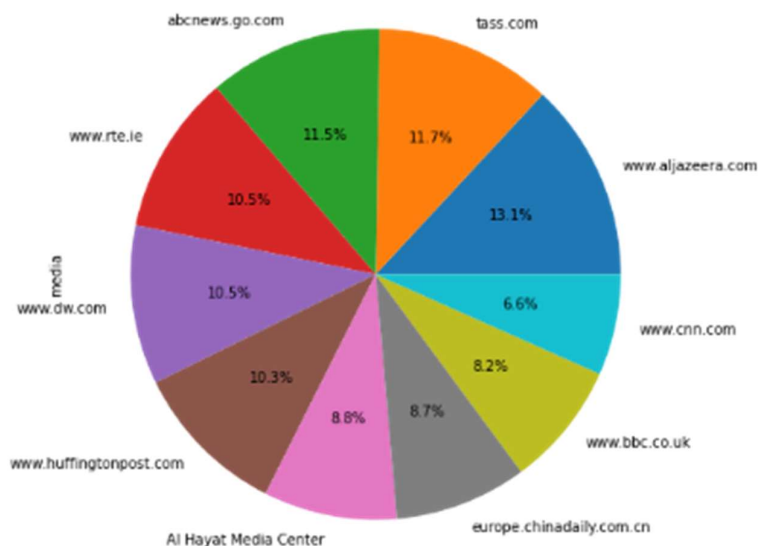
Όλα τα άρθρα υποβλήθηκαν σε κατεργασία με τις ευρέως αποδεκτές τεχνικές επεξεργασίας φυσικής γλώσσας. Ομοίως όπως περιγράφεται στο κεφάλαιο 3.5.3, κάθε άρθρο χωρίστηκε σε λέξεις (tokens), σημεία στίξης και αριθμοί αφαιρέθηκαν, όπως επίσης και κοινές λέξεις (stopwords). Οι εναπομείναντες λέξεις μετασχηματίστηκαν στα λήμματά τους, αφού πρώτα λήφθηκε υπόψη η συντακτική τους θέση (part of speech tagging). Για παράδειγμα, οι λέξεις “running” (ρήμα) και “runner” (υποκείμενο ή αντικείμενο) μετασχηματίστηκαν σε “run” και “runner” αντίστοιχα, αντί σε “run” (ρήμα) και οι δύο.

Στη συνέχεια, έγινε μέτρηση των λέξεων ανά άρθρο και υπολογίστηκαν τα βασικά μέτρα περιγραφικής στατιστικής, ανά πρακτορείο. Μέσα από αυτή τη διαδικασία εντοπίστηκαν άρθρα τα οποία είχαν πολύ μικρό αριθμό λέξεων, ή μηδέν λέξεις. Ύστερα από χειροκίνητη αναθεώρηση διαπιστώθηκε ότι ένα μικρό πλήθος άρθρων είχαν όντως λίγες ή μηδέν λέξεις, για έναν από τους παρακάτω λόγους:

- Ανεπιτυχής προσπάθεια και αποθήκευση του άρθρου,
- Το άρθρο είναι φωτομοντάζ,
- Το άρθρο αφορά ανάρτηση βίντεο,
- Το άρθρο αποτελεί σύντομη δημοσίευση έκτακτης είδησης.

Λαμβάνοντας υπόψη τα στατιστικά στοιχεία που αναφέρθηκαν προηγουμένως, αποφασίστηκε εμπειρικά ένα κατώτατο όριο αριθμού λέξεων, το οποίο κάθε άρθρο έπρεπε να πληρεί ώστε να συμπεριληφθεί στην έρευνα (αριθμός λέξεων ≥ 30).

Όπως φαίνεται στο διάγραμμα πίτας (Εικόνα 25), η επέμβαση αυτή δεν επηρέασε αρνητικά τις αναλογίες αντιπροσώπευσης κάθε πρακτορείου. Τα αναλυτικά αποτελέσματα αυτού του βήματος είναι διαθέσιμα στο Παράρτημα Ι.



ΕΙΚΟΝΑ 25 ΠΟΣΟΣΤΟ ΑΡΘΡΩΝ ΑΝΑ ΠΗΓΗ ΜΕΤΑ ΤΗΝ ΑΦΑΙΡΕΣΗ ΑΡΘΡΩΝ ΜΕ <30 ΛΕΞΕΙΣ

4.5 Εξαγωγή Συναισθήματος

Όπως πριν, για την εξαγωγή των συναισθημάτων χρησιμοποιήθηκε το λεξικό EmoLex. Κάθε λίστα λέξεων ανά άρθρο, αντιστοιχίστηκε σε επίπεδο λέξης με τα οχτώ συναισθήματα που σχετίζονται με αυτή, βάσει του λεξικού. Με αυτό τον τρόπο οι λίστες λέξεων μετασχηματίστηκαν σε λίστες οχταδιάστατων διανυσμάτων. Στη συνέχεια προστέθηκαν τα συναισθήματα κάθε λέξης μεταξύ τους σε επίπεδο συνιστώσας (π.χ. Sum Trust, Sum Joy, Sum Fear, κλπ). Τα αθροίσματα διατηρήθηκαν και σε νέες στήλες υπολογίστηκαν τα κανονικοποιημένα συναισθήματα ανά άρθρο, όπως εξηγείται στο κεφάλαιο 3.6.3, με μόνη

διαφορά μία μικρή προσθήκη στον αλγόριθμο. Η προσθήκη αυτή ουσιαστικά, απλά πολλαπλασιάζει τα κανονικοποιημένα διανύσματα με τον αριθμό 100, ώστε να τα μεταφέρει από την κλίμακα [0, 1] στην κλίμακα [0, 100]. Αυτό έγινε για να αντιμετωπιστεί ένα πρόβλημα κατά τον υπολογισμό της απόστασης Mahalanobis, όπου πολύ μικρές τιμές είχαν ως αποτέλεσμα την παραγωγή αρνητικών αποστάσεων.

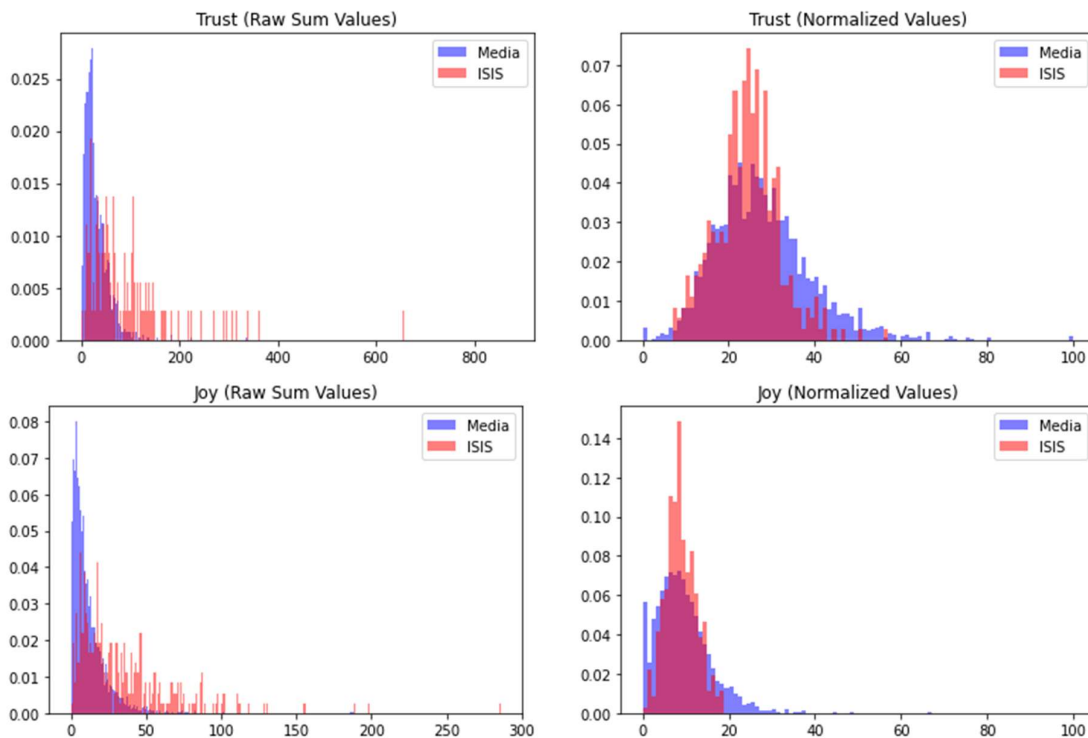
Ο λόγος που αυτή τη φορά κρατούνται και οι δύο μορφές διανυσμάτων (Sum και Normalised) είναι ώστε να εξεταστούν και τα δύο αυτά σενάρια ως features και να γίνει επιλογή του καλύτερου. Στην ανίχνευση ανωμαλιών προτιμάται πάντα η αποφυγή μετατροπής των δεδομένων στην ίδια κλίμακα, καθώς αυτό συμπιέζει κανονικές και ακραίες τιμές σε ένα κοινό εύρος. Αυτό έχει παρατηρηθεί ότι βλάπτει την αποτελεσματικότητα των μοντέλων μηχανικής μάθησης, καθώς η συμπίεση αυτή κάνει τις ακραίες τιμές δυσκολότερα εντοπίσιμες (Thudumu, et al., 2020; Seliya, et al., 2021). Ταυτόχρονα όμως, η έρευνά μας εστιάζει στην κατηγοριοποίηση με βάση συναισθηματικές κανονικότητες. Δηλαδή, παρατηρούνται διαφορετικές κανονικότητες μεταξύ των άρθρων; Υπάρχουν κανονικότητες που παρουσιάζονται σε έναν αριθμό άρθρων, άρα δημιουργούν μία κλάση συναισθηματικού προφίλ; Για να απαντήσουμε σε αυτές τις ερωτήσεις πρέπει τα συναισθήματα κάθε άρθρου να είναι στην ίδια κλίμακα με τα υπόλοιπα. Με αυτό τον τρόπο μπορούμε να βασιστούμε σε ποσοστά αναλογιών. Για παράδειγμα, όπως είδαμε στην προηγούμενη ενότητα οι συναισθηματικές αναλογίες ήταν πάντα ίδιες σε κάθε περιοδικό των ISIS. Εάν δεν γίνει κανονικοποίηση των διανυσμάτων, τότε το μοντέλο μηχανικής μάθησης υποχρεούται να χαρακτηρίσει ως ανώμαλα, τα άρθρα που πιθανά έχουν μεγάλα (ή μικρά) αθροίσματα εξαιτίας της έκτασης του κειμένου (περισσότερες λέξεις σε ένα άρθρο, σημαίνει εν δυνάμει περισσότερο συναίσθημα συνολικά). Έτσι, μέσω της κανονικοποίησης, το μοντέλο μπορεί να εστιάσει στις αναλογίες, αντί στον όγκο συναισθημάτων.

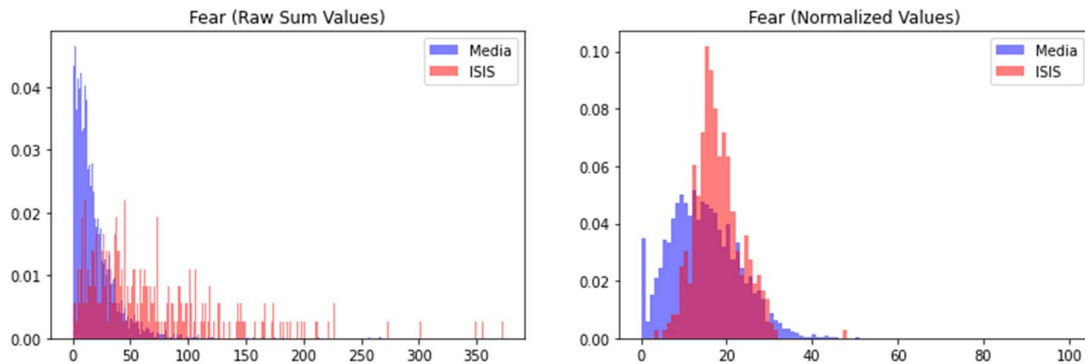
4.6 Διερευνητική Ανάλυση

Έχοντας εξάγει τα συναισθήματα από κάθε άρθρο και με τις δύο μεθόδους (αθροιστικά και κανονικοποιημένα) προχωρήσαμε σε διερευνητική ανάλυση. Σκοπός ήταν να βρούμε τυχόν συναισθηματικές κανονικότητες σε Media και ISIS, ομοιότητες μεταξύ ψυχολογικών προτύπων, κυρίαρχα συναισθήματα και συνδιακυμάνσεις αυτών με τα υπόλοιπα. Αποτυπώνοντας διαγραμματικά το σχετικό ιστόγραμμα ανά συναίσθημα, για τις αθροιστικές και κανονικοποιημένες τιμές, παρατηρούμε:

- Τα κανονικοποιημένα συναισθήματα παρουσιάζουν λιγότερο «θόρυβο», σε αντίθεση με τα αθροιστικά όπου υπάρχουν έντονες διακυμάνσεις στην κατανομή.
- Παρά το μικρότερο πλήθος των άρθρων του ISIS, η κατανομή των κανονικοποιημένων συναισθημάτων είναι εμφανώς πιο συνεπής.
- Στα αθροιστικά συναισθήματα βλέπουμε ότι όπως αναμέναμε, υπάρχουν σημαντικές διαφορές σε επίπεδο εύρους τιμών, το οποίο θα θεωρηθεί ως ακραία τιμή από τα μοντέλα.
- Τα κανονικοποιημένα συναισθήματα μας δίνουν πολύτιμη πληροφορία αναφορικά με τις κατανομές τους βάσει ISIS και Media, καθώς και τα κοινά διαστήματα μεταξύ τους (overlap – μοβ χρώμα).

Παρακάτω (Εικόνα 26) εξετάζουμε τρία εκ των οχτώ συναισθημάτων, ξεκινώντας από την Εμπιστοσύνη (Trust), η οποία ήταν το κυρίαρχο συναίσθημα σύμφωνα με τα αποτελέσματα της προηγούμενης ενότητας. Ακολουθούν η Χαρά (Joy) και ο Φόβος (Fear). Στην αριστερή στήλη είναι τα συναισθήματα στην αθροιστική μορφή και δεξιά στην κανονικοποιημένη.





ΕΙΚΟΝΑ 26 ΙΣΤΟΓΡΑΜΜΑΤΑ ΙΣΧΥΟΣ ΤΩΝ ΤΡΙΩΝ ΚΥΡΙΩΝ ΣΥΝΑΙΣΘΗΜΑΤΩΝ (ΑΘΡΟΙΣΤΙΚΑ ΑΡΙΣΤΕΡΑ, ΚΑΝΟΝΙΚΟΠΟΙΗΜΕΝΑ ΔΕΞΙΑ)

Είναι φανερή η έμφαση που δίνουν και οι δύο πλευρές στα παραπάνω συναισθήματα. Επίσης το overlap των κατανομών των κανονικοποιημένων τιμών, είναι μία πρώτη επιβεβαίωση ότι όντως, ISIS και Media δίνουν όμοιο βάρος σε αυτά. Μία διαφοροποίηση παρατηρείται στον Φόβο, όπου οι ISIS τείνουν να χρησιμοποιούν αυτό το συναίσθημα περισσότερο (overlap στο δεξιό μέρος της κατανομής των Media). Οι έντονες κορυφές που παρατηρούνται στις περιπτώσεις των ISIS υποδηλώνουν την αυστηρή ακολουθία συγκεκριμένων συναισθηματικών προτύπων – το οποίο είναι εύρημα και της προηγούμενης ερευνητικής εργασίας. Όλα τα συναισθήματα βρίσκονται στα παραρτήματα Η και Θ, όπου υπάρχει και αναλυτική απεικόνιση θηκογραμμάτων.

Στη συνέχεια υπολογίσαμε και απεικονίσαμε τις μήτρες συνδιακύμανσης. Αυτό το κάναμε για τα αθροιστικά συναισθήματα σε Media και ISIS, και για κανονικοποιημένα συναισθήματα σε Media και ISIS. Κοιτώντας αρχικά τις μήτρες για τα αθροιστικά, βλέπουμε ότι τα δύο κυρίαρχα συναισθήματα με σειρά φθίνουσας ισχύος, είναι η Εμπιστοσύνη και ο Φόβος. Αυτό φαίνεται από το γεγονός ότι η Εμπιστοσύνη δείχνει να παρουσιάζεται συχνά και να έχει δυνατή σχέση διακύμανσης με όλα τα υπόλοιπα συναισθήματα. Όμοια, το επόμενο πιο ισχυρό αίσθημα είναι ο Φόβος, όπου βλέπουμε λιγότερες μεν, αλλά ισχυρές συσχετίσεις με συγκεκριμένα άλλα συναισθήματα.

Συγκεκριμένα, για τα Media (Πίνακας 16), βλέπουμε ότι η Εμπιστοσύνη έχει μεγάλο βαθμό συσχέτισης με την Προσμονή, τη Χαρά και την Έκπληξη. Ενώ ο Φόβος με το Θυμό και τη Λύπη.

ΠΙΝΑΚΑΣ 16 ΣΥΝΔΙΑΚΥΜΑΝΣΗ ΑΘΡΟΙΣΤΙΚΩΝ ΣΥΝΑΙΣΘΗΜΑΤΩΝ MEDIA

<i>Sum Emotion</i>	Joy	Trust	Fear	Surprise	Sadness	Disgust	Anger	Anticip
Joy	129.19	209.71	85.37	70.47	68.39	33.11	62.92	142.36
Trust	209.71	593.46	272.85	151.55	167.78	86.88	194.25	301.46
Fear	85.37	272.85	356.05	85.61	193.94	88.46	245.61	160.78
Surprise	70.47	151.55	85.61	71.86	60.71	29.38	67.62	93.13
Sadness	68.39	167.78	193.94	60.71	144.47	58.17	142.12	110.21
Disgust	33.11	86.88	88.46	29.38	58.17	41.18	68.3	53.37
Anger	62.92	194.25	245.61	67.62	142.12	68.3	201.34	115.16
Anticipation	142.36	301.46	160.78	93.13	110.21	53.37	115.16	221.75

Σε παρόμοιο μοτίβο (Πίνακας 17), οι ISIS δείχνουν ότι βασίζονται πολύ στην εμπιστοσύνη, συνοδεύοντάς τη κάθε φορά με το «ακόλουθο» συναίσθημα που θέλουν να εγείρουν στον αναγνώστη. Επίσης, ο Φόβος παίζει σημαντικό ρόλο ως το δεύτερο κυρίαρχο συναίσθημα, ακολουθούμενος από το Θυμό, τη Λύπη, αλλά και την Εμπιστοσύνη.

ΠΙΝΑΚΑΣ 17 ΣΥΝΔΙΑΚΥΜΑΝΣΗ ΑΘΡΟΙΣΤΙΚΩΝ ΣΥΝΑΙΣΘΗΜΑΤΩΝ ISIS

<i>Sum Emotion</i>	Joy	Trust	Fear	Surprise	Sadness	Disgust	Anger	Anticip
Joy	1092.67	2720.48	1479.74	457.50	912.05	624.39	1031.64	1270.64
Trust	2720.48	8113.45	4299.76	1223.70	2565.01	1798.42	3004.03	3422.13
Fear	1479.74	4299.76	3598.21	949.26	2011.13	1095.81	2624.96	2061.91
Surprise	457.50	1223.70	949.26	314.01	559.37	293.00	696.28	627.53
Sadness	912.05	2565.01	2011.13	559.37	1278.01	691.28	1487.49	1225.20
Disgust	624.39	1798.42	1095.81	293.00	691.28	569.19	815.87	795.37
Anger	1031.64	3004.03	2624.96	696.28	1487.49	815.87	2050.29	1470.04
Anticipation	1270.64	3422.13	2061.91	627.53	1225.20	795.37	1470.04	1710.76

Ένα ακόμα ενδιαφέρον εύρημα προκύπτει από τις συνδιακυμάνσεις των κανονικοποιημένων συναισθημάτων. Στην περίπτωση των Media (Πίνακας 18) παρατηρείται ξανά η δυνατή σχέση μεταξύ Εμπιστοσύνης και Χαράς, Προσμονής. Παρουσιάζεται όμως και ένα σύμπλεγμα αρνητικών συναισθημάτων (Φόβος, Λύπη, Αηδία, Θυμός) που δείχνουν να συνυπάρχουν ή/και να είναι αλληλεξαρτώμενα. Το σύμπλεγμα αυτό έχει επισημανθεί με κόκκινο στον σχετικό πίνακα παρακάτω.

ΠΙΝΑΚΑΣ 18 ΣΥΝΔΙΑΚΥΜΑΝΣΗ ΚΑΝΟΝΙΚΟΠΟΙΗΜΕΝΩΝ ΣΥΝΑΙΣΘΗΜΑΤΩΝ MEDIA

<i>Norm Emotion</i>	Joy	Trust	Fear	Surprise	Sadness	Disgust	Anger	Anticip
Joy	38.06	7.71	-28.41	4.43	-12.1	-5.26	-19.42	14.99
Trust	7.71	123.68	-47.3	-4.81	-33.81	-15.12	-40.53	10.18
Fear	-28.41	-47.3	63.94	-10.14	16.44	5.88	30.86	-31.28
Surprise	4.43	-4.81	-10.14	21.49	-2.18	-1.74	-6.34	-0.72
Sadness	-12.1	-33.81	16.44	-2.18	30.92	6.16	11.32	-16.74
Disgust	-5.26	-15.12	5.88	-1.74	6.16	11.75	7.18	-8.85
Anger	-19.42	-40.53	30.86	-6.34	11.32	7.18	41.15	-24.22
Anticipation	14.99	10.18	-31.28	-0.72	-16.74	-8.85	-24.22	56.65

Αντίστοιχα αποτελέσματα έχουμε και για τους ISIS (Πίνακας 19), όπου η Εμπιστοσύνη ακολουθείται από Χαρά και Προσμονή, ενώ παρεμφερές σύμπλεγμα (με κίτρινη υπόδειξη στον πίνακα) δημιουργείται μεταξύ των συναισθημάτων του Φόβου, της Λύπης, του Θυμού, της Έκπληξης και της Αηδίας.

ΠΙΝΑΚΑΣ 19 ΣΥΝΔΙΑΚΥΜΑΝΣΗΣ ΚΑΝΟΝΙΚΟΠΟΙΗΜΕΝΩΝ ΣΥΝΑΙΣΘΗΜΑΤΩΝ ISIS

<i>Norm Emotion</i>	Joy	Trust	Fear	Surprise	Sadness	Disgust	Anger	Anticip
Joy	12.75	9.04	-11.41	0.30	-5.58	-1.58	-10.59	7.07
Trust	9.04	55.32	-27.44	-5.85	-11.18	-1.50	-25.32	6.92
Fear	-11.41	-27.44	28.64	0.33	5.53	-2.13	16.72	-10.24
Surprise	0.30	-5.85	0.33	4.91	0.25	-1.84	0.84	1.07
Sadness	-5.58	-11.18	5.53	0.25	13.28	1.61	3.22	-7.13
Disgust	-1.58	-1.50	-2.13	-1.84	1.61	8.23	0.95	-3.73
Anger	-10.59	-25.32	16.72	0.84	3.22	0.95	22.76	-8.58
Anticipation	7.07	6.92	-10.24	1.07	-7.13	-3.73	-8.58	14.61

4.7 Εντοπισμός Ανωμαλιών

Ύστερα από τα έως τώρα αποτελέσματα, προχωρήσαμε στο σχεδιασμό δύο πειραμάτων. Και στα δύο πειράματα χρησιμοποιήσαμε τα μοντέλα μη εποπτευόμενης μάθησης One Class Mahalanobis (Vareldzhan, et al., 2021), One Class SVM (Amer, et al., 2013; Najafabadi, et al., 2017; Khreich, et al., 2017) με αλγόριθμο RBF (Gorokhov, et al., 2017) και Isolation Forest (Ding & Fei, 2013; Xu, D. et al., 2017; Heigl, et al., 2021). Η επιλογή αυτή έγινε με βάση προηγούμενες έρευνες, όπου τα μοντέλα μίας τάξης αποδεδειγμένα υπερτερούν των άλλων μοντέλων σε σενάρια εντοπισμού ανωμαλιών (Zuang & Dai, 2006; Thudumu, et al., 2020; Seliya, et al., 2021). Κάθε μοντέλο εκπαιδεύτηκε (train) 10 φορές (10 fold validation), με την

εκάστοτε αναδίπλωση (fold) να χρησιμοποιεί ένα τυχαίο δείγμα 70% των άρθρων των Media (negative only train set). Οι σχετικές προβλέψεις (test) έγιναν πάνω στο υπολειπόμενο 30% των Media συν το 100% των άρθρων του ISIS. Η μόνη διαφορά μεταξύ των μοντέλων ήταν ότι στο πρώτο σενάριο χρησιμοποιήσαμε αθροιστικά συναισθήματα ως διανύσματα εισόδου, ενώ στη δεύτερη κανονικοποιημένα.

Στον εντοπισμό ανωμαλιών είναι σημαντικό να γίνει κατανοητό ότι η προσέγγιση δημιουργίας του train-test split, είναι εντελώς διαφορετική από ό,τι σε άλλες μεθοδολογίες (Kumar & Ravi, 2017; Schmitt & Spinoso, 2018; Thudumu, et al, 2020; Seliya, et al., 2021). Καταρχήν, ενώ σε άλλες περιπτώσεις απαιτείται ένα ισορροπημένο σύνολο δεδομένων (dataset), το οποίο έχει ίσα αντιπροσωπευτικά παραδείγματα από όλες τις ζητούμενες κατηγορίες, στην περίπτωση εντοπισμού ανωμαλιών το δείγμα μπορεί να αποτελείται κατά συντριπτική πλειοψηφία (ή εξ ολοκλήρου) μόνο από κανονικές παρατηρήσεις. Αυτό εξυπηρετεί σε δύο τομείς:

1. Με αυτό τον τρόπο εκπαιδύουμε το μοντέλο μόνο σε κανονικά δεδομένα, δημιουργώντας έτσι ένα “bias” το οποίο λειτουργεί υπέρ μας όταν δεχθεί ακραίες παρατηρήσεις. Με άλλα λόγια, στην περίπτωση των μοντέλων μίας κλάσης, εκπαιδύουμε μόνο με κανονικά παραδείγματα, αυτά σχηματίζουν τη κλάση και οτιδήποτε ακραίο χαρακτηρίζεται ως «μη μέλος».
2. Η εκπαίδευση μοντέλων με τέτοιο τρόπο εξομοιώνει πραγματικές συνθήκες, όπου δεν είναι απαραίτητο ότι έχουμε στην κατοχή μας ακραίες παρατηρήσεις. Ή μπορεί να έχουμε κάποιες ακραίες παρατηρήσεις ενός είδους και ξαφνικά, να δεχθούμε ως είσοδο κάτι ακραίο αλλά εντελώς διαφορετικό. Με αυτή τη μέθοδο βεβαιώνουμε στο μέγιστο δυνατό βαθμό πως ό,τι ανώμαλα παραδείγματα δεχθούμε, το μοντέλο θα είναι ικανό να τα εντοπίσει, ακόμα κι αν δεν τα έχει ξαναδεί.

Όπως αναφέρθηκε και πριν, κάθε επανάληψη train-test πραγματοποιήθηκε 10 φορές, με τυχαίο δείγμα εκπαίδευσης από τα άρθρα των Media. Παρακάτω, έχουμε τα συγκεντρωτικά αποτελέσματα απόδοσης κάθε μοντέλου για τα δύο σενάρια (Πίνακες 20 και 21). Συγκεκριμένα υπολογίσαμε το Μέσο Όρο Accuracy, Precision και Recall, καθώς και Μήτρα Σύγχυσης για κάθε επανάληψη. Για τις Μήτρες Σύγχυσης, τα ποσοστά που αναγράφονται στους πίνακες υπολογίστηκαν ως εξής:

- Αληθινά Θετικά (**TP**) = Σωστά κατηγοριοποιημένες ανώμαλες παρατηρήσεις / Σύνολο ανωμαλιών
- Ψευδή Αρνητικά (**FN**) = Λάθος κατηγοριοποιημένες ανώμαλες παρατηρήσεις / Σύνολο ανωμαλιών
- Αληθινά Αρνητικά (**TN**) = Σωστά κατηγοριοποιημένες κανονικές παρατηρήσεις / Σύνολο κανονικών
- Ψευδή Θετικά (**FP**) = Λάθος κατηγοριοποιημένες κανονικές παρατηρήσεις / Σύνολο κανονικών

Με βάση τους παραπάνω τύπους, αφού δεν έχουμε ανώμαλες παρατηρήσεις κατά τη φάση της εκπαίδευσης του μοντέλου, δεν είναι δυνατό να υπολογιστεί ποσοστό. Σε αυτές τις περιπτώσεις ο αριθμητής και ο παρονομαστής είναι μηδέν (0).

ΠΙΝΑΚΑΣ 20 ΜΕΣΟΣ ΟΡΟΣ ΔΕΙΚΤΩΝ ΕΠΙΔΟΣΗΣ ΜΟΝΤΕΛΟΥ (ΑΘΡΟΙΣΤΙΚΑ ΣΥΝΑΙΣΘΗΜΑΤΑ)

Μοντέλο	Φάση	Accuracy	Precision	Recall	True Positives	False Negatives	True Negatives	False Positives
Mahalanobis	Train	0.97559	1.00000	0.97559	0.	0.	0.97559	0.02441
	Test	0.83000	0.82864	0.97737	0.37246	0.62754	0.97737	0.02263
ocSVM	Train	0.58748	1.00000	0.58748	0.	0.	0.58748	0.41252
	Test	0.29048	0.97542	0.06353	0.99503	0.00497	0.06353	0.93647
Iso. Forest	Train	0.99980	1.00000	0.99980	0.	0.	0.99980	0.00020
	Test	0.76220	0.76093	0.99973	0.02479	0.97521	0.99973	0.00027

ΠΙΝΑΚΑΣ 21 ΜΕΣΟΣ ΟΡΟΣ ΔΕΙΚΤΩΝ ΕΠΙΔΟΣΗΣ ΜΟΝΤΕΛΟΥ (ΚΑΝΟΝΙΚΟΠΟΙΗΜΕΝΑ ΔΕΔΟΜΕΝΑ)

Μοντέλο	Φάση	Accuracy	Precision	Recall	True Positives	False Negatives	True Negatives	False Positives
Mahalanobis	Train	0.99046	1.00000	0.99046	0.	0.	0.99046	0.00954
	Test	0.74732	0.75425	0.98777	0.00083	0.99917	0.98777	0.01223
ocSVM	Train	0.58267	1.00000	0.58267	0.	0.	0.58267	0.41733
	Test	0.24552	0.60677	0.00613	0.98869	0.01131	0.00613	0.99387
Iso. Forest	Train	0.99960	1.00000	0.99960	0.	0.	0.99960	0.00040
	Test	0.75613	0.75633	0.99964	0.00000	1.00000	0.99964	0.00036

Αναλυτικοί πίνακες για κάθε φάση και αναδίπλωση όλων των μοντέλων βρίσκονται στα παραρτήματα Κ και Λ.

4.8 Αποτελέσματα

Παρήχθησαν πολλά ενδιαφέροντα αποτελέσματα έως τώρα. Προτού όμως επιχειρήσουμε να ερμηνεύσουμε το νόημά τους, οφείλουμε να συζητήσουμε εν συντομία τους δείκτες Accuracy, Precision και Recall. Ξεκινώντας από τον ορισμό του καθενός:

Accuracy

Είναι το ποσοστό που προκύπτει από τις σωστές προβλέψεις, προς το σύνολο όλων των προβλέψεων. Όσο πιο κοντά στο 1 (ή 100%) είναι ο δείκτης, τόσο καλύτερα. Ο αριθμός 1 υποδηλώνει με άλλα λόγια ότι όλες οι προβλέψεις ήταν σωστές, ενώ το 0 μας δείχνει ότι καμία πρόβλεψη δεν ήταν σωστή.

Ο τύπος υπολογισμού είναι:

$$Accuracy = \frac{\text{Αριθμός Σωστών Προβλέψεων}}{\text{Συνολικός Αριθμός Προβλέψεων}}$$

Ή για περιπτώσεις δυαδικής κατηγοριοποίησης:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision

Ο δείκτης αυτός επιχειρεί να μας δώσει εικόνα αναφορικά με την πιθανότητα ένα μοντέλο να κάνει σωστή ανώμαλη πρόβλεψη. Με άλλα λόγια, αν υποθέσουμε ότι είχαμε ένα μοντέλο με ποσοστό 60% Precision, αυτό θα σήμαινε ότι κάθε φορά που το μοντέλο κατηγοριοποιεί μία παρατήρηση ως ανώμαλη, η πιθανότητα αυτή η πρόβλεψη να είναι σωστή είναι 0,6. Το ποσοστό μας δίνεται από τον τύπο:

$$Precision = \frac{TP}{TP + F}$$

Recall

Αυτός ο δείκτης είναι παρόμοιος με τον Precision και πάντα εξετάζονται μαζί. Η συμπεριφορά των δύο αυτών δεικτών είναι αντίστροφη, με την έννοια ότι όταν προσπαθούμε να βελτιώσουμε τον έναν, ο άλλος χειροτερεύει. Η ερμηνεία του Recall μας δείχνει το ποσοστό των ανωμαλιών που έχουν προβλεφθεί σωστά, σε συνάρτηση όλων των ανώμαλων παρατηρήσεων στο σύνολο δεδομένων μας. Δίδεται από τον τύπο:

$$Recall = \frac{TP}{TP+FN}$$

Παρόλο που και οι τρεις δείκτες είναι χρήσιμοι στη μηχανική μάθηση γενικά, αδυνατούν να περιγράψουν σωστά την αποτελεσματικότητα των μοντέλων εντοπισμού ανωμαλιών. Αυτό οφείλεται στο γεγονός ότι, εκ προοιμίου στον εντοπισμό ανωμαλιών το ποσοστό ανώμαλων παρατηρήσεων εντός του συνόλου δεδομένων μας είναι εν γένει μικρό (Thudumu, et al, 2020; Seliya, et al., 2021). Όπως είναι φυσικό, αν λάβουμε υπόψη τους παραπάνω τύπους, μπορεί να έχουμε για παράδειγμα υψηλό Accuracy της τάξης του 98%. Όμως επειδή το σύνολο δεδομένων μας αποτελείται σε ποσοστό 1% από ανώμαλες παρατηρήσεις, είναι πιθανό αυτό το 98% να προκύπτει από τη σωστή πρόβλεψη της πλειοψηφίας των κανονικών παρατηρήσεων (που ίσως αποτελούν το 99% του dataset) και να υποκρύπτει τη λανθασμένη πρόβλεψη ανωμαλιών (σημαντικά λιγότερες αναλογικά).

Ανατρέχοντας στους πίνακες δεικτών επίδοσης 20 και 21, παρατηρούμε το φαινόμενο που μόλις περιγράψαμε. Για το λόγο αυτό, συμπεριλάβαμε τα ποσοστά από τις μήτρες σύγχυσης. Όπως αναφέραμε στο κεφάλαιο 4.3 περί Προετοιμασίας των δεδομένων, υποθέσαμε ότι τα άρθρα των Media αποτελούν κανονικές παρατηρήσεις και τους δόθηκε πραγματικό $y = +1$, ενώ τα άρθρα των ISIS θεωρήθηκαν ανώμαλα και χαρακτηρίστηκαν με πραγματικό $y = -1$. Οι προβλέψεις \hat{y} κάθε μοντέλου συγκρίθηκαν με τα πραγματικά y και έτσι παρήχθησαν οι τιμές True/False Positive/Negative. Για την αξιολόγηση των μοντέλων και ερμηνεία των αποτελεσμάτων, θα εστιάσουμε στις μήτρες σύγχυσης.

4.8.1 Αθροιστικά Δεδομένα

Το μοντέλο Mahalanobis κατανόησε τα κανονικά (αρνητικά) δεδομένα με επιτυχία κατά τη φάση εκπαίδευσης. Ύστερα, κατάφερε να αναγνωρίσει το 37% των ανωμαλιών, χωρίς να τις έχει δει ποτέ ξανά. Σε αυτό το σενάριο δοκιμής, ήταν το μοντέλο με την καλύτερη επίδοση.

Το μοντέλο ocSVM κατάφερε να προβλέψει σωστά το 59% των κανονικών παρατηρήσεων κατά τη φάση εκπαίδευσης. Κατά τη φάση του τεστ ωστόσο, φαίνεται να παρουσίασε ισχυρό bias, χαρακτηρίζοντας σχεδόν όλες της παρατηρήσεις ως ανώμαλες.

Το μοντέλο Isolation Forest επίσης κατανόησε τα κανονικά δεδομένα κατά την εκπαίδευση. Όμως, κατά το τεστ, φάνηκε να έχει ακριβώς αντίθετη συμπεριφορά συγκριτικά με το ocSVM, κατηγοριοποιώντας σχεδόν όλες τις παρατηρήσεις ως κανονικές.

4.8.2 Κανονικοποιημένα Δεδομένα

Στο σενάριο αυτό, τα μοντέλα Mahalanobis και Isolation Forest επέδειξαν παρόμοια συμπεριφορά. Ενώ έδειξαν να καταλαβαίνουν τις κανονικές παρατηρήσεις κατά την εκπαίδευση, δεν κατάφεραν να εντοπίσουν αποτελεσματικά τις ανωμαλίες κατά την εκπαίδευση.

Το μοντέλο ocSVM, εσφαλμένα χαρακτήρισε ως ανώμαλες 42% των παρατηρήσεων κατά την εκπαίδευση. Μετά, κατά τη φάση τεστ παρουσίασε ίδια συμπεριφορά όπως και με τα αθροιστικά δεδομένα στο πρώτο σενάριο, κατηγοριοποιώντας σχεδόν όλες τις παρατηρήσεις ως ανώμαλες.

4.8.3 Ερμηνεία

Όπως είδαμε και κατά το στάδιο της διερευνητικής ανάλυσης, τόσο τα Media όσο και οι ISIS, δείχνουν να χρησιμοποιούν σχεδόν πανομοιότυπες τακτικές συνδυασμού συναισθημάτων. Μία διαφορά είναι ότι οι ISIS φαίνεται να χρησιμοποιούν τα κυρίαρχα αισθήματα που είδαμε (Εμπιστοσύνη, Φόβος, κλπ) σε πιο έντονο βαθμό από ό,τι τα Media. Αυτό είναι ορατό ειδικά όταν κοιτάμε τα αθροιστικά συναισθήματα κάθε άρθρου. Σε αυτό το σενάριο, το μοντέλο Mahalanobis υπερσχύει των υπολοίπων (ocSVM, Isolation Forest), καθώς είναι ικανό να εντοπίσει το έντονο των συναισθημάτων στα κείμενα των ISIS. Με άλλα λόγια, μπορεί να έχουμε ίδια συναισθήματα μεταξύ Media και ISIS, αλλά οι ISIS προσπαθούν να εγείρουν τα αισθήματα αυτά πιο έντονα. Το μοντέλο Mahalanobis μπορεί και συλλαμβάνει αυτή τη διαφορά.

Ειδικά στην περίπτωση του δεύτερου σεναρίου, όπου έχουμε κανονικοποιημένα συναισθήματα και άρα, αναφερόμαστε καθαρά σε αναλογίες, παρατηρείται ένα ενδιαφέρον φαινόμενο. Όπως προκύπτει και από τη διερευνητική ανάλυση, τα Media και οι ISIS κινούνται σε κοινές συναισθηματικές συνιστώσες. Αυτό έχει ως αποτέλεσμα όμοιες συναισθηματικές κανονικότητες. Λαμβάνοντας υπόψη την επίδοση των μοντέλων μας στο δεύτερο σενάριο, θα μπορούσε να επικαλεστεί κανείς βιαστικά ότι τα μοντέλα αποτυγχάνουν σε αυτή την περίπτωση να διαχωρίσουν τα κανονικά άρθρα από τα ανώμαλα. Ωστόσο, στην πραγματικότητα φαίνεται η συμπεριφορά αυτή των μοντέλων, να επαληθεύει τις αρχικές παρατηρήσεις και υποψίες που είχαμε κατά τη διερευνητική ανάλυση. Το ότι τα μοντέλα δεν μπορούν ξεχωρίσουν μεταξύ Media και ISIS, σημαίνει ουσιαστικά ότι όντως και

οι δύο πλευρές κάνουν χρήση πανομοιότυπων συναισθηματικών κανονικοτήτων, για την επικοινωνία των μηνυμάτων τους.

5 Συμπεράσματα

5.1 Ανασκόπηση

Σκοπός της έρευνας μας ήταν να αναλύσουμε τα θέματα που κυριαρχούσαν στη διαδικτυακή προπαγάνδα του ISIS, όπως αυτή εκφραζόταν από τα ηλεκτρονικά περιοδικά-εκδόσεις του Al-Hayat Media Center. Όπως έχει ήδη αναφερθεί στο τρίτο κεφάλαιο, το συγκεκριμένο κέντρο εξέδιδε τα περιοδικά:

- ✓ Dabiq,
- ✓ Rumiayah,
- ✓ Istok,
- ✓ Dar al Islam και
- ✓ Konstantiniyye

στοχεύοντας με το καθένα από αυτά σε ένα διαφορετικό, γλωσσικά, ακροατήριο.

Παράλληλα, είχαμε σκοπό να προσδιορίσουμε μέσα από το αφήγημα της προπαγάνδας του ISIS τα βασικά συναισθήματα που ενεργοποιούνται. Έτσι, μέσα από τεχνικές μηχανικής μάθησης και φυσικής γλώσσας (NLP) αναλύθηκαν τα κείμενα με σκοπό την ανάδειξη εντάσεων συναισθημάτων και πολικότητας. Στα αποτελέσματα της ανάλυσης διακρίνει κανείς ότι η ανάλυση των κειμένων μέσω αποτύπωσης των εντάσεων των συναισθημάτων υπερτερεί σαφώς αυτής της πολικότητας.

Θεωρούμε ότι το μεθοδολογικό πλαίσιο που χρησιμοποιήθηκε όχι μόνο επιβεβαιώνει τους ισχυρισμούς διαφόρων ερευνητών-ειδικών σε θέματα τρομοκρατίας σχετικά με τα κυρίαρχα συναισθήματα που προσπαθούσε να αφυπνίσει η προπαγάνδα του ISIS αλλά, προσφέρει και ένα δομημένο πλαίσιο το οποίο αυξάνει τη διακριτική ικανότητα παρατήρησης του ερευνητή. Έτσι, διαπιστώσαμε ότι τα κυρίαρχα συναισθήματα στα κείμενα των εξετασθέντων ηλεκτρονικών περιοδικών ήταν τα παρακάτω (σε σειρά μειούμενης έντασης):

1. Εμπιστοσύνη (Trust),
2. Φόβος (Fear) και
3. Χαρά (Joy)

Επιπρόσθετα, η LDA προσδιόρισε με ακρίβεια τις λέξεις-θέματα της συλλογής και οι οποίες αντιπροσωπεύουν πλήρως τις έξι συνιστώσες της φέρμας του ISIS, όπως αυτές έχουν αποτυπωθεί από τον Winter (Winter, 2015). Μέσω της εναλλακτικής αυτής προσέγγισης παράχθηκαν ακριβέστερα συναισθηματικά μορφότυπα σχετικά με τα κείμενα, λόγω του ότι λαμβάνονταν υπόψη όλες οι λεξικογραφικές μονάδες.

Ταυτόχρονα, στο δεύτερο σκέλος της έρευνας εμπλουτίσαμε το σύνολο δεδομένων μας με άρθρα δυτικών ειδησεογραφικών πρακτορείων. Ήδη από το στάδιο της διερευνητικής ανάλυσης ήταν ορατό ότι τόσο τα Media όσο και οι ISIS, κάνουν χρήση των ίδιων συναισθημάτων με τον ίδιο τρόπο. Σημαντικά ευρήματα αυτού του πειράματος ήταν ότι, το μοντέλο μηχανικής μάθησης που βασιζόταν στην απόσταση Mahalanobis, κατάφερε σε σημαντικό βαθμό να εντοπίσει τα εντονότερα συναισθηματικά μορφότυπα των ISIS, όταν τα διανύσματα αφορούσαν τα αθροιστικά συναισθήματα του άρθρου. Επιπλέον, όταν χρησιμοποιήσαμε ως εισόδους τα κανονικοποιημένα συναισθήματα, όλα τα μοντέλα απέτυχαν να διαφοροποιήσουν τις ανώμαλες παρατηρήσεις από τις κανονικές. Αυτός ο διαχωρισμός έγινε αρχικά με βάση την υπόθεση ότι τα Media έχουν διαφορετικά συναισθηματικά μορφότυπα από τα άρθρα των ISIS. Αυτό μας υποδεικνύει ότι πιθανότατα υπάρχει ισχυρή ομοιότητα των συναισθηματικών προτύπων που οι δύο πλευρές χρησιμοποιούν στις επικοινωνιακές τους στρατηγικές.

Σε κάθε περίπτωση, τα εξαχθέντα συμπεράσματα είναι άκρως ενδιαφέροντα και εκπλήσσουν. Με βάση την γνώση που διαθέτουμε, δεν υφίσταται κάποια άλλη έρευνα σχετική με την δική μας που να παράγει αυτά τα αποτελέσματα στην έκταση αυτή. Θεωρούμε ότι η εργασία αυτή αποτελεί συνεισφορά στην έρευνα του εν λόγω πεδίου με, εν δυνάμει, εφαρμογές όχι μόνο στον Πληροφοριακό Πόλεμο αλλά και σε πεδία όπως είναι το Μάρκετινγκ, η Πολιτική, η Διαχείριση Συμπεριφοράς σε σχέση με τη διαχείριση κρίσεων κλπ.

Επιπλέον, βάσει των ανωτέρω, απαντάμε στα εναπομείναντα ερευνητικά ερωτήματα:

RQ6. Παρατηρούνται σημαντικές ομοιότητες μεταξύ των συναισθηματικών προτύπων που εντοπίζονται, ανάμεσα σε άρθρα γνωστών ειδησεογραφικών πρακτορείων και αυτά του ISIS. Τα μοντέλα μηχανικής μάθησης που δοκιμάσαμε δεν καταφέρνουν να διαχωρίσουν μεταξύ των δύο ομάδων, ειδικά όταν τα συναισθήματα είναι κανονικοποιημένα.

RQ7. Μπορούμε δυνητικά να χρησιμοποιήσουμε μοντέλα μηχανικής μάθησης για να εντοπίσουμε ανωμαλίες σε μια συλλογή κειμένων (corpus), βασιζόμενοι μόνο στο συναίσθημα. Παρά τη δυσκολία που αντιμετώπισαν τα μοντέλα μηχανικής μάθησης κατά την έρευνά μας, φαίνεται πως υπήρχαν παρατηρήσεις που αναγνωρίζονταν ως ακραίες περιπτώσεις. Επίσης, κατά το πείραμα με τα αθροιστικά συναισθήματα, αν και η μέθοδος επηρεάζεται άμεσα από το σκέλος του κειμένου, το μοντέλο Mahalanobis κατάφερε να διαχωρίσει αρκετές ακραίες παρατηρήσεις σε σχέση με τα άλλα δύο μοντέλα.

RQ8. Το μοντέλο Mahalanobis έχει μεγαλύτερες δυνατότητες σε σύγκριση με τα ocSVM και Isolation Forest. Ειδικά όταν τα συναισθήματα δεν έχουν κανονικοποιηθεί. Ένα πιο διακριτά διαφοροποιημένο corpus (σε συναισθηματικό επίπεδο) ίσως βοηθούσε το μοντέλο Mahalanobis περαιτέρω, ακόμα και στο σενάριο όπου τα συναισθήματα έχουν κανονικοποιηθεί.

5.2 Μελλοντική έρευνα

Αναφορικά με την έρευνα του τρίτου κεφαλαίου, ένα πρόβλημα που αντιμετωπίσαμε αφορούσε το μεγάλο αριθμό των περιεχόμενων αραβικών λέξεων μέσα στα αναλυόμενα κείμενα της συλλογής των περιοδικών και είναι σημαντικές μια και υποδεικνύονται ως λέξεις-θέματα από την LDA (πχ *allah, islam, jihad, khilafah, mujahidin*). Οι λέξεις αυτές, όπως καταδεικνύεται από σχετικές έρευνες, εμφανίζουν πολλαπλές και διαφορετικές ερμηνείες ανάλογα την ομάδα των ανθρώπων που τη χρησιμοποιεί (Ozegen, et al., 2018). Το λεξικό EmoLex δεν επιτρέπει την αντιστοίχιση των λέξεων αυτών σε συναισθήματα. Στα μελλοντικά μας ερευνητικά σχέδια είναι να πραγματοποιηθεί η παραπάνω ανάλυση μέσω μιας πολυγλωσσικής έκδοσης του EmoLex. Ένα ακόμη ζήτημα ήταν ότι αν και το λεξικό EmoLex κάλυπτε σε πολύ καλό βαθμό τις λέξεις που υπήρχαν μέσα στο κάθε κείμενο, σίγουρα υπάρχει περιθώριο βελτίωσης. Επίσης, είναι σημαντικό να αναλυθεί και κατανοηθεί ο βαθμός στον οποίο περίπλοκες έννοιες δεν συλλαμβάνονται σωστά. Για παράδειγμα, οι φράσεις “good”, “very good”, “not very good”, “not as good” είναι περιπτώσεις με διαφορετικά νοήματα. Θα άξιζε λοιπόν να διερευνηθεί αν και σε ποιο βαθμό η χρήση τεχνικών, όπως n-grams ή word vectors, βελτιώνουν τα αποτελέσματα. Τέλος, είδαμε ότι το μοντέλο Mahalanobis υπόσχεται καλές επιδόσεις όταν υπάρχει διαφορά στο έντονο των συναισθημάτων. Η μέθοδος αυτή καταφέρνει με επιτυχία, και μόνο με αθροιστικά συναισθήματα ως είσοδο, να εντοπίσει συναισθηματικά φορτισμένα κείμενα και να τα διαχωρίσει από άλλα. Ωστόσο πρέπει να σημειωθεί ότι αυτή η μέθοδος επηρεάζεται άμεσα

από το μέγεθος του κειμένου και το ποσοστό επιτυχών ευρέσεων λέξεων στο λεξικό. Για παράδειγμα, είναι αναμενόμενο ένα κείμενο Α δεκαπλάσιο σε μέγεθος από ένα κείμενο Β, να έχει αναλόγως, δέκα φορές μεγαλύτερα συναισθηματικά αθροίσματα από το Β. Επίσης, αν έχουμε δύο κείμενα με μη σημαντική στατιστική διαφορά στο πλήθος των λέξεων που τα απαρτίζουν, συνεχίζει να υπάρχει πρόβλημα εάν στο πρώτο επιτευχθεί εξαγωγή συναισθημάτων σε σημαντικά μικρότερο ποσοστό λέξεων από το δεύτερο. Θα μπορούσαμε να πούμε αναφορικά με αυτό το ενδεχόμενο, ότι η αποτελεσματικότητα του μοντέλου εξαρτάται και από το έμμεσο bias που παρουσιάζει το λεξικό, μέσα από τις λέξεις που περιέχει ή/και δεν περιέχει. Παρόλα αυτά, είναι αδιαπραγμάτευτο ότι υπάρχει πλούτος πληροφοριών στην εξαγωγή συναισθημάτων από λέξεις. Ενδιαφέρον κομμάτι μελλοντικής έρευνας θα ήταν και ο συνδυασμός συναισθημάτων με άλλες τεχνικές κατηγοριοποίησης, καθώς και η βελτίωση των παραμέτρων και υπερπαραμέτρων του μοντέλου Mahalanobis.

Κλείνοντας, θεωρούμε ότι η παρούσα θέση επιτυγχάνει να καινοτομήσει και να συμβάλλει στην υπάρχουσα βιβλιογραφία με πολύτιμα αποτελέσματα, αλλά και να εντοπίσει παραμέτρους οι οποίες χρίζουν βελτίωσης και προσοχής για τη σωστή διεξαγωγή ερευνών.

6 Εφαρμογές

Στην παρούσα διατριβή παρουσιάσαμε μία έρευνα και σχετικά αποτελέσματα, εστιάζοντας στην ανάλυση άρθρων ειδησεογραφικών πρακτορείων (Al-Hayat Media Center του ISIS και δημοφιλή ΜΜΕ). Παρά τον περιορισμό σε αυτό το δείγμα κειμένων, οι εφαρμογές του μεθοδολογικού πλαισίου που περιγράφηκε στα προηγούμενα κεφάλαια είναι ποικίλες. Μπορούμε να διακρίνουμε μεγάλο αριθμό σχετικών περιπτώσεων με τις πιο σημαντικές, οι οποίες σχετίζονται με θέματα διοίκησης, να δίνονται στη συνέχεια:

1. Εξαγωγή συναισθηματικής πληροφορίας από δημοσιεύσεις χρηστών του διαδικτύου, αναφορικά με ένα προϊόν, υπηρεσία, ανακοίνωση, ενέργεια μιας επιχείρησης. Περαιτέρω λήψη αποφάσεων με βάση το συναίσθημα του κοινού.
2. Ex-ante και post αξιολόγηση ενός στρατηγικού σχεδίου marketing, (re)branding και λοιπά, μέσω ποσοτικοποίησης ποιοτικών δεδομένων (όπως είναι τα συναισθήματα και το κείμενο) και δημιουργίας δεικτών.

3. Ζωντανή σφυγμομέτρηση τμημάτων υποστήριξης πελατών αναφορικά με το κυρίαρχο συναίσθημα των πελατών που επικοινωνούν. Προσαρμογή της επικοινωνιακής στρατηγικής δυναμικά, αντί στατικά.
4. Σφυγμομέτρηση συναισθηματικού κλίματος των εργαζομένων μιας εταιρείας, μέτρηση στρες, αξιολόγηση κουλτούρας, well-being και λοιπά. Δημιουργία εσωτερικού επικοινωνιακού σχεδίου, λήψη μέτρων, σχεδιασμός κινήτρων, κλπ.
5. Αντιμετώπιση κρίσεων σε διάφορα γεωγραφικά επίπεδα. Για παράδειγμα, κατανόηση της συναισθηματικής κατάστασης των πολιτών αναφορικά με θέματα όπως η πανδημία, η κλιματική αλλαγή, ο πόλεμος, η οικονομική κρίση και άλλες διαταραχές. Ύστερα, δημιουργία πλάνου ανταπόκρισης και διαχείρισης.
6. Αξιολόγηση πολιτικής καμπάνιας ή/και πολιτικών ανακοινώσεων/ενεργειών.
7. Εθνική άμυνα, συγκεκριμένα στο μέρος του πληροφοριακού πολέμου και της εξωτερικής προπαγάνδας. Αντίμετρα αντιμετώπισης και εντοπισμός/περιορισμός της προπαγάνδας.

Οι τομείς εφαρμογών δεν είναι μόνο οι παραπάνω. Εφαρμογές υπάρχουν επίσης στην εγκληματολογία, στη δημόσια υγεία, την κλινική έρευνα και ψυχολογία, τη λογοτεχνία, τη γλωσσολογία, την ανάλυση χρηματαγορών και λοιπά. Θεωρούμε ότι η παρούσα έρευνα και τα ευρήματά της επιτυγχάνουν σημαντική και καταλυτική συμβολή, σε όλες τις προαναφερθείσες περιοχές, τόσο σε ακαδημαϊκό, όσο και εμπορικό/κυβερνητικό επίπεδο.

7 Βιβλιογραφία

1. Abdul-Fattah, H. M., Masum-al-Masba, M. & Azharul-Hasan, K. M., 2018. Sentiment Clustering by Mahalanobis Distance. s.l., 4th International Conference on Electrical Engineering and Information and Communication Technology.
2. Acheampong, F., Wenyu, C. & Nunoo-Mensah, H., 2020. Text-based emotion detection: Advances, challenges, and opportunities. *Engineering Reports*, Volume 2, pp. 1-24.
3. Adena, M. και συν., 2015. Radio and the Rise of The Nazis in Prewar Germany. *The Quarterly Journal of Economics*, 130(4), pp. 1885-1939.
4. Ahmad, S., Asghar, M. Z., Alotaibi, F. M. & Awan, I., 2019. Detection and classification of social media based extremist affiliations using sentiment analysis techniques. *Human-centric Computing and Information Sciences*, pp. 9-24.
5. Akioyamen, P., Nicklas, L. C. & Sanchez-Arias, R., 2020. A Framework for Intelligent Navigation Using Latent Dirichlet Allocation on Reddit Posts about opiates. s.l., 4th International Conference on Computer and Data Analysis.
6. Al-Amin, M., Islam, S. & Uzzal, A. A., 2017. Comprehensive study on Sentiment of Bengali Text. s.l., International Conference on Electrical, Computer and Communication Engineering.
7. Aldgate, A. & Richards, J., 2007. *Britain Can Take It: The British Cinema in the Second World War*. 3rd Edition. London: I. B. Tauris.
8. Alfifi, M. & Caverlee, J., 2017. Badly Evolved? Exploring Long-Surviving Suspicious Users on Twitter. s.l., International Conference on Social Informatics, Springer.
9. Alfifi, M., Kaghazgaran, P. & Morstatter, F., 2018. Measuring the Impact of ISIS Social Media Strategy. Available at:

https://snap.stanford.edu/mis2/files/MIS2_paper_23.pdf [Πρόσβαση 20 10 2021].
10. Alm, E. C. O., 2008. *Affect in Text and Speech* (PhD thesis). Urbana Champaign: University of Illinois.
11. Al-Tamimi, A. J., 2015. *The Isis papers: A masterplan for consolidating power*. Full Text and Translation at:

<https://www.theguardian.com/world/2015/dec/07/islamic-state-document-masterplan-for-power> [Πρόσβαση 20 6 2021]

12. Altheide, D. L. & Grimes, J. N., 2005. War Programming: The Propaganda Project and the Iraq War. *The Sociological Quarterly*, 46(4), pp. 617-643.
13. Alyousef, Y. & Hasmah Zanuddin, H., 2018. Mining online communities to inform strategic messaging: practical methods to identify community-level insights. *International Journal of Engineering & Technology*, 7(2.29), pp. 633-638.
14. Amati, G., Angelini, S., Bianchi, M., Constantini, L., Marcone, G., 2016. A cumulative approach to quantification for sentiment analysis. [Ηλεκτρονικό] Available at: <https://arxiv.org/abs/1610.01366> [Πρόσβαση 25 7 2021].
15. Amer, M., Goldstein, M. & Abdennadher, S., 2013. Enhancing one-class Support Vector Machines for unsupervised anomaly detection. s.l., Amer, Mennatallah & Goldstein, Markus & Abdennadher, Slim. (2013). Enhancing one-class Support Vector Machines for unsupervised anomaly detection. *Proceedings ACM SIGKDD Workshop on Outlier Detection and Description*.
16. Ansaria, M., Aziz, M., Mehraa, S. H. & Singha, K., 2020. Analysis of Political Sentiment Orientations on Twitter. *Procedia Computer Science*, Volume 167, pp. 1821-1228.
17. Archetti, C., 2015. Terrorism, Communication and New Media: Explaining Radicalization in the Digital Age. *Perspectives on Terrorism*, 9(1), pp. 49-59.
18. Azaz, P., Navimipour, N. J., Rahmani, A. M. & Sharifi, A., 2019. The role of structured and unstructured data management mechanisms in the Internet of Things. *Cluster Computing*, Volume 23, pp. 1185-1198.
19. Kumar B. S., Ravi V. (2017) Text Document Classification with PCA and One-Class SVM. In: Satapathy S., Bhateja V., Udgata S., 5th International Conference on Frontiers in Intelligent Computing.
20. Babbie, E., 2014. *The Basics of Social Research. International Edition. 6th Ed.* s.l.: Cengage.
21. Badaway, A. & Ferrara, E., 2018. The rise of Jihadist propaganda on social networks. *Journal of Computational Social Science*, Volume 1, pp. 453-470.
22. Balahur, A. & Turchi, M., 2012. Multilingual Sentiment Analysis using Machine Translation? *Association for Computational Linguistics*, pp. 52 - 60.
23. Balasubramanian, V., Gupta, S. & Veerappagoundar, P., 2016. Mahalanobis distance-the ultimate measure for sentiment analysis. *The International Arab Journal of Information Technology*, 13(2), pp. 252-257.

24. Barbalet, J., 2004. *Emotion, Social Theory and Social Structure*. Cambridge: Cambridge University Press.
25. Barron-Estrada, M. L., Zatarain-Cabada, R. & Oramas-Bustillos, R., 2019. Emotion Recognition for Education using Sentiment Analysis. *Research in Computer Science*, 148(5), pp. 71-80.
26. Barton, D., 2016. Data mining in the deep web. *The innovation enterprise*, 14 July.
27. Basant, A., Namita, M., Pooja, B. & Garg, S., 2015. Sentiment analysis using common-sense and context information. *Computational Intelligence and Neuroscience*, pp. 1-9.
28. Basile, A., Franco-Salvador, M., Pawar, N., Stajner, S., 2019. SymantoResearch at SemEval-2019 task 3: combined neural models for emotion classification in human-chatbot conversations. Minneapolis, 13th international workshop on semantic evaluation.
29. Baziotis, C., Athanasiou, N., Chronopoulou, A., Kolovou, A., Paraskevopoulos, G., Ellinas, N., Narayanan, S., Potamianos, A., 2018. NTUA-SLP at SemEval-2018 task 1: predicting affective content in tweets with deep attentive rnns and transfer learning. s.l., 12th International workshop on semantic evaluation.
30. Baziotis, C., Pelekis, N. & Doulkeridis, C., 2017. Datastories at semeval-2017 task 4: deep LSTM with attention for message-level and topic-based sentiment analysis. s.l., 11th international workshop on semantic evaluation (SemEval-2017).
31. Beigi, G., Hu, X., Maciejewski, R. & Liu, H., 2016. An Overview of Sentiment Analysis in Social Media and its Applications in Disaster Relief. *Sentiment Analysis and Ontology Engineering*. s.l.:Springer, pp. 313-340.
32. Bengio, Y. & Grandvalet, Y., 2004. No Unbiased Estimator of the Variance of K-Fold Cross-Validation. *Journal of Machine Learning Research*, Volume 5, pp. 1089–1105.
33. Benigni, M. & Carley, K. M., 2016. From Tweets to Intelligence: Understanding the Islamic Jihad Supporting Community on Twitter. Washington, DC, Cultural, and Behavioral Modeling: 9th International Conference, SBP-BRiMS 2016.
34. Benigni, M., Joseph, K., Carley & M., K., 2018. Mining online communities to inform strategic messaging: practical methods to identify community-level insights. *Computational and Mathematical Organization Theory*, Springer, 24(2), pp. 224-242.
35. Berger, C. 1959. *An Introduction to Wartime Leaflets*, Washington, DC: Washington, DC: Department of the Army

36. Berger, J. M. & Morgan, J., 2015. The ISIS Twitter Census: Defining and describing the population of ISIS supporters on Twitter, s.l.: The Brookings Project on US Relations with the Islamic World.
37. Bernays, E. L., 1928. Propaganda 1st Edition. New York: Ig Publishing.
38. Bhargav, S., 2014. Efficient Features for Movie. Recommendation Systems, Stockholm, Sweden: Master's Degree Project XR-EE-KT.
39. Birjali, M., Beni-Hssane, A. & Erritali, M., 2017. Machine Learning and Semantic Sentiment Analysis based Algorithms for Suicide Sentiment Prediction in Social Networks. s.l., Procedia Computer Science, pp. 65-72.
40. Blair, C., 1987. The Forgotten War: America in Korea, 1950–1953. New York: Times Books.
41. Blei, D. M., Ng, A. Y. & Jordan, M. I., 2003. Latent Dirichlet allocation. Journal of Machine Learning Research, 3(4 - 5), pp. 993 - 1022.
42. Bloom, M. & Daymon, C., 2018. Assessing the Future Threat: ISIS's Virtual Caliphate. Orbis, May, 62(3), pp. 372 - 388.
43. Bodine-Baron, E., Helmus, T., Magnuson, M. & Winkelman, Z., 2016. Examining ISIS Support and Opposition Networks on Twitter, Santa Monica: RAND Corporation.
44. Borau, S. & Wamba, S., 2019. Social Media, Evolutionary Psychology, and ISIS: A Literature Review and Future Research Directions. s.l., New Knowledge in Information Systems and Technologies. WorldCIST'19 2019.
45. Boyd-Graber, J. & Resnik, P., 2010. Holistic Sentiment Analysis Across Languages: Multilingual Supervised Latent Dirichlet Allocation. s.l., Boyd-Graber, J.; Resnik, P. Holistic Sentiment Analysis Across Languages: Multilingual Supervised Conference on Empirical Methods in Natural Language Processing.
46. Bradley, M. & Lang, P., 1999. Affective norms for english words (anew): In struction manual and affective ratings, Florida: The Center for Research in Psychophysiology, University of Florida.
47. Brereton, R., 2015. The Mahalanobis distance and its relationship to PCA. Journal of Chemometrics, Volume 29, pp. 143-145.
48. Cafarella, J. & Zhou, J., 2017. ISIS's Expanding Campaign in Europe. [Ηλεκτρονικό] Available at:

<http://www.understandingwar.org/sites/default/files/ISIS%20in%20Europe%20Update%20September-2017.pdf> [Πρόσβαση 20 6 2020].

49. Cao, Z., Zheng, M., Vorobyeva, Y., Song, C., Johnson, N. F., 2017. Dynamical Patterns in Individual Trajectories Toward Extremism. ArXiv, Physics and Society.
50. Capuano, N., Greco, L., Ritrovato, P. & Vento, M., 2020. Sentiment analysis for customer relationship management: an incremental learning approach. Applied Intelligence, pp. 1-14.
51. Carafella, J., Wallace, B. & Zhou, J., 2019. Isis's Second Comeback: Assessing the next ISIS Insurgency 1st Edition. Washington, DC: Institute for the Study of the War.
52. Carey, A., 1997. Taking the risk out of democracy: Corporate propaganda versus freedom and liberty 1st Edition. Urbana: University of Illinois Press.
53. Carley, C. M., Malik, M., Landwehr, P. M., Pfeffer, J., Kowalchuck, M., 2016. Crowd sourcing disaster management: The complex nature of Twitter usage in Padang Indonesia. Safety Science, Volume 90, pp. 48-61.
54. Casey, R., 1944. EM2: What is propaganda. 1st εκμ. Washington, DC: American Historical Association.
55. Casey, S., 2008. Selling the Korean War. Propaganda, Politics, and Public Opinion 1950-1953. Oxford: Oxford University Press.
56. Chandler, R., 1981. War Of Ideas: The U.s. Propaganda Campaign In Vietnam. s.l.:Rutledge.
57. Chatterjee, A., Narahari, K., Joshi, M. & Agrawal, P., 2019. SemEval-2019 Task 3: EmoContext contextual emotion detection in text. s.l., 13th International Workshop on Semantic Evaluation.
58. Chatzakou, D., Vakali, A. & Kafetsios, K., 2017. Detecting variation of emotions in online activities. Expert Systems with Applications, Volume 89, pp. 318 - 332.
59. ChengXiang, Z. & Masung, S., 2016. Text Data Management and Alaysis: A practical introduction to Information retrieval and text mining. 1st εκμ. s.l.:ACM, Morgan and Claypool.
60. Cherry, C., Mohammad, S. & de Bruijn, B., 2012. Binary Classifiers and Latent Sequence Models for Emotion Detection in Suicide Notes. Biomedical Informatics Insights, 5(1), pp. 147-154.

61. Choi, S. & Park, H. W., 2014. An exploratory approach to a Twitter-based community centered on political goal in south Korea: Who organized it, what they shared, and how they acted. *New Media and Society*, 16(1), pp. 129-148.
62. Cohen-Almagor, R. 2012. In Internet's Way: Radica, Terrorist Islamists on the Free Highway. *International Journal of Cyber Warfare and Terrorism*, 2(3), pp. 39-58.
63. Cong, Q., Feng, Z., Li, F., Xiang, Y., Rao, G., Tao, C., 2018. X-A-BiLSTM: a Deep Learning Approach for Depression Detection in Imbalanced Data. s.l., DIEEE International Conference on Bioinformatics and Biomedicine.
64. Courty, A., Halim, R., Kasun, U., 2019. Blood and ink: the relationship between Islamic State propaganda and Western media. *The Journal of International Communication*, 25(1), pp. 69-94.
65. Cullen, P. & Reichborn-Kjennerud, E., 2017. MCDC Countering Hybrid Warfare Project: Understanding Hybrid warfare, s.l.: Multinational Capability Development Campaign.
66. Cull, N. J., Gatov, V., Pomerantsev, P., Applebaum, A., Shawcross, A., 2017. *Soviet Subversion, Disinformation and Propaganda: How the West Fought Against it. An Analytic History, with Lessons for the Present*, London: London School of Economics.
67. Cunliffe, E. & Curini, L., 2018. ISIS and heritage destruction: A sentiment analysis. *Antiquity*, 92(364), pp. 1094-1111.
68. Cunningham, D., Everton, S. F. & Schroeder, R., 2014. *Social Media and the ISIS Narrative*. [Ηλεκτρονικό] Available at:

https://calhoun.nps.edu/bitstream/handle/10945/53059/Social_Media_and_the_ISIS_Narrativ_%28Everton_et_al%29.pdf?sequence=1&isAllowed=y [Πρόσβαση 18 10 2021].
69. Dadvar, M., Trieschnigg, D. & de Jong, F., 2014. Experts and Machines against Bullies: A Hybrid Approach to Detect Cyberbullies. S. M. & v. B. P., *Advances in Artificial Intelligence. Canadian AI 2014. Lecture Notes in Computer Science*, vol 8. s.l.: s.n., pp. 275-281.
70. Dai, T., 2017. *News Articles, Version 1*, Harvard Dataverse
71. Dai, S. & Man, H., 2018. Integrating Visual and Textual Affective Descriptors for Sentiment Analysis of Social Media Posts. s.l., *IEEE Conference on Multimedia Information Processing and Retrieval*.
72. Davis, C. A., Ferrara, E., Flammini, A., Menczer, F., 2016. BotOrNot: A system to evaluate social bots. s.l., *25th International Conference Companion on World Wide Web*.

73. Davis, H. & Mohammad, S. M., 2014. Generating Music from Literature. Gothenburg, EACL, Whorkshop on Computational Liguistics for Literature.
74. Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., Harshman, R., 1990. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6), pp. 391-407.
75. Deloitte, 2017. Dark analytics: Illuminating opportunities hidden within unstructured data. [Ηλεκτρονικό] Available at: <https://www2.deloitte.com/us/en/insights/focus/tech-trends/2017/dark-data-analyzing-unstructured-data.html?id=us:2el:3dc:dup3707:awa:cons:tt17#endnote-2> [Πρόσβαση 23 6 2020].
76. Ding, Z. & Fei, M., 2013. An Anomaly Detection Approach Based on Isolation Forest Algorithm for Streaming Data using Sliding Window. s.l., Zhiguo Ding, Minrui Fei, An Anomaly Detection Approach Based on Isolation ForeIFAC Proceeding.
77. Doherty, M. A., 2000. Nazi wireless prpaganda Lord Haw-Haw and British Public Opinion in the Second World War. 1st Edition. Edinburgh: Edinburgh University Press.
78. Doob, L. W., 1948. Public opinion and propaganda. New York: Henry Holt.
79. Dovring, K. & Lasswell, H., 1959. Road of propaganda. New York: Philosophical Library, Inc.
80. Dunaway, J., 2013. Media ownership and storytone in campaign news. *American Politics Research*, 41(1), pp. 24-53.
81. Edwards, D. & Cromwell, D., 2003. Mass Deception: How the Media Helped the Government Deceive the People. *Tell me Lies: Propaganda and Media Distortion in the Attack on Iraq*. London: Pluto Press, pp. 2010-2014.
82. Ekman, P., 1999. Basic emotions. *Handbook of Cognition and Emotions*. s.l.: John Wiley, pp. 45-60.
83. Ellul, J. 1965. *Propaganda: The formation of men's attitudes*. Trans. By Konrad Kellen and Jean Learner. New York: Random House Vintage Books.
84. Engelberg, J. & Parsons, C., 2011. The causal impact of media in financial markets. *The Journal of Finance*, 66(1), pp. 67-97.
85. Enge, R. & Ng, V., 1993. Measuring and testing the impact of news on volatility. *The Journal of Finance*, 48(5), pp. 1749-1778.
86. Erfani, A. M., Rajasegarar, S., Karunasekera, S. & Leckie, C., 2016. High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning. *Pattern Recognition*, Volume 58, pp. 121-134.

87. Eshbaugh-Soha, M., 2010. The tone of local presidential news coverage. *Political Communication*, 27(2), pp. 121-140.
88. Esuli, A. & Sebastiani, F., 2006. SentiWordNetQ A Publicly Available Lexical Resource for Opinion Mining. Genoa, International Conference on Language Resources and Evaluation.
89. Esuli, A. & Sebastiani, F., 2014. Sentiwordnet: a publicly available lexical resource for opinion mining. s.l., LREC.
90. Fadel, I. A. & Öz, C., 2020. A Sentiment Analysis. *Sakarya University Journal of Science*, 24(6), pp. 1294-1302.
91. Farwell, J. P., 2014. The Media Strategy of ISIS. *Survival*, 56(6), pp. 49-55.
92. Fernandez, A., 2015. Here to stay and growing: Combating ISIS propaganda networks, Washington: Brooking Institution.
93. Ferrara, A., 2015. Manipulation and Abuse on social media. *ACM SIGWEB Newsletter*, 4(Spring), pp. 1-9.
94. Ferrara, E., 2017. Contagion dynamics of extremist propaganda in social networks. *Information Sciences*, Volume 418-419, pp. 1-12.
95. Ferrara, E., Wang, W., Varol, O., Flammini, A., Galstyan, A., 2016. Predicting Online Extremism, Content Adopters, and Interaction Reciprocity. s.l., *Social Informatics 2016. Lecture Notes in Computer Science*, vol 10047. Springer.
96. Fisher, A., 2015. Swarmcast: How Jihadist Networks Maintain a Persistent Online Presence. *Perspectives on Terrorism*, 9(3), pp. 3-20.
97. Gambhir, H., 2016. *The Virtual Caliphate: ISIS's Information Warfare*, Washington, DC: Institute for the Study of War.
98. Gambhir, H., Zimmerman, K. & Cafarella, J., 2016. Jabhat al Nusra and ISIS: Sources of Strength. Available at: <https://www.understandingwar.org/report/jabhat-al-nusra-and-isis-sources-strength-0>
99. Gao, H., Tate, M., Zhang, H., Chen, S., Liang, B., 2018. Social media ties strategy in international branding: An application of resource-based theory. *Journal of International Marketing*, 26(3), pp. 45-69.
100. Gates, S. & Sukanya P., 2015. Social Media, Recruitment, Allegiance and the Islamic State. *Perspectives on Terrorism*, 9(4), pp. 107-116.

101. Ghajar-Khosravi, S., Kwantes, P., Derbentseva, N. & Huey, L., 2016. Quantifying Salient Concepts Discussed in Social Media Content: An Analysis of Tweets Posted by ISIS Fangirls. *Journal of Terrorism Research*, 7(2), pp. 79-90.
102. Ghanbari-Adivi, F. & Mosleh, M., 2019. Text emotion detection in social networks using a novel ensemble classifier based on Parzen Tree Estimator (TPE). *Neural Computing and Applications*, pp. 1-13.
103. Gialampoukidis, I., Kalpakis, G., Tsirikika, T., Vrochidis, S., Kompatsiaris, I., 2016. Key player identification in terrorism-related social media networks using centrality measures. s.l., 2016 European Intelligence and Security Informatics Conf. (EISIC).
104. Gohil, S., Vuik, S. & Darzi, A., 2018. Sentiment Analysis of Health Care TweetsQ Review of the Methods Used. *JMIR Public Health and Surveillance*, 4(2), pp. 72-81.
105. Gonzalez-Castro, V., Rodriguez, R. A., Alegre, E., 2013. Class distribution estimation based on the Hellinger distance. *Information Sciences*, Volume 2018, pp. 146-164.
106. Goodman, M., 2015. Dark web revealed. *Popular Science*, 1 April.
107. Gorokhov, O., Petrovskiy, M. & Mashechkin, I., 2017. Convolutional Neural Networks for Unsupervised Anomaly Detection in Text Data. Berlin, Springer.
108. Greenwald, T., 2013. Data Won the U.S. Election. Now Can It Save the World? [Ηλεκτρονικό] Available at:

<https://www.technologyreview.com/2013/05/29/178302/data-won-the-us-election-now-can-it-save-the-world/> [Πρόσβαση 20 6 2020].
109. Griesser, S. E. & Gupta, N., 2019. Triangulated Sentiment Analysis of Tweets for Social CRM. s.l., 6th Swiss Conference on Data Science.
110. Griffith, E., 2017. 90 percent of the Big Data we generate is an unstructured mess. [Ηλεκτρονικό] Available at: <https://www.pcmag.com/news/90-percent-of-the-big-data-we-generate-is-an-unstructured-mess> [Πρόσβαση 20 6 2020].
111. Grijalva, E., Newman, D. A., Louis, T., Brent, D. M., Harms, P. D., Richard W., R., Taiyi, Y., 2015. Gender differences in narcissism: A meta-analytic review. *Psychological Bulletin*, 141(2), pp. 261-310.
112. Grover, P., Kar, A. K., Dwivedi, Y. K. & Janssen, M., 2017. The untold story of USA presidential elections in 2016 - Insights from twitter analytics. s.l., Digital Nations – Smart

Cities, Innovation, and Sustainability - 16th IFIP Conference on eBusiness, e-Services, and e-Society.

113. Guardian, 2018. Leaked: Cambridge Analytica's blueprint for Trump victory. [Ηλεκτρονικό] Available at:

<https://www.theguardian.com/uk-news/2018/mar/23/leaked-cambridge-analyticas-blueprint-for-trump-victory> [Πρόσβαση 20 6 2020].

114. Guesalaga, R., 2016. The use of social media in sales: Individual and organizational antecedents, and the role of customer engagement in social media. *Industrial Marketing Management*, Volume 54, pp. 71-79.

115. Gupta, N., Girbert, M. & Di Fabrizio, G., 2012. Emotion detection in email customer care. *Computational Intelligence*, 29(3), pp. 10-16.

116. Hamilton, M., Kaltcheva, V. D. & Rohm, A. J., 2016. Social media and value creation: The role of interaction satisfaction and interaction immersion. *Journal of Interactive Marketing*, Volume 36, pp. 121-133.

117. Haselmayer, M. & Jenny, M., 2017. Sentiment analysis of political communication: combining a dictionary approach with crowdcoding. *Quality and Quantity*, Volume 51, pp. 2623-2646.

118. Heigl, M., Anand, K. A., Urmann, A., Fiala, D., Schramm, M., Hable, R., 2021. On the Improvement of the Isolation Forest Algorithm for Outlier Detection with Streaming Data. *On the Improvement of the Isolation Forest. Electronics*, Volume 10, pp. 1-26.

119. Henderson, E. H., 1943. Toward definition of propaganda. *Journal of Social Psychology*, Volume 18, pp. 71-87.

120. Heuser, R., Moretti, F. & Steiner, E., 2016. *The emotions of London*, s.l.: Stanford Literary Lab Pamphlets.

121. Hitler, A., 1925. *Mein Kampf*. 1η έκδοση (2006) Αθήνα: Κάκτος.

122. Hoffman, 2019. ISIS' Shifting Focus. [Ηλεκτρονικό] Available at: https://www.thecipherbrief.com/column_article/isis-shifting-focus [Πρόσβαση 23 2 2021].

123. Hofmann, T., 1999. *Probabilistic Latent Semantic Analysis*. Stockholm, Uncertainty in Artificial Intelligence UAI99.

124. Hong, S. & Kim, S. H., 2016. Political polarization on Twitter: Implications for the use of social media in digital governments. *Government Information Quarterly*, 33(4), pp. 777-782.
125. Huang, J., Lin, Z. & Liu, X., 2019. Episodic Memory Network with Self-attention for Emotion Detection. s.l., *International Conference on Database Systems for Advanced Applications*.
126. Hu, M. & Liu, B., 2004. Mining and summarizing customer reviews. s.l., *ACM SIGKDD international conference on Knowledge Discovery and Data Mining*.
127. Hutto, C. J. & Gilbert, E. E., 2014. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. s.l., *Eighth International Conference on Weblogs and Social Media (ICWSM-14)*.
128. Impara, E., 2018. A social semiotics analysis of Islamic State's use of beheadings: Images of power, masculinity, spectacle and propaganda. *International Journal of Law, Crime and Justice*, pp. 15 - 45.
129. Impara, E., 2018. A social semiotics analysis of Islamic State's use of beheadings: Images of power, masculinity, spectacle and propaganda. *International Journal of Law, Crime and Justice*, Volume 53, pp. 25-45.
130. Ingram, H., 2017. Learning from ISIS's Virtual Propaganda War for Western Muslims: A comparison of Inspire and Dabiq. *NATO Science for Peace and Security Series-Human and Societal Dynamics*, Volume 136, pp. 170-181.
131. Jacobs, A. M., 2019. Sentiment Analysis for Words and Fiction Characters from the Perspective of Computational (Neuro-)Poetics. *Frontiers in Robotics and AI*, Volume 6, pp. 1-53.
132. Jacobs, A. M., Schuster, S., Xue, S. & Lüdtke, J., 2017. What's in the brain that ink may character: a quantitative narrative analysis of Shakespeare's 154 sonnets for use in neurocognitive poetics. *7. Scientific Study of Literature*, Volume 7, pp. 4-51.
133. Jain, P. N. & Vaidya, A. V., 2021. Analysis of Social Media Based on Terrorism | A Review. *Vietnam Journal of Computer Science*, 8(1), pp. 1-21.
134. Janczewski, L. & Colarik, A. M., 2005. Managerial guide for handling Cyber-terrorism and information Warfare. Hershey: IDEA Group.
135. Janczewski, L. & Colarik, A., 2008. *Cyber Warfare and Cyber Terrorism*. Hersey: Information Science Reference.

136. Jefferson, T., 2007. The use of ICTs in regional conflicts, war, and terrorism. *VINE Journal of Information and Knowledge Management Systems*, 37(1), pp. 14-17.
137. Jhavar, H. & Mirza, P., 2018. EMOFIEL: mapping emotions of relationships in a story. Geneva, The WEB Conference 2018.
138. Jones, C. W., 2018. Understanding ISIS's Destruction of Antiquities as a Rejection of Nationalism. *Journal of Eastern Mediterranean Archaeology & Heritage Studies*, 6(1-2), pp. 31-58.
139. Joselson, N. & Hallén, R., 2019. Emotion classification with natural language processing (Comparing BERT and Bi-Directional LSTM models for use with Twitter conversations). Lund: Lund University.
140. Joseph, K., Matthew-Benigni, W. & Carley, K. M., 2016. A social event based approach to sentiment analysis of identities and behaviors in text. *The Journal of Mathematical Sociology*, Volume 40, pp. 13-166.
141. Jowett, G. S. & O'Donnell, V., 2012. *Propaganda and Persuasion*. 5th Edition, Thousand Oaks, California: SAGE.
142. Kang, D. & Park, Y., 2014. Review-based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach. *Expert Systems with Applications*, pp. 1041-1050.
143. Kannan, R., Woo, H., Aggarwal, C. C., Park, H., 2017. Outlier Detection for Text Data: An Extended Version. [Ηλεκτρονικό] Available at: <https://arxiv.org/abs/1701.01325> [Πρόσβαση 22 6 2021].
144. Kaplan, A. M. & Haenlein, M., 2010. Users of the World unite! The challenges and opportunities of social media. *Business Horizons*, Volume 53, pp. 59-68.
145. Karczmarek, P., Kiersztyn, A., Pedrycz, W. & Al, E., 2020. K-Means-based isolation forest. *Knowledge-Based Systems*, Volume 195, pp. 1-15.
146. Kauffmann, E., Peral, J. & Ferrandez, A., 2019. Managing Marketing Decision-Making with Sentiment Analysis: An Evaluation of the main product features using text data mining. *Sustainability*, 11(15), pp. 1-19.
147. Kazemian, S., Zhao, S. & Penn, G., 2016. Evaluating sentiment analysis in the context of securitising. s.l., Annual meeting of Association of Computational Linguistics.
148. Kenez, P., 1985. *The Birth of the Propaganda State: Soviet Methods of Mass Mobilization, 1929*, Cambridge: Cambridge University Press.

149. Khreich, W., Khosravifar, B., Hamou-Lhadj, A. & ChamseddineTalhi, C., 2017. An anomaly detection system based on variable N-gram features and one-class SVM. *Information and Software Technology*, Volume 91, pp. 186-197.
150. Kim, E. & Klinger, R., 2018. Who feels, what, and why? Annotation of a literature corpus with semantic roles of emotions. Santa Fe, 27th International Conference on Computational Linguistics.
151. Kim, E. & Klinger, R., 2019. An analysis of emotion communication channels in fan-fiction: Towards emotional storytelling. Florence, 2nd Workshop of Storytelling.
152. Kim, S., Li, F., Lebanon, G. & Essa, I., 2013. Beyond SentimentQ: The Manifold of Human Emotions. Kim, S.; Li, F.; Lebanon, G.; Essa, I. Beyond SentimentQ: The Manifold of Human Emotions. 2013. Proceedings of the 16th InternationalScottsdale, AZ, USA, Kim, S.; Li, F.; Lebanon, G.; Essa, I. Beyond SentimentQ: The Manifold of Hum16th International Conference on Artificial Intelligence and Statistics (AISTATS).
153. Kinney, A. B., Davis, A. P. & Zhang, Y., 2018. Theming for terror: Organizational adornment in terrorist propaganda. *Poetics*.
154. Kiritchenko, S., Zhu, X., Cherry, C. & Mohammad, S., 2014. Nrc-canada-2014: Detecting aspects and sentiment in customer reviews. s.l., 8th International Workshop on Semantic Evaluation.
155. Klausen, J., Marks, C. E. & Zaman, T., 2018. Finding Extremists in Online Social Networks. *Operations Research*, 66(4), pp. 957-976.
156. Kohavi, R., 1995. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. s.l., Morgan Kaufmann Publishers Inc., pp. 1137-1143.
157. Kontostathis, A., Reynolds, K., Garron, A. & Edwards, L., 2013. Detecting cyberbullying: query terms and techniques. s.l., 5th Annual ACM Web Science Conference, pp. 195-204.
158. Krstovski, K., Smith, D., Wallach, H. & McGregor, A., 2013. Efficient Nearest-Neighbor Search in the Probability Simplex. s.l., Conference on the Theory of Information Retrieval.
159. Krstovski, K. & Smith, D., 2013. Online Polylingual Topic Models for Fast document Translation Detection. s.l., Eighth Workshop on Statistical Machine Translation.
160. Lasswell, H. D., 1927. The theory of political propaganda. *The American Political Science Review*, 21(3), pp. 627-631.
161. Le Bon, G., 1895. *Psychologie des foules [The Crowd]*. Paris: F. Alcan.

162. Lebet, R. & Collobert, R., 2014. Word Embeddings through Hellinger PCA. Gothenburg, Sweden, Association for Computational Linguistics.
163. Lieberman, A. V., 2017. Terrorism, The Internet, and Propaganda: A deadly combination. *Journal of national Security Law and Policy*, 9(95), pp. 95-124.
164. Li, S., Song, S. & Huang, G., 2017. Prediction Reweighting for Domain Adaptation. *IEEE Transactions on Neural Networks and Learning Systems*, 28(7), pp. 1682-1695.
165. Liu, B., 2010. Sentiment Analysis and Subjectivity. *Handbook of Natural Language Processing*, 2nd Edition.
166. Liu, B., 2012. Sentiment Analysis and Opinion Mining. *Synthesis lectures on human language technologies*, 5(1), pp. 1-167.
167. Liu, B., 2015. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. New York: Cambridge University Press.
168. Liu, B., 2017. Many facets of Sentiment Analysis. Στο: E. Cambria, D. Das, S. Bandyopadhyay & A. Feraco, επιμ. *A Practical Guide to Sentiment Analysis, Socio-Affective Computing 5*. s.l.:Springer, pp. 11-39.
169. Loyola-Gonzalez, O., Monroy, R., Rodriguez, J., Lopez-Cuevas, A., Mata-Sanchez, J. I., 2019. Contrast Pattern-Based Classification for bot detection on twitter. *IEEE Access*, 45817(7), p. 45800.
170. Luo, L., Ao, X., Pan, F., Wang, J., Zhao, T., Yu, N., He, Q., 2018. Beyond Polarity: interpretable Financial Sentiment Analysis with Hierarchical query-driven attention. s.l., 27th Joint Confernece on Artificial Intelligence.
171. Mac Kim, S. & Calvo, R. A., 2010. Sentiment Analysis in Student Experiences of Learning. Pittsburg, 3rd International Conference on Educational Data Mining.
172. Ma, D., Li, S., Zhang, X. & Wang, H., 2017. Interactive attention networks for aspect-level sentiment classification. s.l., IJCAI.
173. Magdy, W., Darwish, K. & Weber, I., 2015. #FailedRevolutions: Using Twitter to Study the Antecedents of ISIS Support. [Ηλεκτρονικό] Available at: <https://arxiv.org/abs/1503.02401> [Πρόσβαση 20 8 2021].
174. Mahalanobis, P. C., 1936. On the generalised distance in statistics. Calcutta, National Instiute of Sciences of India, pp. 49-55.
175. Mahood, S. & Rane, H., 2017. Islamist narratives in ISIS recruitment propaganda. *The Journal of International Communication*, 23(1), pp. 15-35.

176. Malik, S., 2015. The Isis papers: leaked documents show how Isis is building its state, s.l.: The Guardian.
177. Mangold, W. G. & Faulds, D., 2009. Social Media: The new hybrid element of the promotion mix. *Business Horizons*, Volume 52, pp. 357-365.
178. Mansour, S., 2018. Social Media Analysis of User's Responses to Terrorism using Sentiment Analysis and Text Analysis. *Procedia Computer Science*, Volume 140, pp. 95-103.
179. Martin-Valdivia, M.-T., Martínez-CámaraJ, E., Perea-Ortega, J.-M. & Ureña-López, L. A., 2013. Sentiment polarity detection in Spanish reviews combining supervised and unsupervised approaches. *Expert Systems with Applications*, 40(10), pp. 3934-3942.
180. McCabe, T. R., 2016. A Strategy for the ISIS Foreign Fighter Threat. *Orbis*, pp. 140 - 153.
181. McClelland, J. S., 1989. *The Crowd and the Mob*. London: Unwin. London: Unwin.
182. McElreath, D. H., Doss, D. A., McElreath, L. S. & Lindsley, A., 2020. The Communicating and Marketing of Radicalism: A Case Study of ISIS and Cyber Recruitment. *International Journal of Cyber Warfare and Terrorism*, 8(3), pp. 26-45.
183. McFate, J., 2015. The ISIS Defense in Iraq and Syria: Countering an Adaptive Enemy. [Ηλεκτρονικό] Available at: <http://understandingwar.org/report/isis-defense-iraq-and-syria-countering-adaptive-enemy> [Πρόσβαση 10 1 2021].
184. McLaine, I., 1979. *Ministry of Morale*. London: Allen & Unwin.
185. Meiselman, H. L. επιμ., 2016. *Emotion Measurement*. 1 επιμ. s.l.:Woodhead Publishing.
186. Merriam & Webster, 2021. *Merriam Webster Dictionary*. [Ηλεκτρονικό] Available at: (<https://www.merriam-webster.com/dictionary> [Πρόσβαση 2021].
187. Merry, M., 2016. Making friends and enemies on social media: the case of gun policy organizations. *Online Information Review*, 40(5), pp. 624-642.
188. Metaxas, P. & Mustafaraj, E., 2012. Social Media and the elections. *Science*, Volume 338, pp. 472-473.
189. Miller, C. R., 1939. *How to detect and analyze propaganda*. New York: Town Hall, Inc.
190. Minamikawa, A. & Yokohama, H., 2011. Personality Estimation Based on Weblog Text Classification. s.l., *International Conference on Industrial Engineering and Other Applications of Applied Intelligent Systems*.

191. Mitchum, R., 2013. Rayid Ghani, Obama campaign chief data scientist, joins UChicago. [Ηλεκτρονικό] Available at: <https://news.uchicago.edu/story/rayid-ghani-obama-campaign-chief-data-scientist-joins-uchicago> [Πρόσβαση 20 6 2020]
192. Moghaddam, S., 2015. Beyond sentiment analysis: mining defects and improvements from customer feedback. s.l., European Conference on Information Retrieval.
193. Mohammad, S. M., 2016. Sentiment Analysis: Detecting Valence, Emotions, and Other Affectual States from Text. Στο: H. Meiselman, επιμ. Emotion Measurement. s.l.:Elsevier, pp. 201-237.
194. Mohammad, S. M. & Turney, P. D., 2010. Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon. LA, California, NAACL - HLT.
195. Mohammad, S. M. & Turney, P. D., 2013. Crowdsourcing a Word-Emotion Association Lexicon. Computational Intelligence, Volume 29, pp. 436 - 465.
196. Mohammad, S. M., Zhu, X., Kiritchenko, S. & Martin, J. D., 2014. Sentiment, emotion, purpose, and style in electoral tweets. Information Processing & Management, 51(4), pp. 1-38.
197. Moody-Ramirez, M. & Cole, H., 2018. Victim Blaming in Twitter Users' Framing of Eric Garner and Michale Brown. Journal of Black Studies, 49(4), pp. 383-407.
198. Morales, A. J., Borondo, J., Losada, J. & Benito, R., 2015. Measuring political polarization: Twitter shows the two sides of Venezuela. Chaos: An Interdisciplinary journal on nonlinear sciences, 25(1), pp. 1-20.
199. Nahar, V., Al-Maskari, V., Li, X. & Pang, C., 2014. Semi-supervised Learning for Cyberbullying Detection in Social Networks. Στο: W. H. & S. M.A., In: Wang H., Sharaf M.A. (eds) Databases Theory and Applications. Lecture Notes in Computer Science. s.l.: Springer, pp. 160-171.
200. Nahar, V., Unankard, S., Li, X. & Pang, C., 2012. Sentiment Analysis for Effective Detection of Cyber Bullying. In: S. Q.Z., W. G., J. C.S. & X. G., Lecture Notes in Computer Science, vol 7235. Web Technologies and Applications. s.l.:Springer, pp. 767-774.
201. Najafabadi, M., Khoshgoftaar, T., Calvert, C. & Kemp, C., 2017. A Text Mining Approach for Anomaly Detection in Application Layer DDoS Attacks. Florida, Thirtieth International Florida Artificial Intelligence Research Society.

202. Nepalli, V. K., Caragea, C. S. A., Tapia, A. & Stehle, S., 2017. Sentiment analysis during hurricane Sandy in emergency response. *International Journal of Disaster Risk Reduction*, Volume 21, pp. 213-222.
203. Nesser, P., Stenetsen, A. & Oftedal, E., 2016. Jihadi Terrorism in Europe: The IS-Effect. *Perspectives on Terrorism*, 10(6), pp. 1-24.
204. NewYorkTimes, 2018. Cambridge Analytica and Facebook: The Scandal and the Fallout So Far. [Ηλεκτρονικό] Available at:

<https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html>
205. NIST, 2019. NIST Big Data Interoperability framework: Volume 1, Definitions. [Ηλεκτρονικό] Available at:

<https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1r2.pdf> [Πρόσβαση 6 1 2020].
206. Nopp, C. & Hanbury, A., 2015. Detecting risks in banking system by sentiment analysis. s.l., Conference on Empirical Methods in Natural Language Processing.
207. Nuñez, M., 2019. to-facebook-exakolouthei-na-diarreei-dedomena-ena-xrono-meta-to-skandalo-tis-cambridge-analytica. [Ηλεκτρονικό] Available at:

<https://www.capital.gr/forbes/3391918/to-facebook-exakolouthei-na-diarreei-dedomena-ena-xrono-meta-to-skandalo-tis-cambridge-analytica> [Πρόσβαση 7 3 2021].
208. Oates, B., 2003. The potential contribution of ICTs to the political process. *Electronic Journal of e-Government*, 1(1), pp. 31-39.
209. Oliveira, D. J. S., Bermejo, P. H. d. S. & dos Santos, P. A., 2017. Can social media reveal the preferences of voters? A comparison between sentiment analysis and traditional opinion polls. *Journal of Information Technology & Politics*, 14(1), pp. 34-45.
210. Oramas, R., Barron-Estrada, M. L. & Cabada, R. Z. R.-A. S. L., 2018. A Corpus for Sentiment Analysis and Emotion Recognition for a Learning Environment. s.l., IEEE 18th International Confernece on Advanced Learning Technologies.
211. Ortony, A., Clore, G. & Collins, A., 1990. *The Cognitive Structure of Emotions*. Cambridge: Cambridge University Press.

212. Ott, B. L., 2017. The age of Twitter: Donald J. Trump and the politics of debasement. *Critical Studies in Media Communication*, 34(1), pp. 59-68.
213. Ozeren, S., Hekimb, H., Salih-Elmasc, M. & Canbegid, H. I., 2018. An Analysis of ISIS Propaganda and Recruitment Activities Targeting the Turkish-Speaking population. *International Annals of Criminology*, Volume 56, pp. 105-121.
214. Pang, B., Lee, L., 2008. Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*. Volume 2, pp. 1-135.
215. Parry-Giles, S. J., 2002. *The rhetorical presidency, propaganda, and the Cold War: 1945-1955*. Westport, CT: Praeger..
216. Pathak, X. & Pathak-Shelat, M., 2017. Sentiment analysis of virtual brand communities for effective tribal marketing. *Journal of Research in Interactive Marketing*, 11(1), pp. 16-38.
217. Paul, M. & Dredze, M., 2011. You Are What You Tweet: Analysing Twitter for Public Health. s.l., Fifth International AAAI Conference on Weblogs and Social Media.
218. Peng, L., Geng, C., Mengzhou, Z., Chunyu, L., 2014. What do seller manipulations of online product reviews mean to consumer? Hong Kong: Hong Kong Institute of Business Studies 070-1314.
219. Plutchik, R., 1980. *Theories of Emotion*. 2nd Edition. Michigan: Academic Press.
220. Ponciano, J., 2021. Capital.gr. [Ηλεκτρονικό] Available at: <https://www.capital.gr/forbes/3527820/i-tesla-exase-170-dis-kefalaioiisi-meta-tin-ependusi-sto-bitcoin> [Πρόσβαση 1 3 2021].
221. Poria, S., Gelbukh, A., Cambria, E. & Hussain, A. H. G., 2014. EmoSenticSpace: a novel framework for affective common-sense reasoning. *Knowledge Based Systems*, Volume 69, pp. 108-123.
222. Prucha, N., 2016. IS and the Jihadist Information Highway – Projecting Influence and Religious Identity via Telegram. *Perspectives on Terrorism*, 10(6), pp. 48-58.
223. Qualter, T. H. 1962. *Propaganda and psychological warfare*. New York: Random House.
224. Ragini, J. R., Rudesh- Anand, P. M. & Bhaskar, V., 2018. Big data analytics for disaster response and recovery through sentiment analysis. *International Journal of Information Management*, Volume 42, pp. 13-24.
225. Rana, T. & Cheah, Y., 2016. Aspect extraction in sentiment analysis: A comparative analysis and survey. *Artificial Intelligence Review*, Volume 46, pp. 459-463.

226. Rangaswamy, A., Moch, N., Felten, C., Bruggen, G., Wieringa, J. E., Wirtz, J., 2020. The Role of Marketing in Digital Business Platforms. *Journal of Interactive Marketing*, Volume 51, pp. 72-90.
227. Rao, V., 2019. Suicide Prediction on Social Media by implementing Sentiment Analysis along Machine Learning. *International Journal of Recent Technology and Engineering*, 8(2), pp. 4833-4837.
228. Rapp, A. B. L. S., Grewal, D. & Hughes, D. E., 2013. Understanding social media effects across seller, retailer, and consumer interactions. *Journal of the Academy of Marketing Science*, 41(5), pp. 547-566.
229. Ratkiewicz, P., Conover, M., Meiss, M., Goncalves, B., 2011. Detecting and tracking political abuse in social media. s.l., 5th International AAAI Conference on Weblogs and Social Media.
230. Ravi, K. & Ravi, V., 2015. A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-Based Systems*, Volume 89, pp. 14-46.
231. Reagan, A. J., Mitchell, L., Kiley, D., Danforth, C. M., Dodds, P. S., 2016. The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Science*, 5(1), pp. 31-43.
232. Rehurek, R. & Sojka, P., 2010. Software Framework for Topic Modelling with Large Corpora. Valletta, Malta, ELRA, pp. 45 - 50.
233. Rekabsaz, N., Lupu, M., Baklanov, A., Hanbury, A., Duer, A., Anderson, L., 2017. Volatility prediction using financial disclosures sentiments with word embedding-based models. s.l., Annual meeting of ACL.
234. Richards, E. G., 2012. Calendars. Στο: S. E. Urban & P. K. Seidelmann, επιμ. Explanatory Supplement to the Astronomical Almanac. 3 επιμ. Mill Valley, CA: University Science Books, pp. 585 - 624.
235. Russell, J. A., 2003. Core affect and the psychological construction of emotion. *Psychological Review*, pp. 145-172.
236. Sakaki, T., Okazaki, M. & Matsuo, Y., 2010. Earthquake shakes Twitter users: real-time event detection by social sensors. s.l., ACM. 19th International conference on WWW.
237. Salton, G. & McGill, M., 1983. Introduction to Modern Information Retrieval 1st Edition. New York.: McGraw-Hill Book Co.
238. Sanders, M. L., 1975. Wellington House and British Propaganda during the First World War. *Historical Journal*, 18(1), pp. 119-146.

239. Schmitt, M. F. L. & Spinosa, E. J., 2018. Outlier Detection on Semantic Space for Sentiment Analysis with Convolutional Neural Networks. s.l., 2018 International Joint Conference on Neural Networks (IJCNN), pp. 1-8.
240. Schwartz, H., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Agrawal, M., Park, G. J., Lakshminanth, S. K., Jha, S., Seligman, M. E. P., Ungar, L., 2013. Characterizing geographic variation in well-being using tweets. s.l., AAAI Conference on weblogs and social media.
241. Seroussi Y., Zukerman, I., Bohnert, F., 2010. Collaborative Inference of Sentiments from Texts. User Modeling, Adaptation, and Personalization, 18th International Conference, UMAP 2010, Big Island, HI, USA, June 20-24, 2010. Proceedings
242. Seliya, N., Abdollah Zadeh, A. & Khoshgoftaar, T.M. A literature review on one-class classification and its potential applications in big data. *Big Data* 8, 122 (2021). <https://doi.org/10.1186/s40537-021-00514-x>
243. Shabad, R., 2016. John Kerry insists U.S. is "winning" ISIS fight despite Brussels attacks. [Ηλεκτρονικό] Available at: <https://www.cbsnews.com/news/john-kerry-insists-u-s-is-winning-isis-fight-despite-brussels-bombings/> [Πρόσβαση 10 3 2021].
244. Khawaja, A. S. & Khan, A. H., 2016. Media Strategy of ISIS: An Analysis. *Strategic Studies*, 36(2), pp. 104-121.
245. Smith, T. J., 1989. Propaganda: A pluralistic perspective. 3rd Edition. New York: Praeger.
246. Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A. Y., Potts, C., 2013. Recursive Deep Models for Semantic Compositionality over a Sentiment Treebank. Seattle, Conference on Empirical Methods in Natyral Processing.
247. Soroka, S. & McAdams, S., 2015. News, politics, and negativity. *Political Communication*, 32(1), pp. 1-22.
248. Speckhard, A., Shajkovci, A., Wooster, C. & Neima Izadi, N., 2018. Mounting a Facebook Brand Awareness and Safety Ad Campaign to Break the ISIS Brand in Iraq. *Perspectives on Terrorism*, 12(3), pp. 50-66.
249. Staiano, J., Guerini, M. & arXiv:1405.1605., 2. a. p., 2014. Depechemood: a lexicon for emotion analysis from crowd-annotated news. [Ηλεκτρονικό] Available at: <https://arxiv.org/abs/1405.1605> [Πρόσβαση 24 6 2021].
250. Strapparava, C. & Valitutti, A., 2004. Wordnet affect: an affective extension of wordnet. Lisbon, 4th International Conference on Language Resources and Evaluation, pp. 1083-1086.

251. Suero-Montero, C. & Suhonen, J., 2014. Emotion analysis meets learning analytics: Online learner profiling beyond numerical data. Koli, 14th Koli Calling International Conference on Computing Education Research.
252. Svete, U., 2008. Use and impact of Information-Communication Technology in modern conflict: The war in Iraq. G. K. G. Caforio & B. Purkayastha, επιμ. *Armed Forces and Conflict Resolution: Sociological Perspectives (Contributions to Conflict Management, Peace Economics and Development, Vol. 7)*. Bingley: Emerald, pp. 75-96.
253. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M., 2011. Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, 37(2), pp. 267-307.
254. Talafha, B., Al-Ayyoub, M. & Abuammar, A. J. Y., 2019. Outperforming State-of-the-Art Systems for Aspect-Based Sentiment Analysis. s.l., 16th International Conference on Computer Systems and Applications.
255. Tate-Goins, E., 2011. Promoting Unity Through Propaganda: How the British Government Utilized Posters During the Second World War. [Ηλεκτρονικό] Available at: https://digitalcommons.wku.edu/cgi/viewcontent.cgi?article=1346&context=stu_hon_tt_hese [Πρόσβαση 20 March 2021].
256. Taylor, P. M., 1990. *Munitions of the Mind: A History of Propaganda 1st Edition*, Manchester University Press: Manchester University Press.
257. Thudumu, S., Branch, P., Jin, J., Singh, J. J., 2020 A comprehensive survey of anomaly detection techniques for high dimensional big data. *Journal of Big Data* 7, 42. <https://doi.org/10.1186/s40537-020-00320-x>
258. Van Der Meer, T. & Verhoeven, P., 2013. Public framing organizational crisis situations: Social media versus news media. *Public Relation Review*, Volume 39, pp. 229-231.
259. Vareldzhan, G., Yurkov, K., Ushenin, K., 2021. Anomaly Detection in Image Datasets Using Convolutional Neural Networks, Center Loss, and Mahalanobis Distance. [Ηλεκτρονικό] Available at: <https://arxiv.org/abs/2104.06193> [Πρόσβαση 5 10 2021]
260. Vergeer, M. R. M. & Hermans, E. A. H. M., 2013. Campaigning on Twitter: Microblogging and online social networking as campaign tools in the 2010 general elections in the Netherlands. *Journal of Computer-mediated Communication*, 18(4), pp. 399-419.
261. Vesset, D. & Schubmehl, 2016. *IDC FutureScape: Worldwide big data, business analytics, and cognitive*, s.l.: International Data Corporation.

262. Wachsmuth, H., Kiesel, J. & Stein, B., 2015. Sentiment Flow – A General Model of Web Review Argumentation. s.l., Wachsmuth, H.; Kiesel, J.; Stein, B. Conference on Empirical Methods in Natural Language Processing.
263. Wang, Z., Tong, J., Xin, X., Vhin, H., 2014. Anomaly Detection through Enhanced Sentiment Analysis on Social Media Data. s.l., The International Conference on Cloud Computing Technology and Science, CloudCom.
264. Wasid, M. & Ali, R., 2018. An improved recommender system based on multi-criteria clustering approach. *Procedia Computer Science*, Volume 131, pp. 93-101.
265. Wei-Fan, C. & Lun-Wei, K., 2016. A Deep Learning Model of Stance Classification on Social Media. Osaka, Conference on Computational Linguistics 2016.
266. Weimann, G., 2006. *Terror on the Internet: The New Arena, the New Challenges*, Washington DC: United States Institute of Peace Press.
267. Welch, D., 2016. *Persuading the People British Propaganda in World War II*. 1st Edition. London: British Library Publishing.
268. Wells, C. και συν., 2016. How Trump Drove Coverage to the Nomination: Hybrid Media Campaigning. *Political Communication*, 33(4), pp. 669-676.
269. Wertz, J., 2018. Why Sentiment Analysis Could Be Your Best Kept Marketing Secret. [Ηλεκτρονικό] Available at: <https://www.forbes.com/sites/jiawertz/2018/11/30/why-sentiment-analysis-could-be-your-best-kept-marketing-secret/#6defded12bbe>
270. Whitton, J. B., 1951. "Cold War Propaganda.". *The American Journal of International Law*, 45(1), pp. 151-153.
271. Wiebe, J. M., 1994. (1994). Tracking point of view in narrative. *Computational Linguistics*, 20(2), pp. 233-287.
272. Windsor, L., 2020. The Language of Radicalization: Female Internet Recruitment to Participation in ISIS Activities. Leah Windsor (2020) The Language of Radicalization: Female Internet Recruitment and Political Violence, 32(3), pp. 506-538.
273. Winkler, A., 1978. *The Politics of Propaganda: Office of War Information, 1942-1945*. New Heaven: Yale University Press.
274. Winter, C., 2015. *The Virtual 'Caliphate': Understanding Islamic State's Propaganda Strategy*. London: Quilliam.
275. Winter, C., 2020. Redefining 'Propaganda': The Media Strategy of the Islamic State. *The RUSI Journal*, 165(1), pp. 38-42.

276. Xiong, S., Zhang', Y., Ji, D. & Lou, Y., 2016. Distance metric learning for aspect phrase grouping. s.l., International Conference on Computational Linguistics (COLING 2016).
277. Xu, D., Wang, Y., Meng, Y. & Zhang, Z., 2017. An Improved Data Anomaly Detection Method Based on Isolation Forest. s.l., 10th International Symposium on Computational Intelligence and Design (ISCID).
278. Xu, Y., Pan, S. J., Xiong, H., Wu, Q., Luo, R., Min, H., Song, H., 2017. A Unified Framework for Metric Transfer Learning. IEEE Transactions on Knowledge and Data Engineering, 29(6), pp. 1158-1171.
279. Yaakub, M. R. & Jinglan-Zhang, Y. L., 2013. Integration of Sentiment Analysis into Customer Relational Model: The Importance of Feature Ontology and Synonym. Procedia Technology, vol11, Volume 11, pp. 495-501.
280. Yadollahi, A., Shahraki, A. G. & Zaiane, O. R., 2017. Current State of Text Sentiment Analysis from Opinion to Emotion Mining. ACM Computing Surveys, 50(2), pp. 1-33.
281. Yu, B., 2008. An evaluation of text classification methods for literary study. Literary and Linguistics Computing, 23(3), pp. 327-343.
282. Zehe, A., Becker, M., Hettinger, L., Hotho, A., Reger, I., Jannidis, F., 2016. Prediction of happy ending in german novels based on sentiment information. Riva de Garla, Workshop on Interactions between Data Mining and Natural Language Processing.
283. Zelikman, E. & Socher, R., 2020. Contextual Saliency for Fast and Accurate Sentence Vectors. [Ηλεκτρονικό] Available at: <https://arxiv.org/abs/1803.08493> [Πρόσβαση 22 8 2021].
284. Zhang, Y. N., Wells, C., Wang, S. & Rohe, K., 2018. Attention and amplification in the hybrid media system: The composition and activity of Donald Trump's Twitter following during the 2016 presidential election. New Media & Society, 20(9), pp. 3161-3182.
285. Zhuang, L. & Dai, H., 2006. Parameter Estimation of One-Class SVM on Imbalance Text Classification. In: Lamontagne L., Marchand M. (eds) Advances in Artificial Intelligence. Canadian AI 2006. Lecture Notes in Computer Science, vol 4013. Springer, Berlin, Heidelberg
286. Zuorba, H. D., Olan, C. L. O. & Cantara, A. D., 2017. A Framework for Identifying Excessive Sadness in Students through Twitter and Facebook in the Philippines. Barcelona, Spain, ICBRA 2017, pp. 52 - 56.

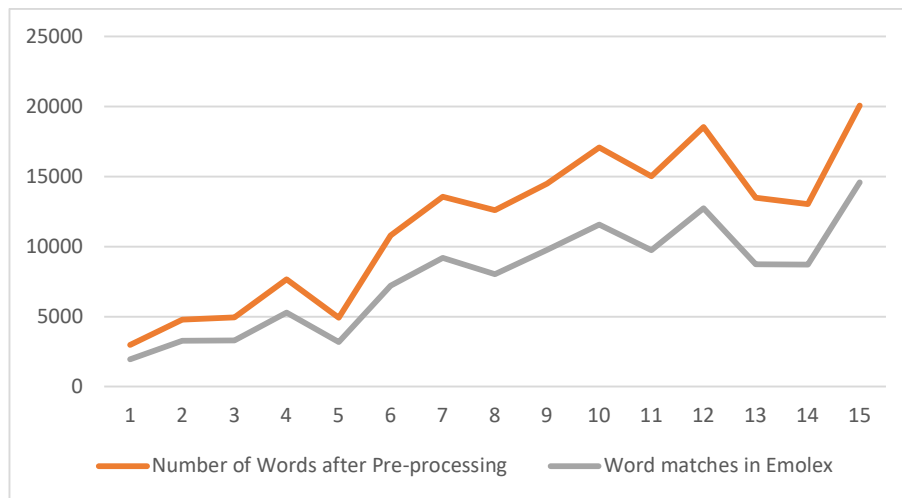
8 Παράρτημα Α – Πλήθος λέξεων και κάλυψη ανά περιοδικό και έκδοση

ΠΙΝΑΚΑΣ 22 ΠΛΗΘΟΣ ΛΕΞΕΩΝ ΚΑΙ ΚΑΛΥΨΗ ΑΠΟ ΤΟ ΕΜΟΛΕΧ ΑΝΑ ΠΕΡΙΟΔΙΚΟ ΚΑΙ ΕΚΔΟΣΗ

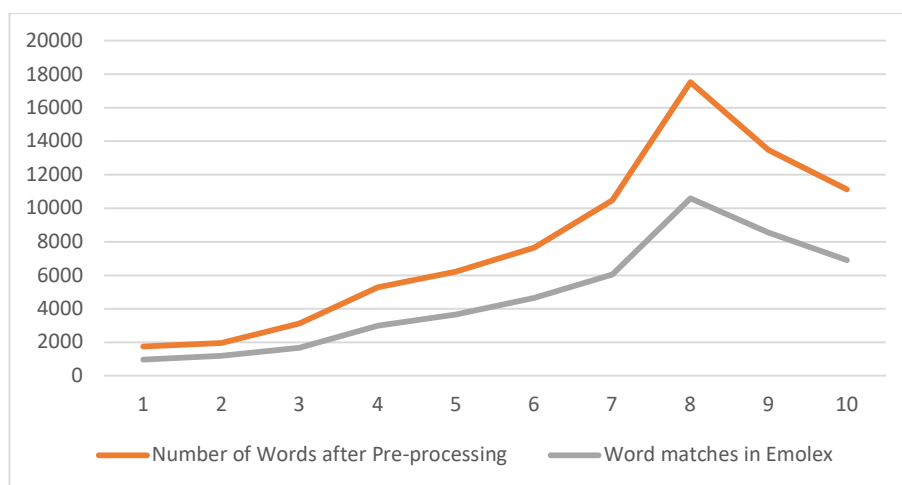
Magazine, Issue	Words	Words after Pre-processing	Word matches in EmoLex	Coverage Ratio
Dabiq 1	7100	2978	1944	65%
Dabiq 2	11564	4790	3289	69%
Dabiq 3	11961	4930	3301	67%
Dabiq 4	18896	7681	5291	69%
Dabiq 5	11723	4915	3198	65%
Dabiq 6	25655	10795	7222	67%
Dabiq 7	33118	13575	9202	68%
Dabiq 8	30518	12605	8026	64%
Dabiq 9	33570	14500	9764	67%
Dabiq 10	42743	17084	11559	68%
Dabiq 11	34514	15027	9734	65%
Dabiq 12	42814	18532	12739	69%
Dabiq 13	31812	13512	8746	65%
Dabiq 14	30670	13020	8725	67%
Dabiq 15	50515	20069	14591	73%
Daral-Islam 1	5390	1744	964	55%
Daral-Islam 2	6457	1952	1187	61%
Daral-Islam 3	10092	3113	1689	54%
Daral-Islam 4	17314	5280	2998	57%
Daral-Islam 5	21019	6220	3660	59%
Daral-Islam 6	28134	7656	4655	61%
Daral-Islam 7	34927	10476	6044	58%
Daral-Islam 8	58376	17524	10598	60%
Daral-Islam 9	43036	13475	8549	63%
Daral-Islam 10	34263	11127	6892	62%
Istok 1	11562	3110	2103	68%
Istok 2	27339	7371	5208	71%
Istok 3	43349	12521	8771	70%
Istok 4	22243	6543	4752	73%
Konstantiniyye 1	13661	5368	2928	55%
Konstantiniyye 2	19223	7751	4317	56%
Konstantiniyye 3	17740	6762	3761	56%
Konstantiniyye 4	22931	9255	5331	58%
Konstantiniyye 5	24621	9785	5354	55%
Konstantiniyye 6	22904	9209	5090	55%
Konstantiniyye 7	30207	12082	6595	55%
Rumiyah 1	23792	9435	6394	68%
Rumiyah 2	25455	10163	6983	69%
Rumiyah 3	34900	13845	9533	69%
Rumiyah 4	22973	9071	6335	70%
Rumiyah 5	29085	11914	8008	67%
Rumiyah 6	20182	8252	5552	67%
Rumiyah 7	16320	6273	4376	70%
Rumiyah 8	21189	8407	6117	73%
Rumiyah 9	32488	12485	9086	73%

Magazine, Issue	Words	Words after Pre-processing	Word matches in EmoLex	Coverage Ratio
Rumiyah 10	20705	8484	6034	71%
Rumiyah 11	36054	13839	10070	73%
Rumiyah 12	23824	9503	6749	71%
Rumiyah 13	21855	8478	6042	71%

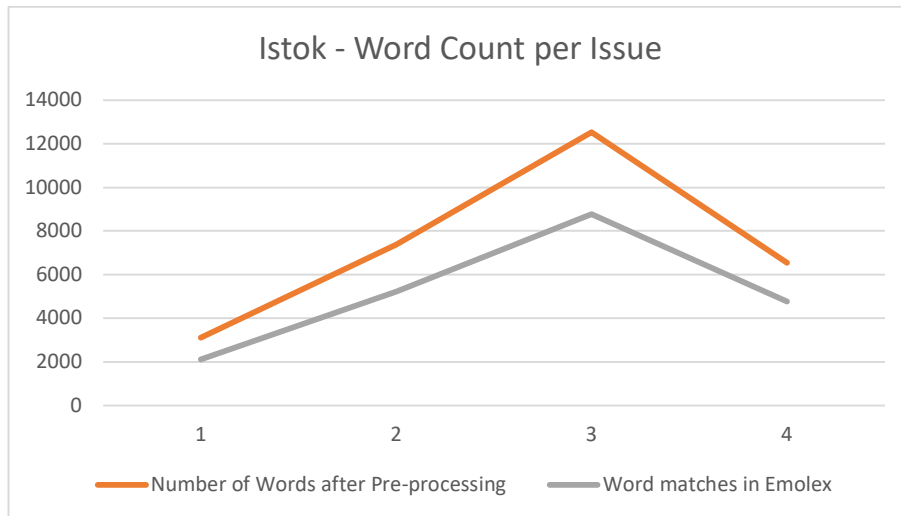
9 Παράρτημα Β – Πλήθος λέξεων και διαγράμματα κάλυψης



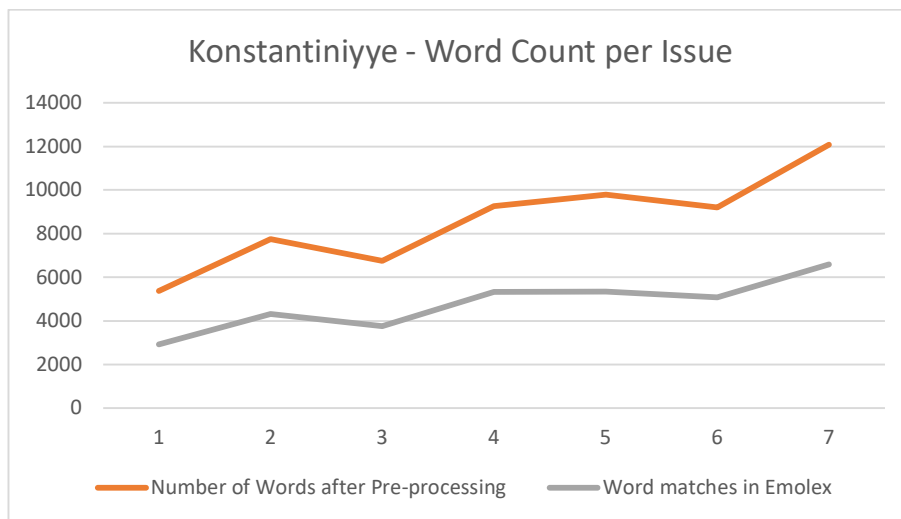
ΕΙΚΟΝΑ 27 ΠΛΗΘΟΣ ΛΕΞΕΩΝ ΜΕΤΑ ΤΗΝ ΠΡΟΠΕΞΕΡΓΑΣΙΑ ΚΑΙ ΠΛΗΘΟΣ ΛΕΞΕΩΝ ΠΟΥ ΕΝΥΠΑΡΧΟΥΝ ΣΤΟ ΕΜΟΛΕΧ ΑΝΑ ΕΚΔΟΣΗ ΤΟΥ DABIQ



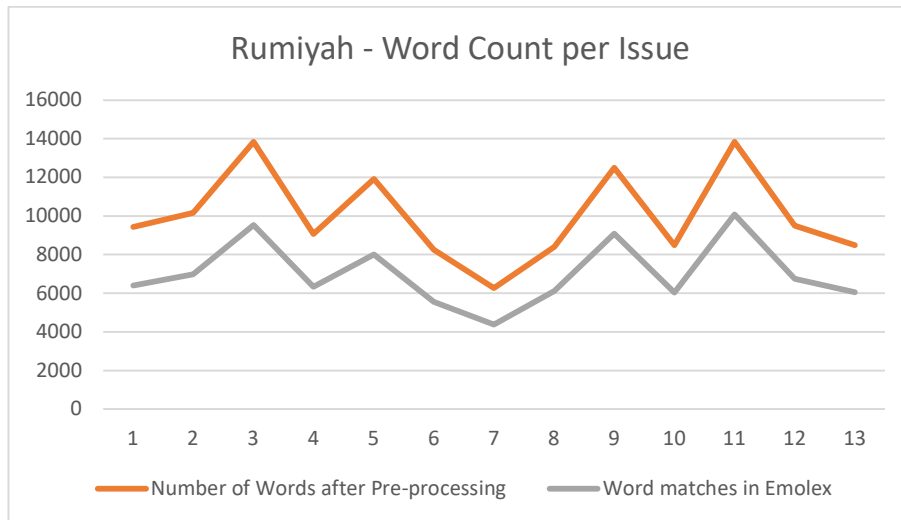
ΕΙΚΟΝΑ 28 ΠΛΗΘΟΣ ΛΕΞΕΩΝ ΜΕΤΑ ΤΗΝ ΠΡΟΠΕΞΕΡΓΑΣΙΑ ΚΑΙ ΠΛΗΘΟΣ ΛΕΞΕΩΝ ΠΟΥ ΕΝΥΠΑΡΧΟΥΝ ΣΤΟ ΕΜΟΛΕΧ ΑΝΑ ΕΚΔΟΣΗ ΤΟΥ DARAL-ISLAM



ΕΙΚΟΝΑ 29 ΠΛΗΘΟΣ ΛΕΞΕΩΝ ΜΕΤΑ ΤΗΝ ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΚΑΙ ΠΛΗΘΟΣ ΛΕΞΕΩΝ ΠΟΥ ΕΝΥΠΑΡΧΟΥΝ ΣΤΟ ΕΜΟΛΕΧ ΑΝΑ ΕΚΔΟΣΗ ΤΟΥ ΙΣΤΟΚ



ΕΙΚΟΝΑ 30 ΠΛΗΘΟΣ ΛΕΞΕΩΝ ΜΕΤΑ ΤΗΝ ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΚΑΙ ΠΛΗΘΟΣ ΛΕΞΕΩΝ ΠΟΥ ΕΝΥΠΑΡΧΟΥΝ ΣΤΟ ΕΜΟΛΕΧ ΑΝΑ ΕΚΔΟΣΗ ΤΟΥ ΚΟΝΣΤΑΝΤΙΝΙΥΓΕ

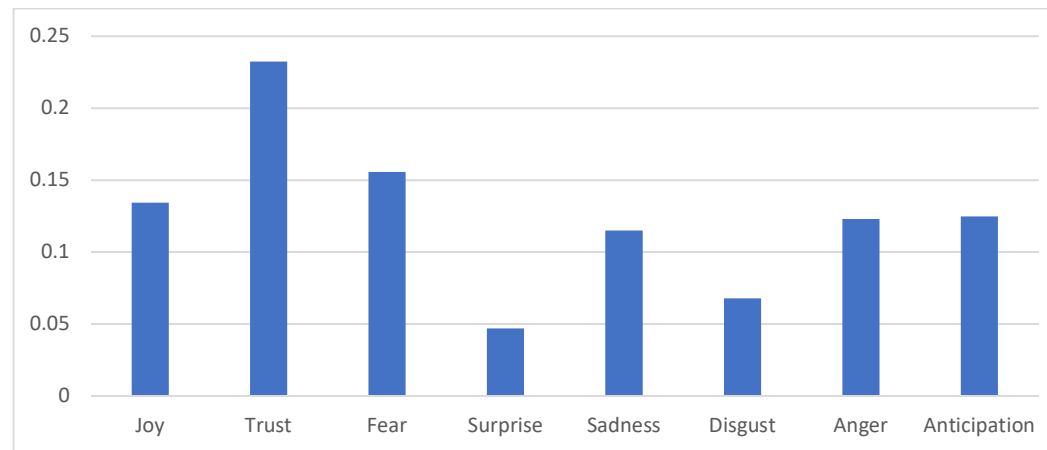


ΕΙΚΟΝΑ 31 ΠΛΗΘΟΣ ΛΕΞΕΩΝ ΜΕΤΑ ΤΗΝ ΠΡΟΠΕΞΕΡΓΑΣΙΑ ΚΑΙ ΠΛΗΘΟΣ ΛΕΞΕΩΝ ΠΟΥ ΕΝΥΠΑΡΧΟΥΝ ΣΤΟ ΕΜΟΛΕΧ ΑΝΑ ΕΚΔΟΣΗ ΤΟΥ RUMIYAH

10 Παράρτημα Γ – Μέση εκτίμηση συναισθήματος και πολικότητας ανά περιοδικό (Overall)

ΠΙΝΑΚΑΣ 23 ΜΕΣΕΣ ΤΙΜΕΣ ΕΚΤΙΜΗΣΗΣ ΣΥΝΑΙΣΘΗΜΑΤΟΣ ΚΑΙ ΠΟΛΙΚΟΤΗΤΑΣ ΑΝΑ ΠΕΡΙΟΔΙΚΟ

Περιοδικό	Emotion								Polarity	
	Joy	Trust	Fear	Surprise	Sadness	Disgust	Anger	Anticipation	Positive	Negative
Dabiq	0.116962	0.223052	0.16653	0.04703	0.114484	0.07168	0.135932	0.124329	0.516079	0.483921
Daral-Islam	0.143684	0.254213	0.145704	0.04663	0.110508	0.068437	0.120079	0.110746	0.537315	0.462685
Istok	0.142034	0.251951	0.147567	0.048303	0.105864	0.066865	0.116692	0.120725	0.57037	0.42963
Konstantiniyye	0.146791	0.22158	0.150293	0.045022	0.11646	0.069282	0.103952	0.14662	0.555658	0.444342
Rumiyah	0.122174	0.211433	0.16873	0.047626	0.127519	0.062795	0.138485	0.121237	0.513427	0.486573
Μέση εκτίμηση	0.134329	0.232446	0.155765	0.046922	0.114967	0.067812	0.123028	0.124732	0.53857	0.46143

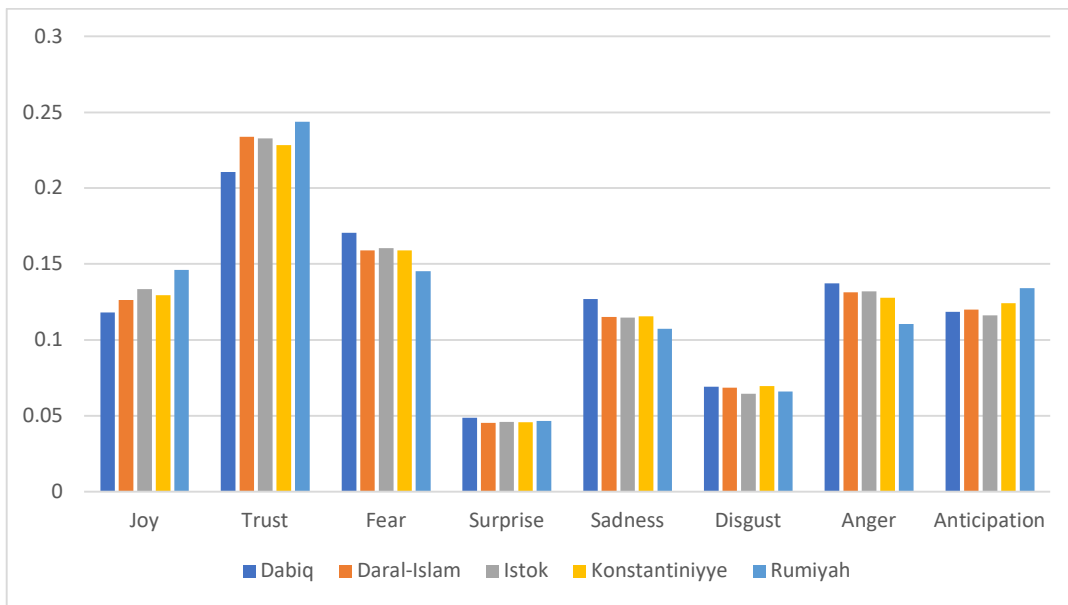


ΕΙΚΟΝΑ 32 ΜΕΣΗ ΕΚΤΙΜΗΣΗ ΣΥΝΑΙΣΘΗΜΑΤΟΣ ΥΠΟΛΟΓΙΣΜΕΝΗ ΣΕ ΟΛΑ ΤΑ ΠΕΡΙΟΔΙΚΑ

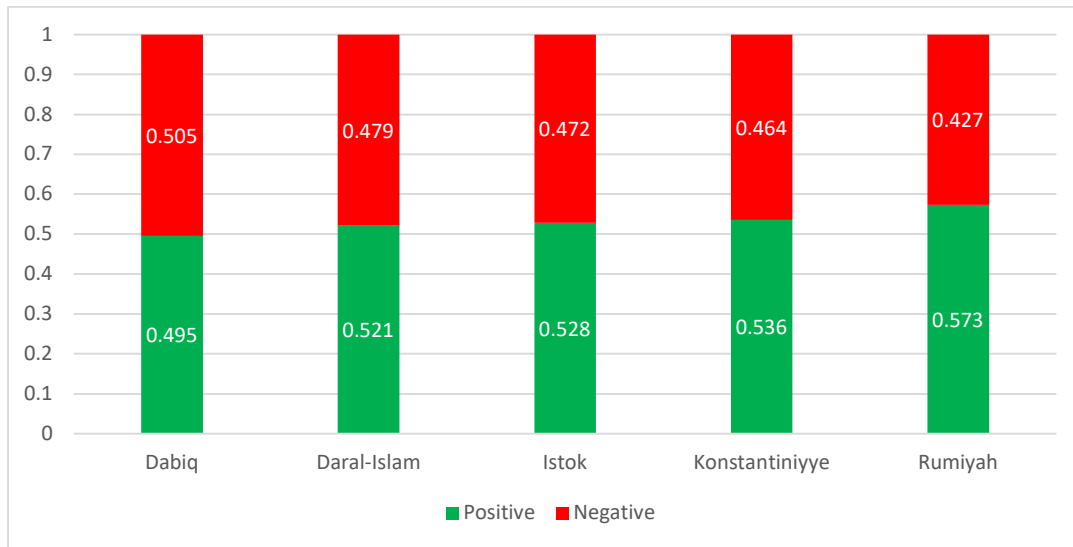


ΕΙΚΟΝΑ 33 ΜΕΣΗ ΠΟΛΙΚΟΤΗΤΑ ΣΤΟ ΣΥΝΟΛΟ ΤΩΝ ΠΕΡΙΟΔΙΚΩΝ

11 Παράρτημα Δ – Μέση εκτίμηση συναίσθηματος και πολικότητας σε επίπεδο περιοδικού.



ΕΙΚΟΝΑ 34 ΜΕΣΗ ΕΚΤΙΜΗΣΗ ΣΥΝΑΙΣΘΗΜΑΤΟΣ ΑΝΑ ΠΕΡΙΟΔΙΚΟ



ΕΙΚΟΝΑ 35 ΜΕΣΗ ΤΙΜΗ ΠΟΛΙΚΟΤΗΤΑΣ ΑΝΑ ΠΕΡΙΟΔΙΚΟ

12 Παράρτημα Ε – Εκτίμηση Συναισθήματος και πολικότητας ανά έκδοση περιοδικού.

Στα κελιά του πίνακα που ακολουθεί εμφανίζεται διαβάθμιση χρώματος του φόντου από το κόκκινο στο πράσινο, καθώς η τιμή του κελιού ολισθαίνει από το ελάχιστο της στήλης στο μέγιστο της.

TABLE 12-1. ΕΚΤΙΜΗΣΗ ΣΥΝΑΙΣΘΗΜΑΤΟΣ ΚΑΙ ΠΟΛΙΚΟΤΗΤΑΣ ΑΝΑ ΕΚΔΟΣΗ ΠΕΡΙΟΔΙΚΟΥ.

Περιοδικό	Joy	Trust	Fear	Surprise	Sadness	Disgust	Anger	Anticipation	Positive	Negative
Dabiq 01	0.1222496 49	0.2311540 7	0.1607119 96	0.0484404 94	0.1145179 66	0.0603286 77	0.1270693 55	0.1355277 93	0.5435450 13	0.4564549 87
Dabiq 02	0.1209063 45	0.2367104 78	0.1516669 67	0.0465954 48	0.1052074 22	0.0736085 02	0.1344082 93	0.1308965 46	0.5225554 96	0.4774445 04
Dabiq 03	0.0957348 12	0.1923578 29	0.1853298 02	0.0473750 14	0.1426187 32	0.0844377 5	0.1457290 44	0.1064170 17	0.4506384 65	0.5493615 35
Dabiq 04	0.1248271 28	0.2090797 57	0.1683973 91	0.0482608 88	0.1119396 15	0.0674574 51	0.1366291 47	0.1334086 22	0.4951741 06	0.5048258 94
Dabiq 05	0.1348468 17	0.2160561 55	0.1549226 9	0.0462099 4	0.1072180 98	0.0731410 01	0.1290790 25	0.1385262 74	0.5308482 45	0.4691517 55
Dabiq 06	0.1129242 03	0.231803	0.1601379 82	0.0481940 02	0.1092300 99	0.0655323 34	0.1309856 92	0.1411926 88	0.5190312 39	0.4809687 61
Dabiq 07	0.1186107 17	0.2157275 9	0.1671375 33	0.0473981 23	0.1100562 14	0.0779131 91	0.1373986 68	0.1257579 64	0.4944921 74	0.5055078 26
Dabiq 08	0.1150432 31	0.2246415 8	0.1667500 8	0.0437321 17	0.1176854 51	0.0846951 15	0.1315308 31	0.1159215 94	0.5009989 94	0.4990010 06
Dabiq 09	0.1197453 23	0.2127988 15	0.1769339 64	0.0480655 52	0.1051068 48	0.0677999 79	0.1435909 25	0.1259585 94	0.5171013 21	0.4828986 79

Περιοδικό	Joy	Trust	Fear	Surprise	Sadness	Disgust	Anger	Anticipation	Positive	Negative
Dabiq 10	0.12008116	0.235381968	0.163099362	0.044957751	0.106875341	0.067521961	0.143281785	0.118800672	0.53036665	0.46963335
Dabiq 11	0.116129662	0.23556087	0.17163919	0.043319219	0.106705626	0.070496443	0.13302905	0.123119939	0.545327623	0.454672377
Dabiq 12	0.109704763	0.224008282	0.170702669	0.047886579	0.112706364	0.0720144	0.147021248	0.115955695	0.524494717	0.475505283
Dabiq 13	0.104872113	0.20365516	0.181017718	0.048525422	0.136278225	0.071089113	0.142730816	0.111831434	0.478368228	0.521631772
Dabiq 14	0.108579566	0.237461563	0.169616651	0.047213266	0.11794833	0.069712025	0.139329755	0.110138844	0.53702467	0.46297533
Dabiq 15	0.130181937	0.239375746	0.149892894	0.049271887	0.1131725	0.069448518	0.117169389	0.131487129	0.551218088	0.448781912
Daral-Islam 01	0.181363228	0.305921168	0.107451268	0.044144677	0.090524147	0.063702811	0.08709652	0.119796181	0.637946023	0.362053977
Daral-Islam 02	0.162905312	0.245326037	0.14468579	0.052651271	0.092480208	0.074691801	0.123174558	0.104085023	0.567111105	0.432888895
Daral-Islam 03	0.1330464	0.230086554	0.163227922	0.039674539	0.140489037	0.078560283	0.123085156	0.091830108	0.515005831	0.484994169
Daral-Islam 04	0.149452724	0.260763644	0.148246988	0.04195872	0.106100731	0.053307808	0.11855353	0.121615857	0.519175378	0.480824622
Daral-Islam 05	0.146161716	0.258897042	0.145448714	0.038332473	0.108436898	0.066211632	0.128658028	0.107853498	0.529250996	0.470749004

Περιοδικό	Joy	Trust	Fear	Surprise	Sadness	Disgust	Anger	Anticipation	Positive	Negative
Daral-Islam 06	0.135540372	0.242416196	0.129774385	0.059512738	0.147731608	0.070207304	0.097989773	0.116827623	0.511654174	0.488345826
Daral-Islam 07	0.136943733	0.246923457	0.151535354	0.0473382	0.088820728	0.077396529	0.131513241	0.119528758	0.533544762	0.466455238
Daral-Islam 08	0.129245439	0.240653393	0.161345017	0.043131035	0.116637671	0.06746072	0.133084807	0.108441917	0.514036578	0.485963422
Daral-Islam 09	0.134492533	0.239682793	0.156773509	0.051565788	0.100788232	0.070580501	0.131939858	0.114176786	0.520079677	0.479920323
Daral-Islam 10	0.127684513	0.271454875	0.148552772	0.047987116	0.113070988	0.062248845	0.125693525	0.103307365	0.525343831	0.474656169
Istok 1	0.14912688	0.313343802	0.113289878	0.047568495	0.094866418	0.062582472	0.097016587	0.122205469	0.630745486	0.369254514
Istok 2	0.128065537	0.239751743	0.168904813	0.045179683	0.112178003	0.070284423	0.127876459	0.10775934	0.533972511	0.466027489
Istok 3	0.129465658	0.218012858	0.166337272	0.054814586	0.119533732	0.066901103	0.126729078	0.118205713	0.52748897	0.47251103
Istok 4	0.161477672	0.236696644	0.141735106	0.045648914	0.096876566	0.067690151	0.115144808	0.13473014	0.589273073	0.410726927
Konstantiniy ye 1	0.147749608	0.213236373	0.144511133	0.053401099	0.127585603	0.066333581	0.105232298	0.141950306	0.551181142	0.448818858
Konstantiniy ye 2	0.143953323	0.214992925	0.157827297	0.043074295	0.118855239	0.070113002	0.10830202	0.142881899	0.538259496	0.461740504

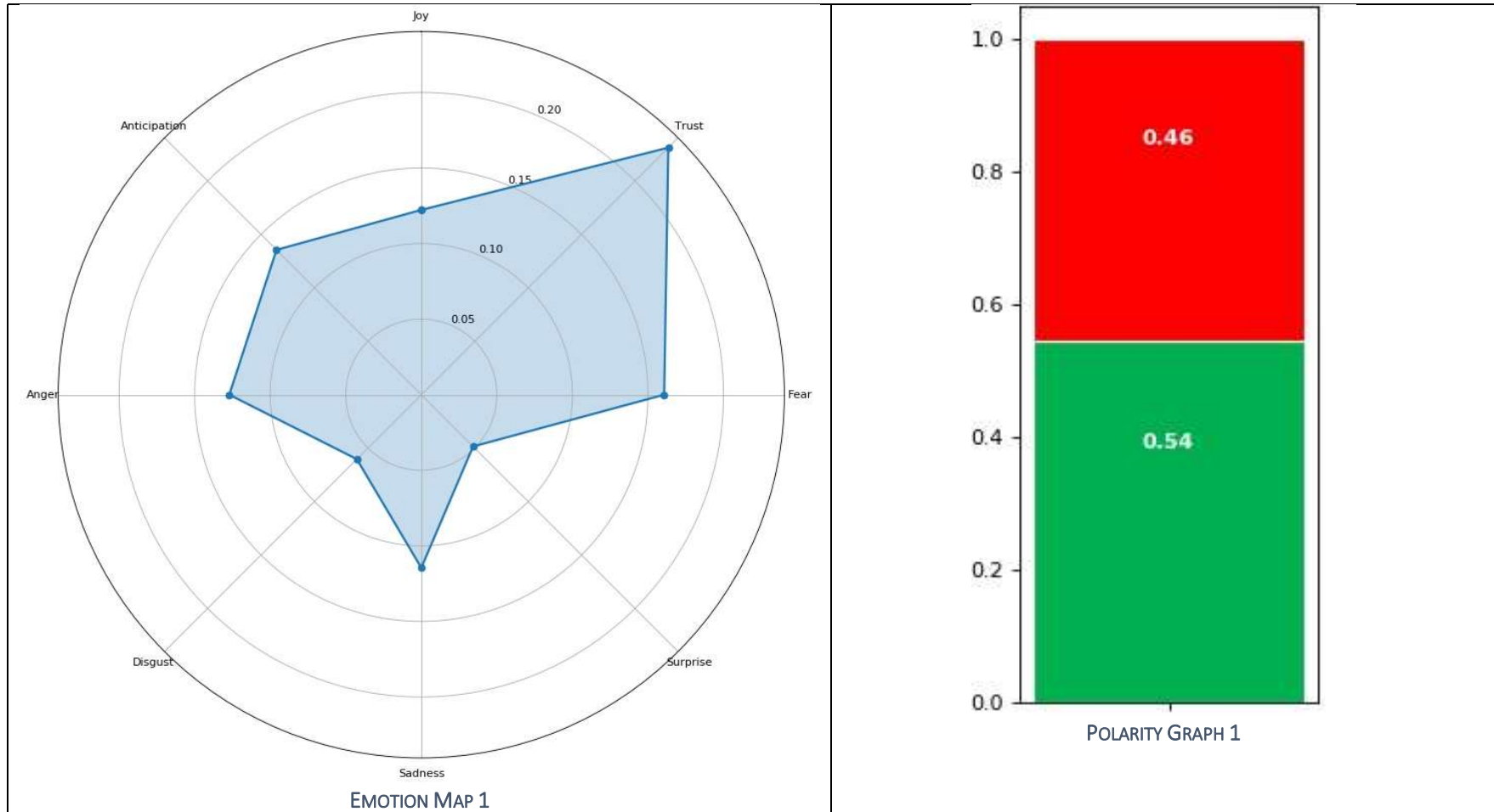
Περιοδικό	Joy	Trust	Fear	Surprise	Sadness	Disgust	Anger	Anticipation	Positive	Negative
Konstantiniy ye 3	0.138604166	0.214148687	0.149901049	0.039693059	0.12658072	0.081683546	0.106719914	0.142668859	0.519986593	0.480013407
Konstantiniy ye 4	0.16889095	0.214324332	0.144679647	0.04700665	0.106458855	0.070292765	0.099936921	0.148409879	0.559186155	0.440813845
Konstantiniy ye 5	0.156507728	0.230024804	0.147607563	0.039590708	0.112080511	0.065073771	0.095198212	0.153916702	0.579336139	0.420663861
Konstantiniy ye 6	0.13634312	0.235482588	0.150021414	0.041466881	0.107333498	0.069359848	0.113333496	0.146659155	0.575605579	0.424394421
Konstantiniy ye 7	0.135491065	0.228851276	0.157501089	0.050919164	0.11632548	0.062114269	0.098941935	0.149855723	0.566050959	0.433949041
Rumiyah 01	0.117333187	0.217108258	0.163145144	0.043159253	0.131987467	0.061451897	0.15259157	0.113223224	0.525705278	0.474294722
Rumiyah 02	0.138339976	0.218636062	0.166321975	0.045919452	0.123711647	0.057480521	0.129109463	0.120480904	0.526862668	0.473137332
Rumiyah 03	0.120854669	0.211670056	0.173031506	0.045909118	0.119967758	0.065587952	0.13140723	0.13157171	0.505588161	0.494411839
Rumiyah 04	0.130586775	0.204214687	0.17383209	0.042194309	0.136506974	0.065661169	0.135963963	0.111040032	0.484259158	0.515740842
Rumiyah 05	0.112180065	0.188093355	0.188272327	0.050120916	0.13284494	0.064141643	0.145898014	0.118448739	0.4921067	0.5078933
Rumiyah 06	0.114999734	0.201729388	0.179210595	0.055326302	0.120695063	0.057387211	0.148167268	0.122484439	0.498975831	0.501024169

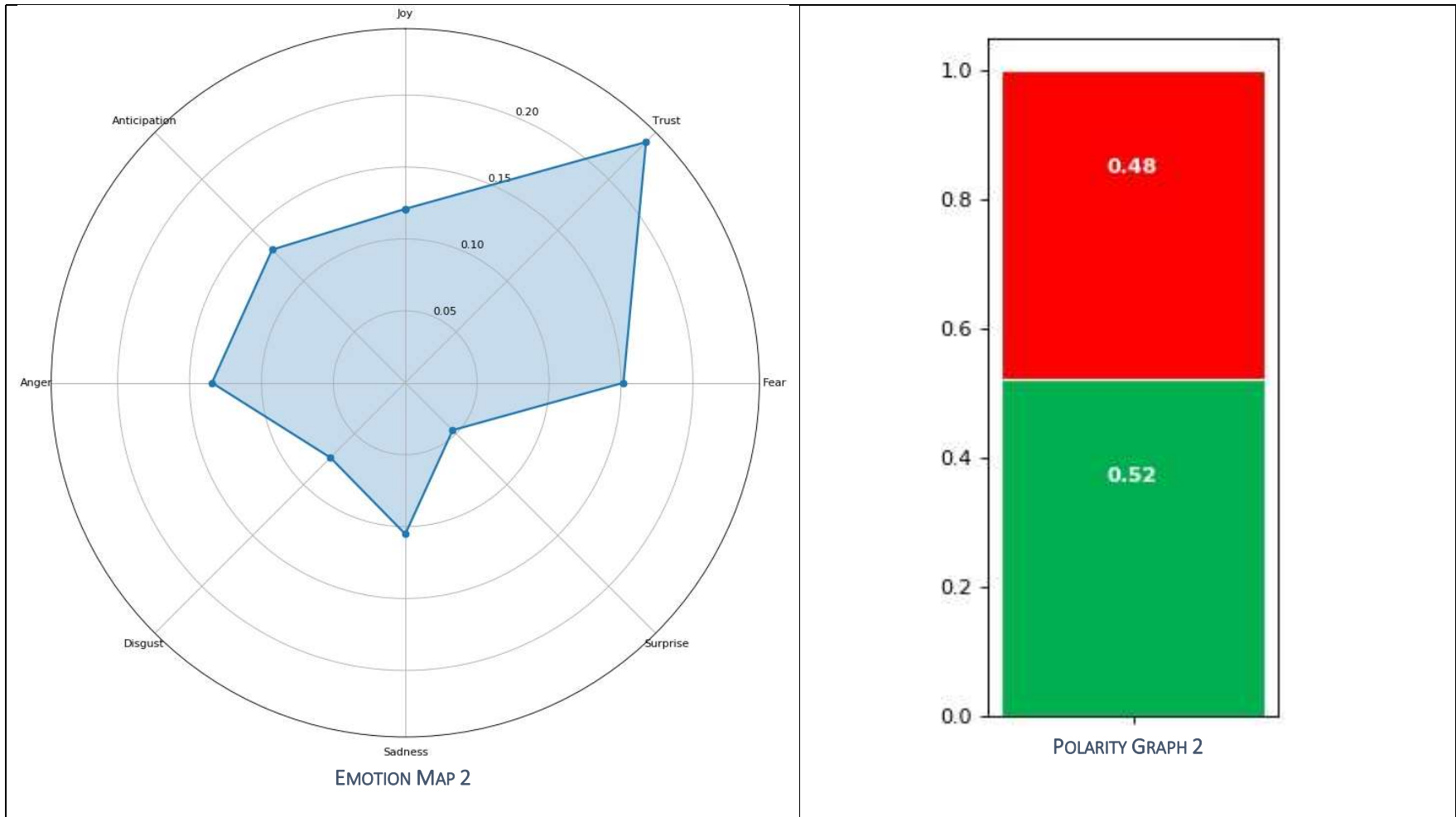
Περιοδικό	Joy	Trust	Fear	Surprise	Sadness	Disgust	Anger	Anticipation	Positive	Negative
Rumiyah 07	0.114098021	0.215511677	0.166748271	0.048572535	0.12872609	0.072908739	0.134176765	0.119257902	0.508354572	0.491645428
Rumiyah 08	0.126913598	0.223389596	0.150395575	0.0426597	0.134991241	0.068975086	0.122244883	0.130430323	0.538812866	0.461187134
Rumiyah 09	0.125367318	0.219214401	0.167794382	0.044595189	0.130145531	0.06302301	0.134507488	0.115352682	0.5044131	0.4955869
Rumiyah 10	0.111539715	0.207776232	0.17696102	0.052445783	0.122599342	0.061586038	0.151039119	0.116052751	0.50821781	0.49178219
Rumiyah 11	0.131469872	0.23905092	0.153947895	0.043221137	0.117338299	0.05653425	0.125466116	0.132971511	0.558780841	0.441219159
Rumiyah 12	0.118404004	0.181275055	0.173604986	0.055508344	0.130167386	0.0635688	0.153426814	0.12404461	0.485126805	0.514873195
Rumiyah 13	0.126171218	0.220963903	0.16022581	0.049511915	0.128071549	0.058030017	0.136303789	0.120721799	0.537347934	0.462652066
Ελάχιστο	0.0957	0.1813	0.1075	0.0383	0.0888	0.0533	0.0871	0.0918	0.4506	0.3621
Μέγιστο	0.1814	0.3133	0.1883	0.0595	0.1477	0.0847	0.1534	0.1539	0.6379	0.5494
Μέση τιμή	0.1301	0.2285	0.1590	0.0469	0.1167	0.0679	0.1272	0.1236	0.5298	0.4702
Τυπική απόκλιση	0.0171	0.0244	0.0158	0.0045	0.0136	0.0068	0.0162	0.0135	0.0345	0.0345

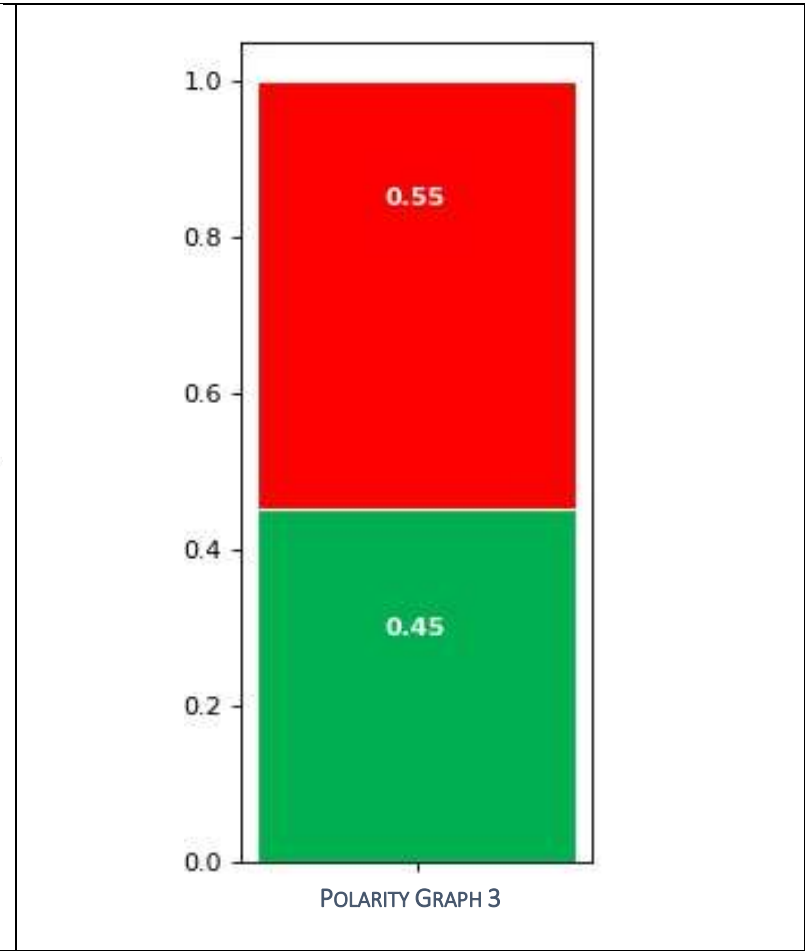
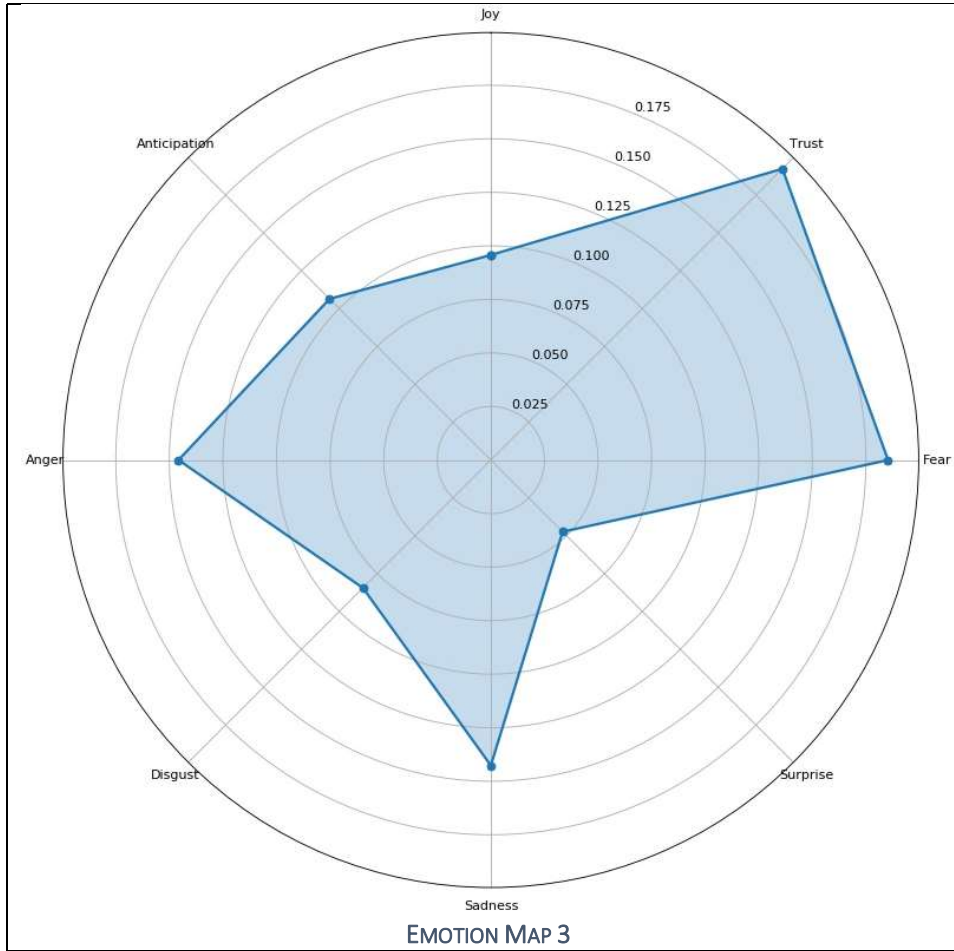
13 Παράρτημα ΣΤ – Διαγράμματα συναισθήματος και πολικότητας ανά έκδοση περιοδικού.

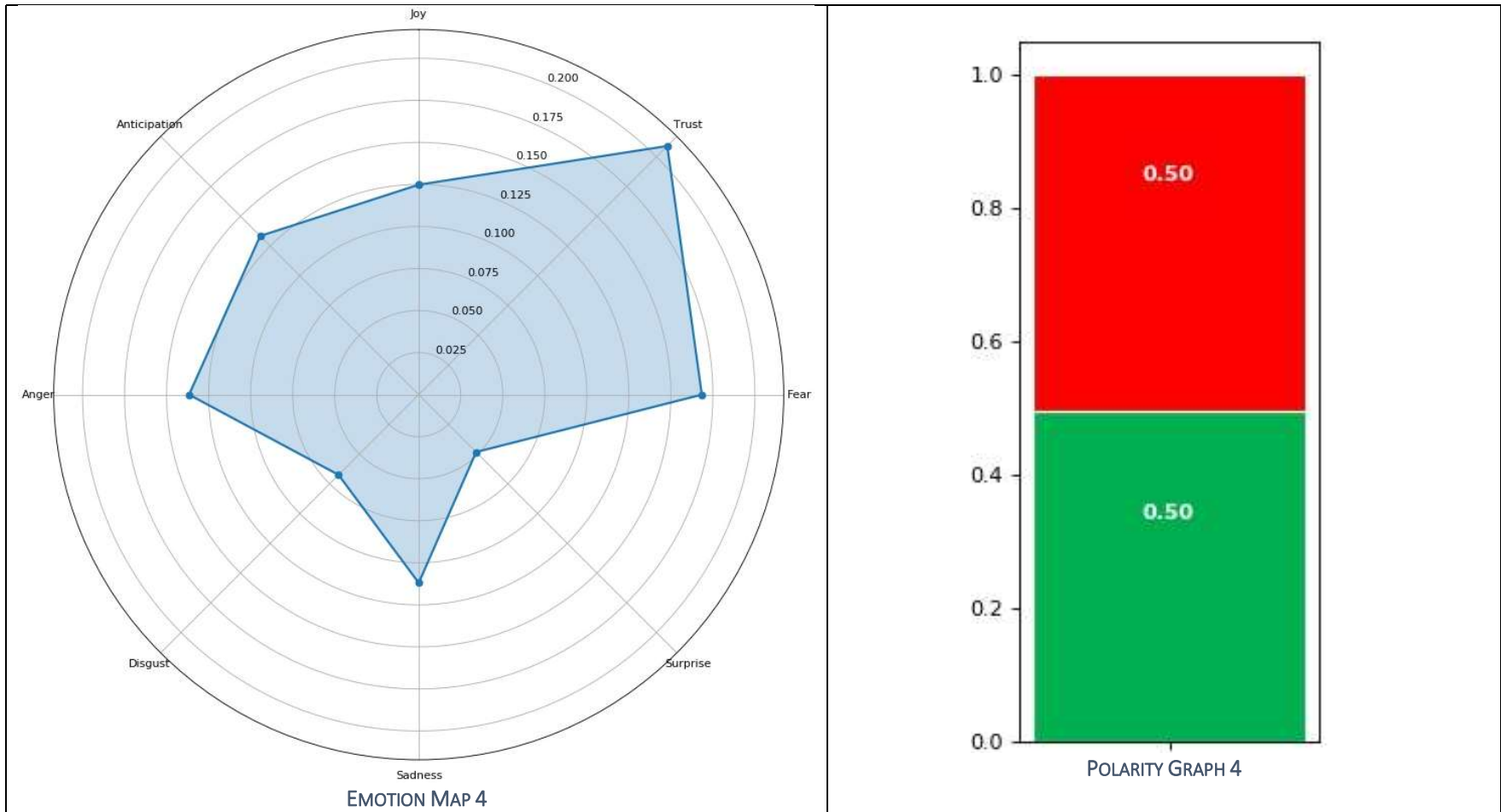
13.1 Dabiq

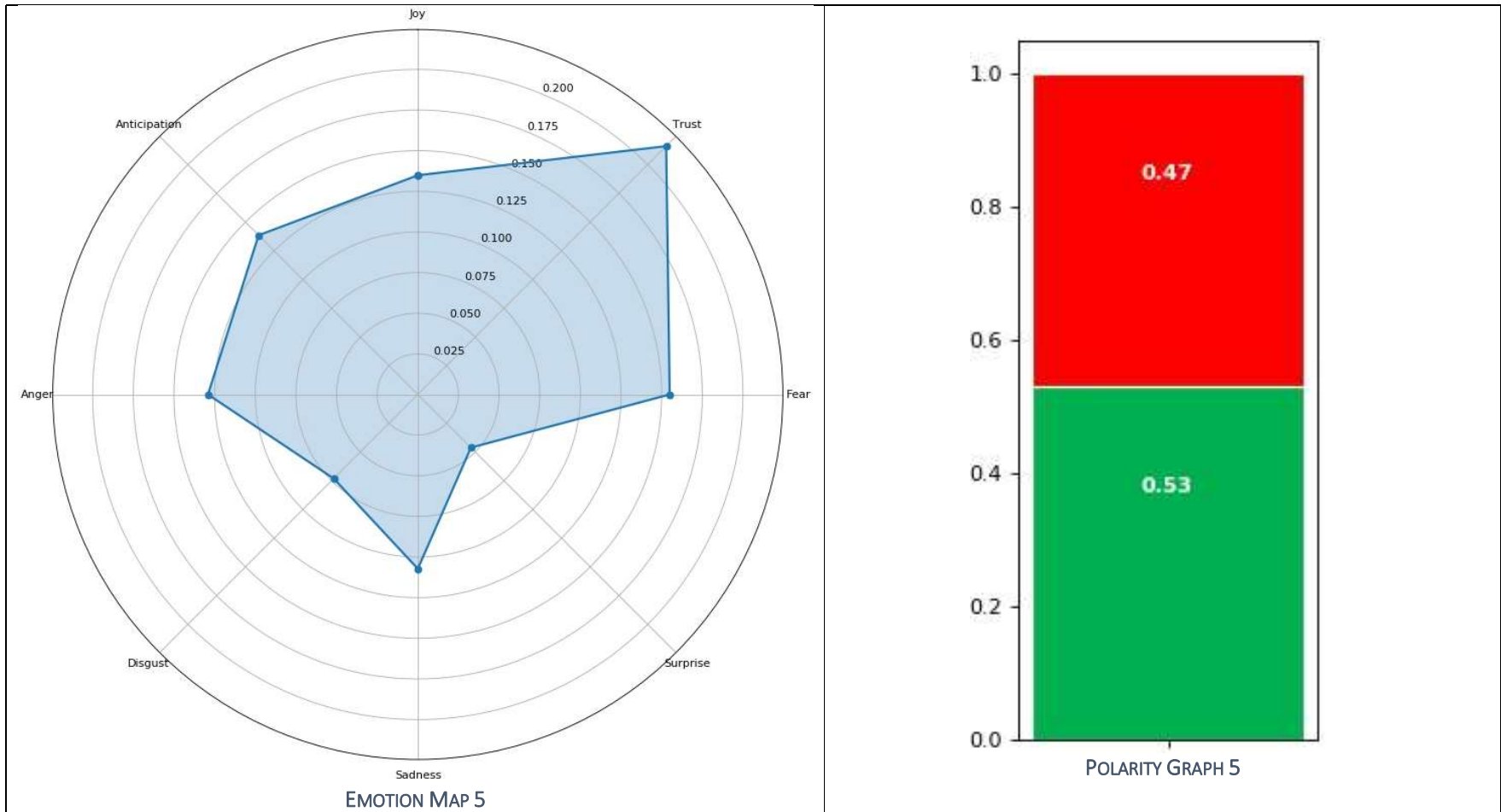
Issue 1

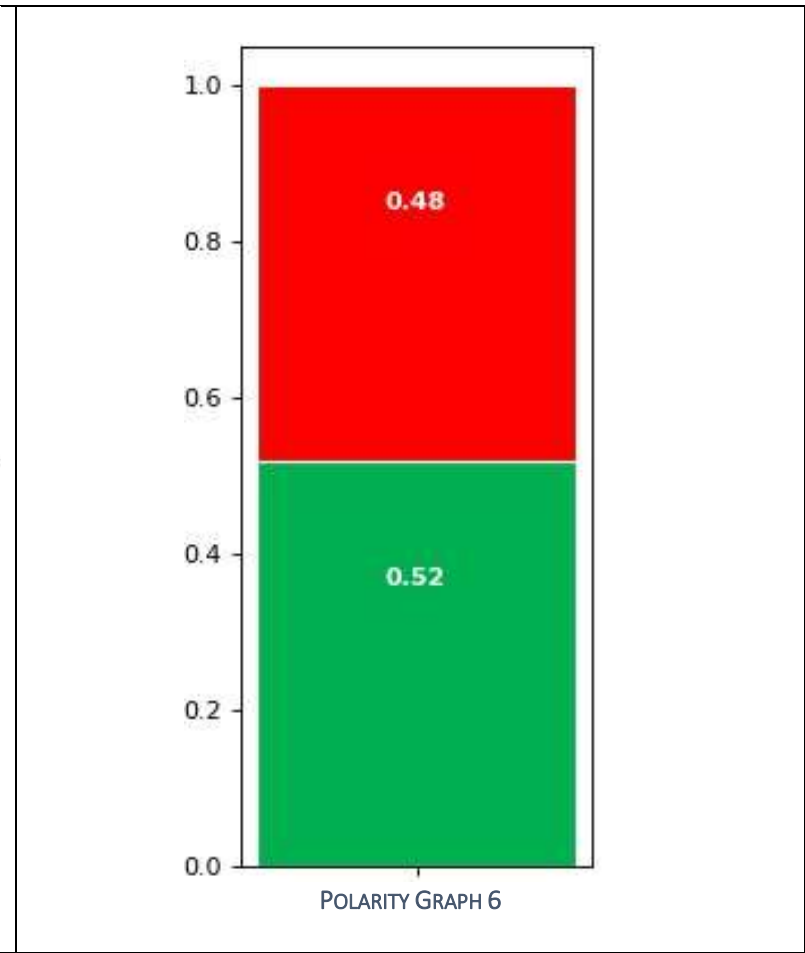
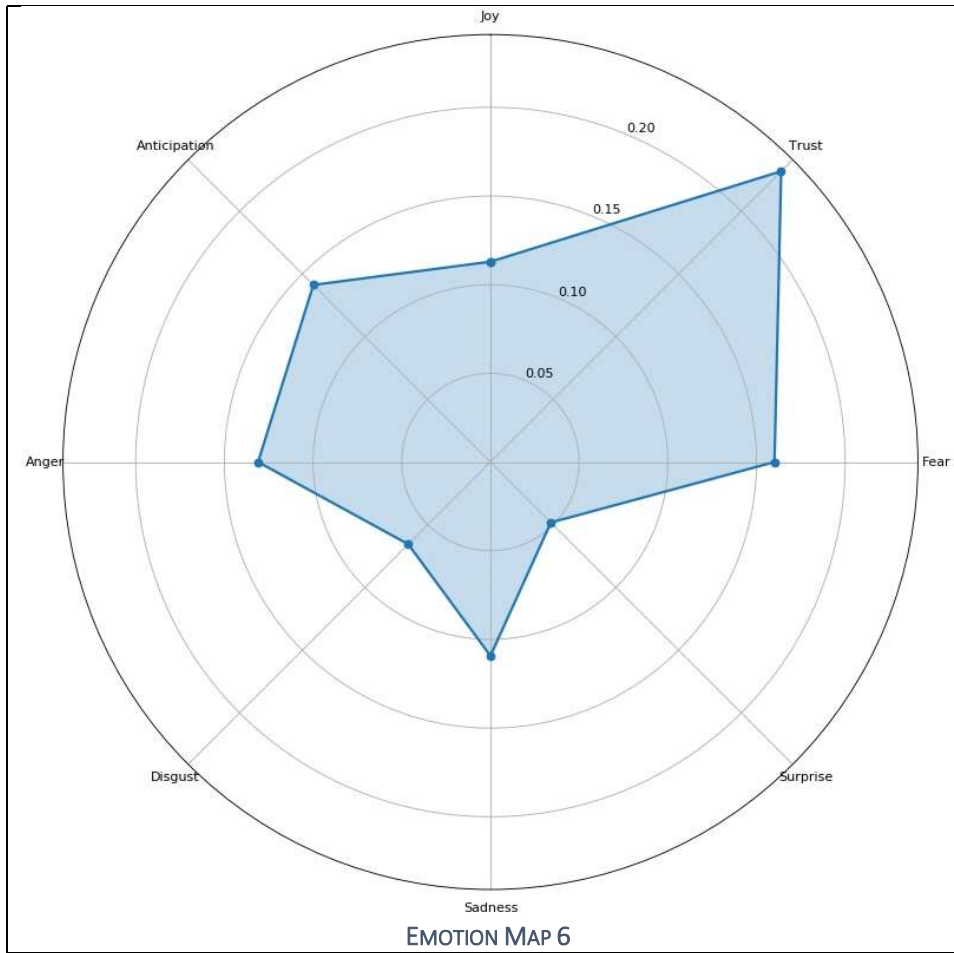


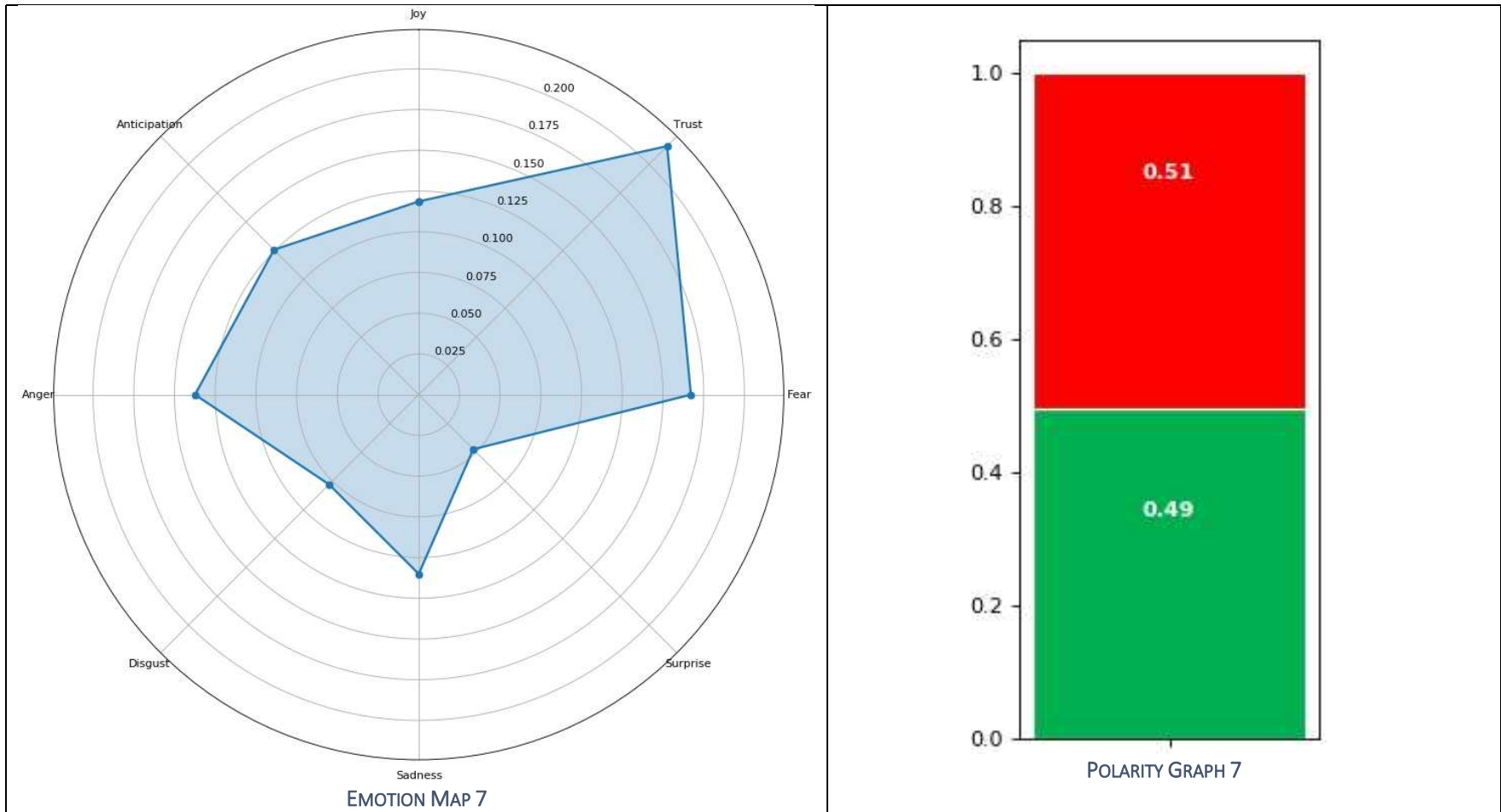


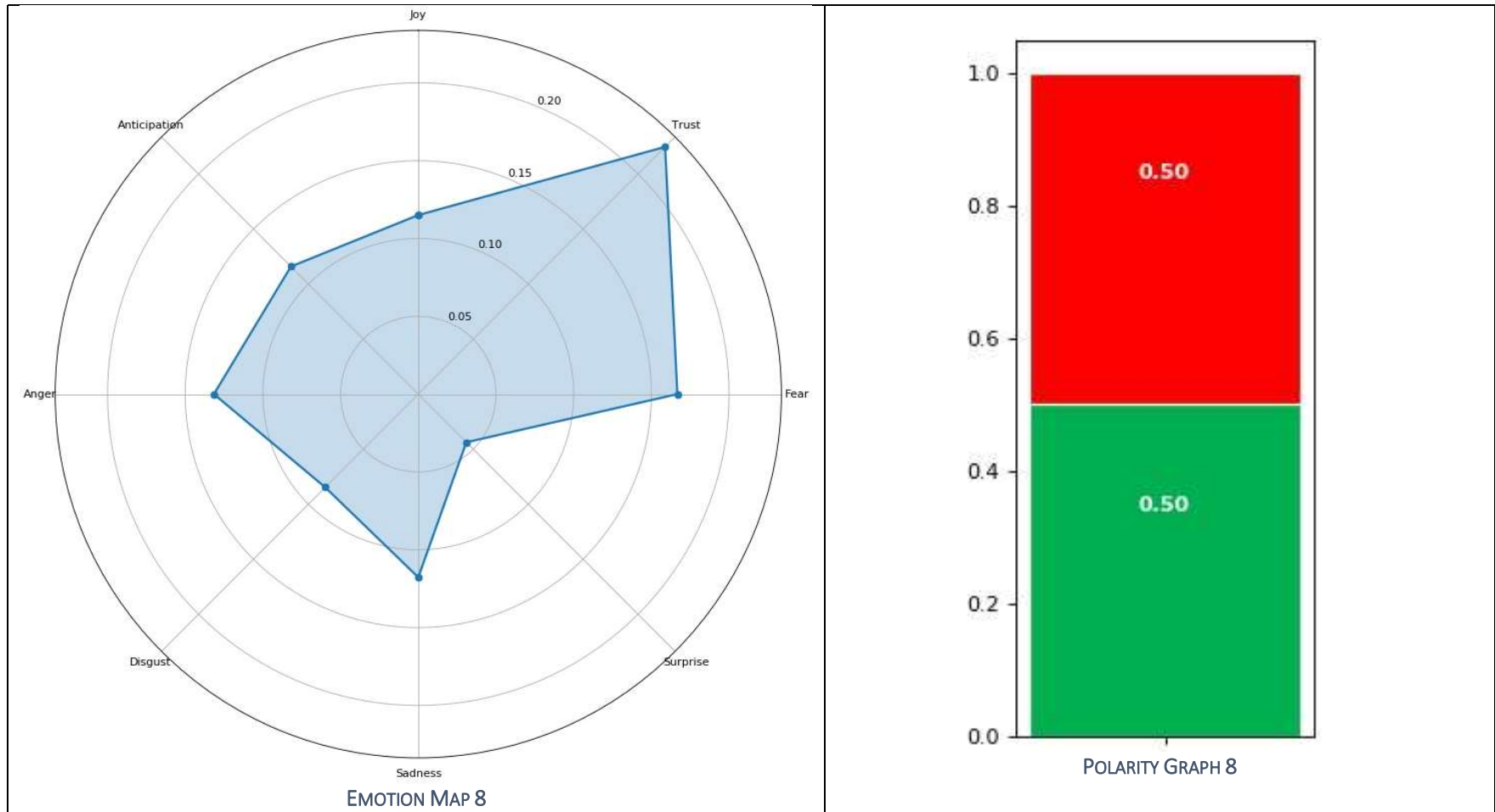


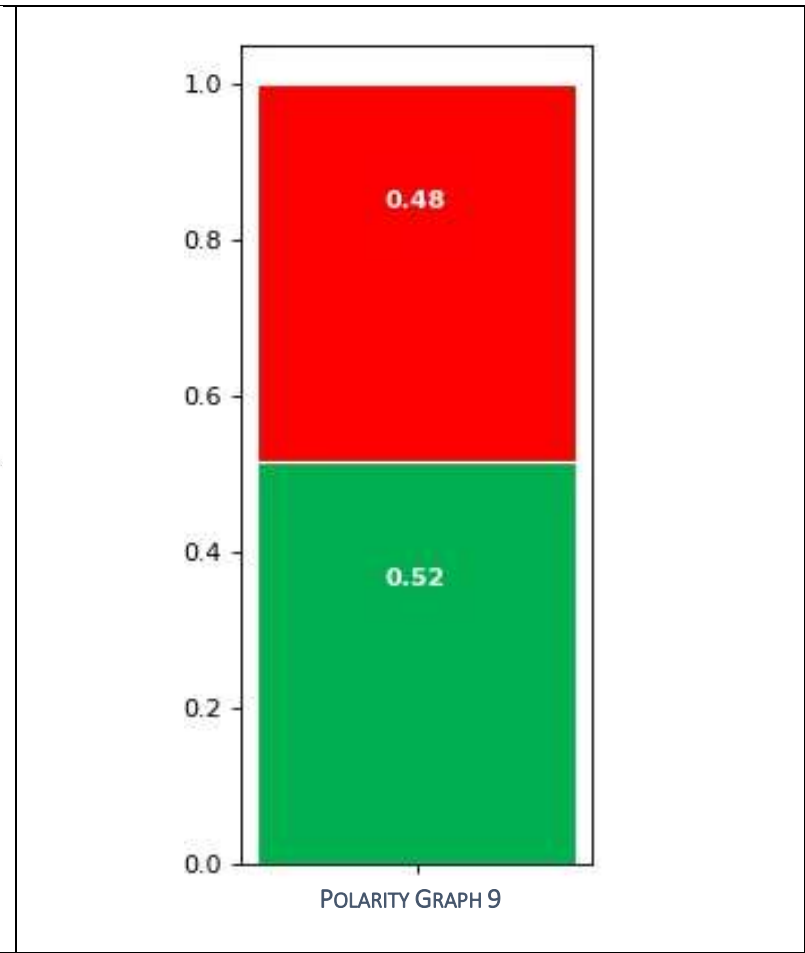
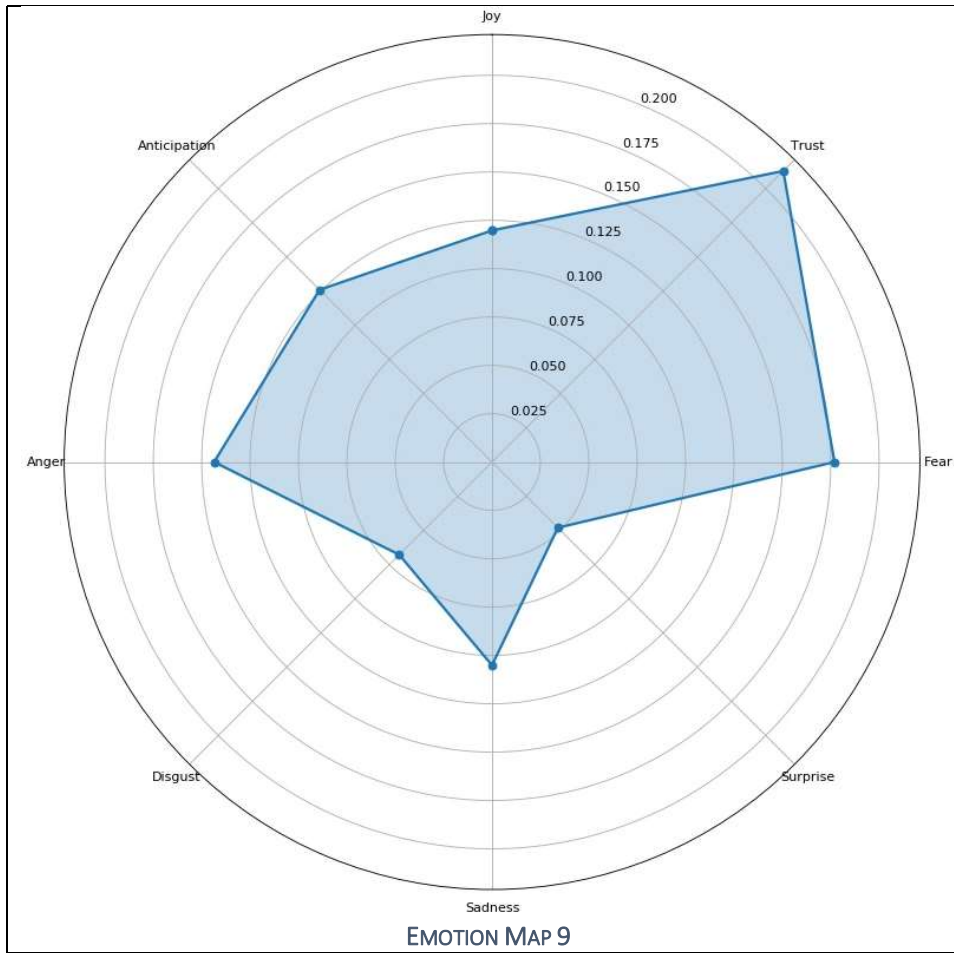




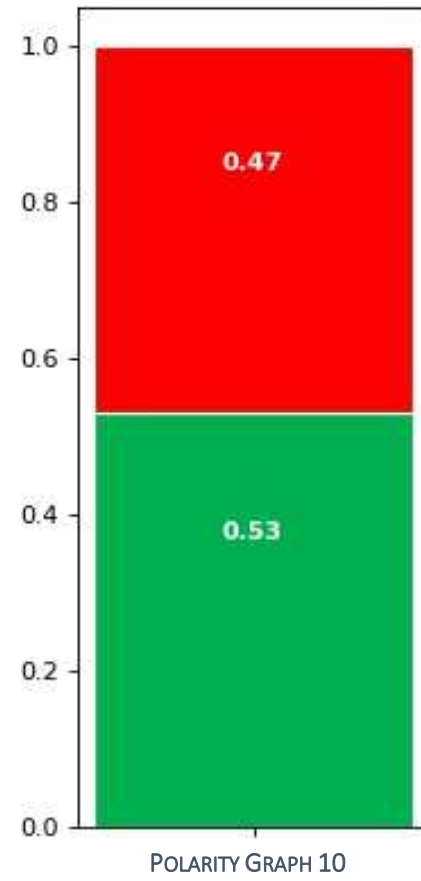
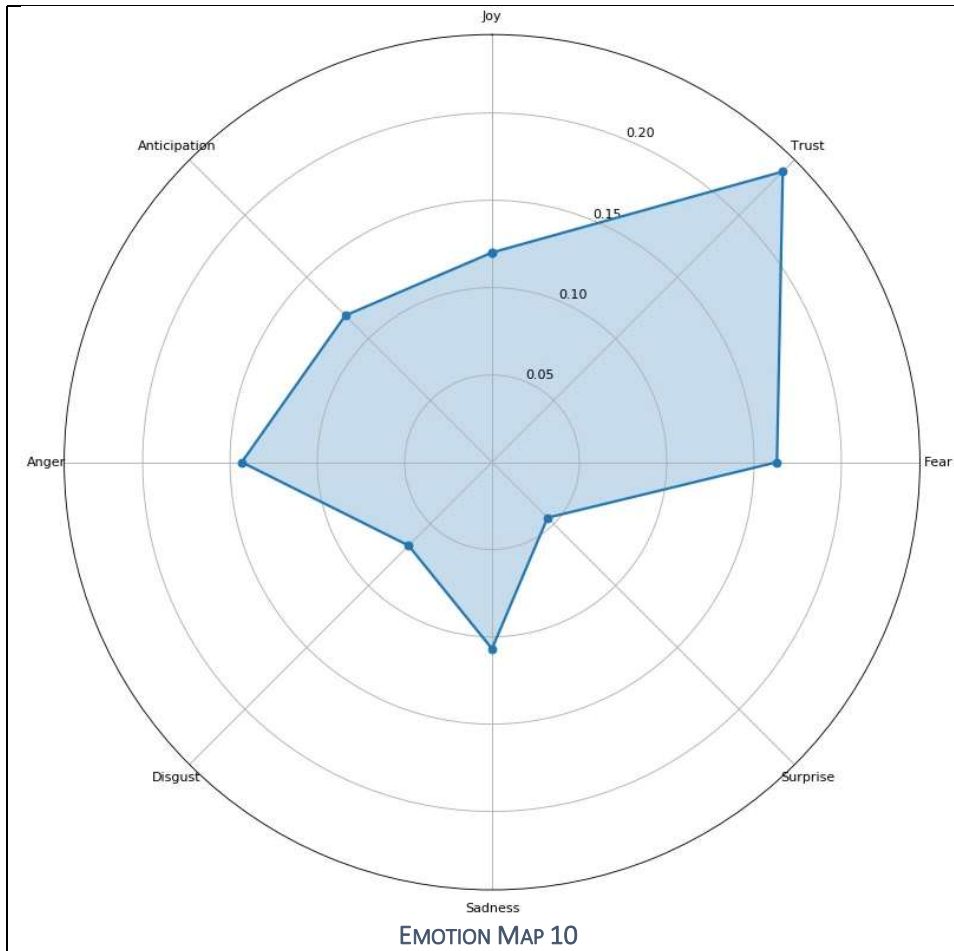




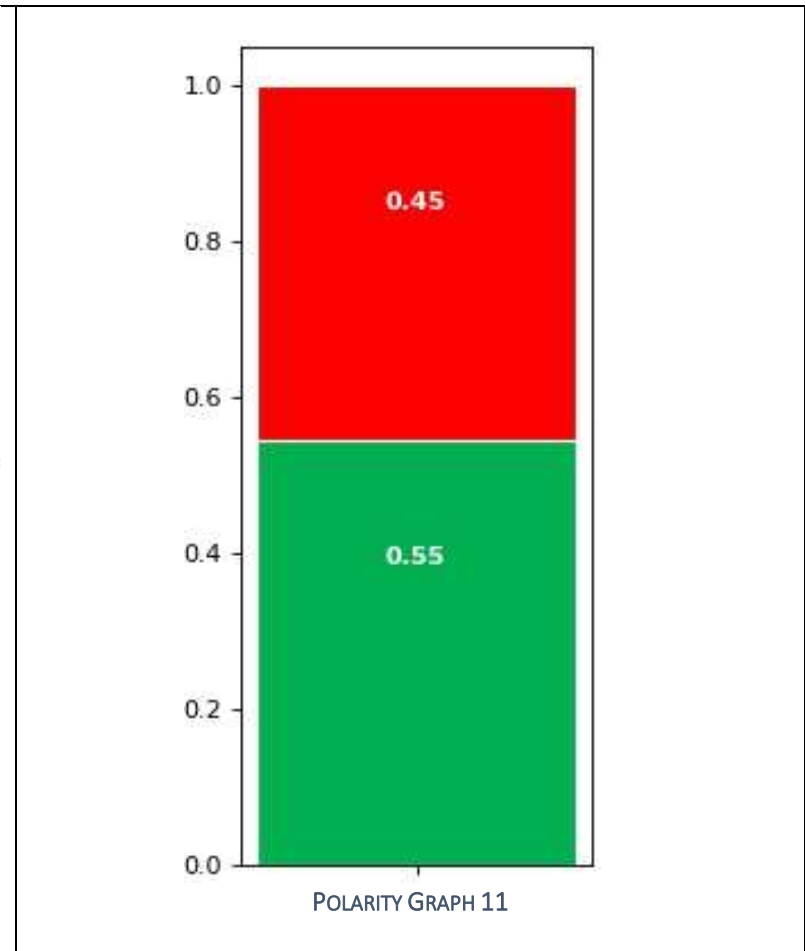
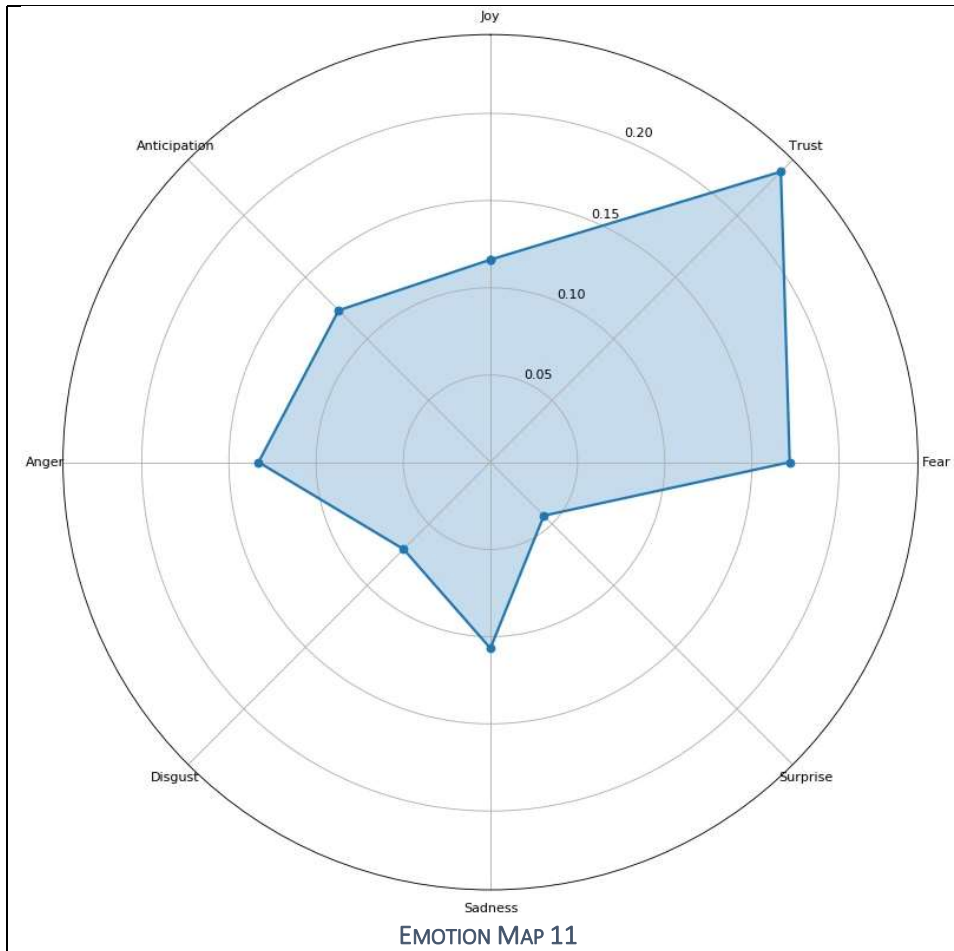


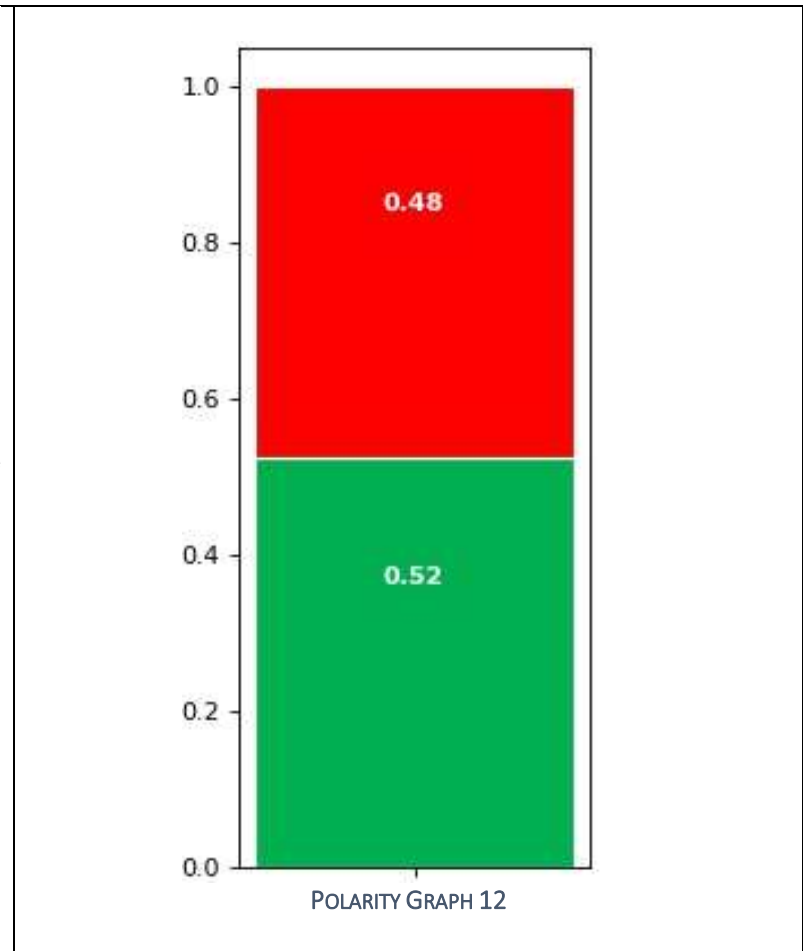
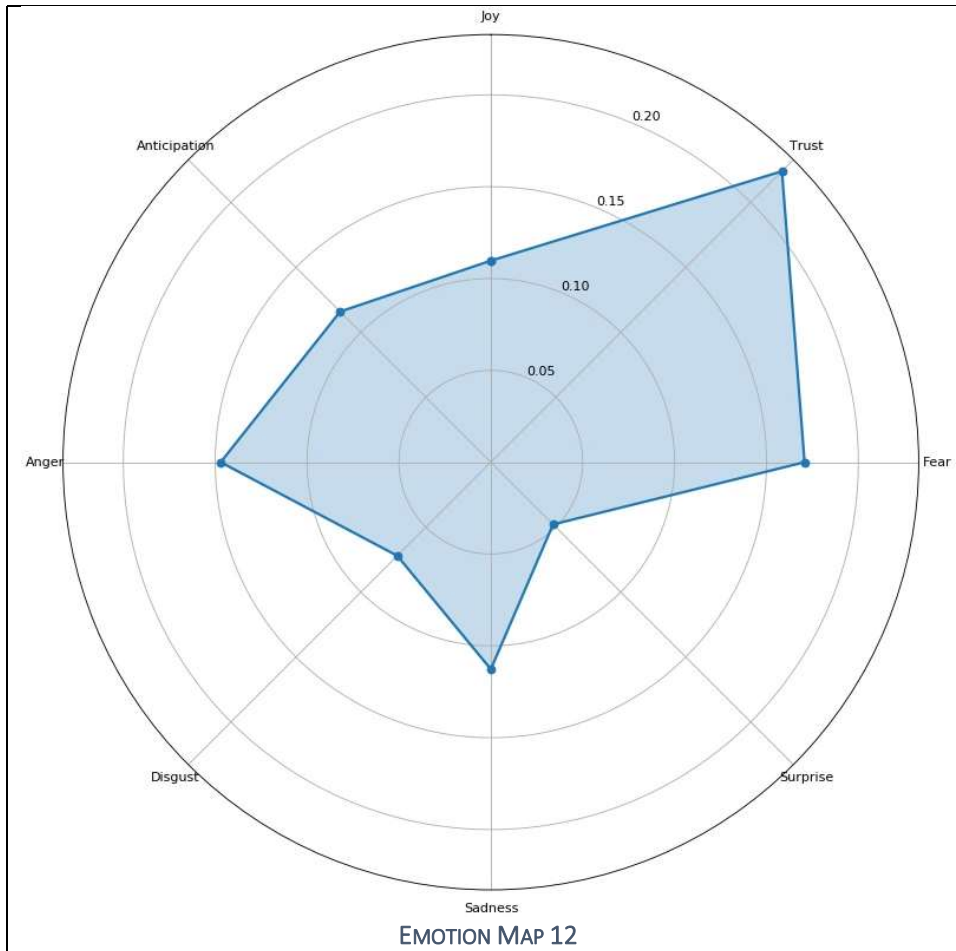


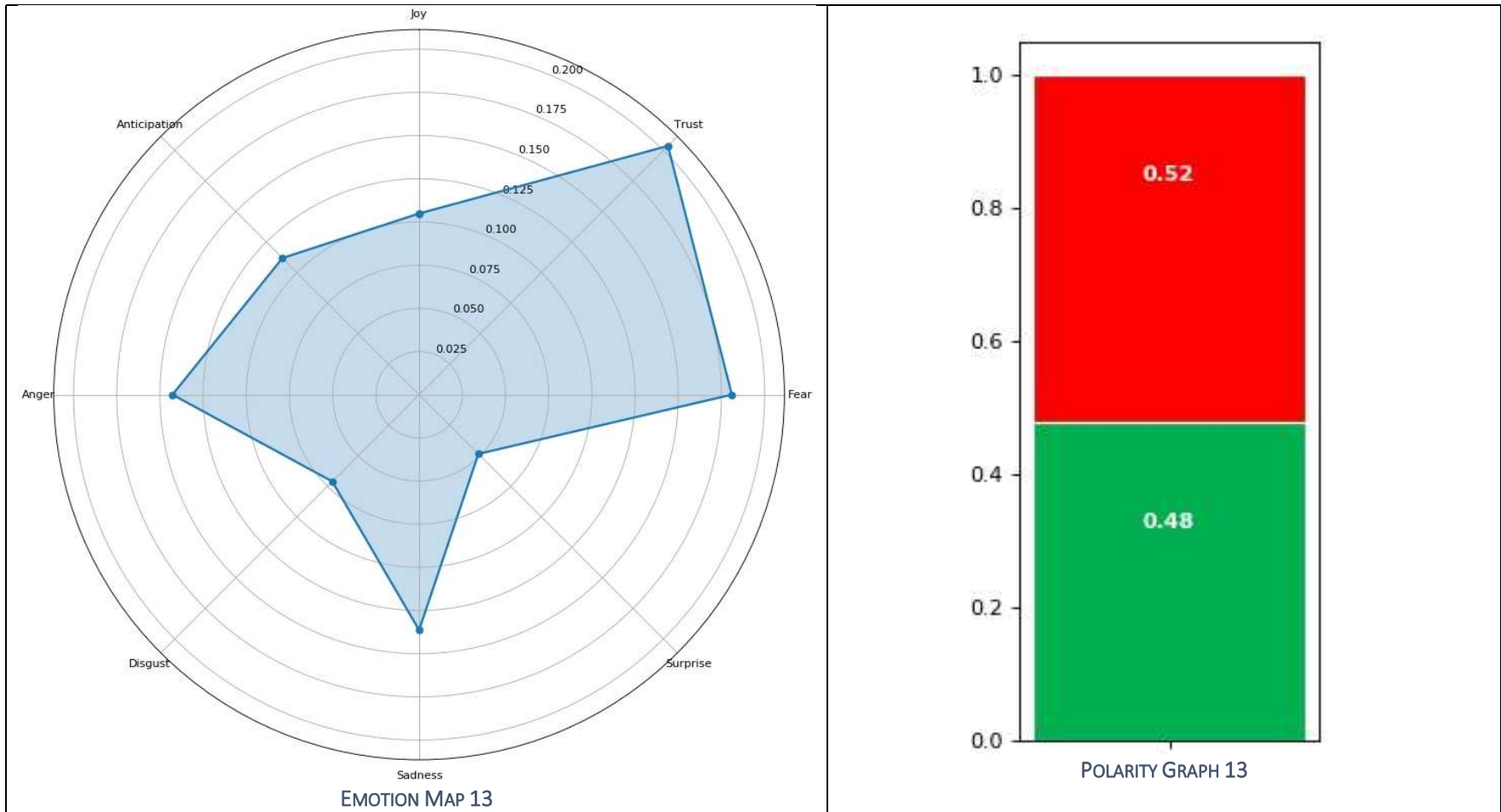
Dabiq – Issue 10

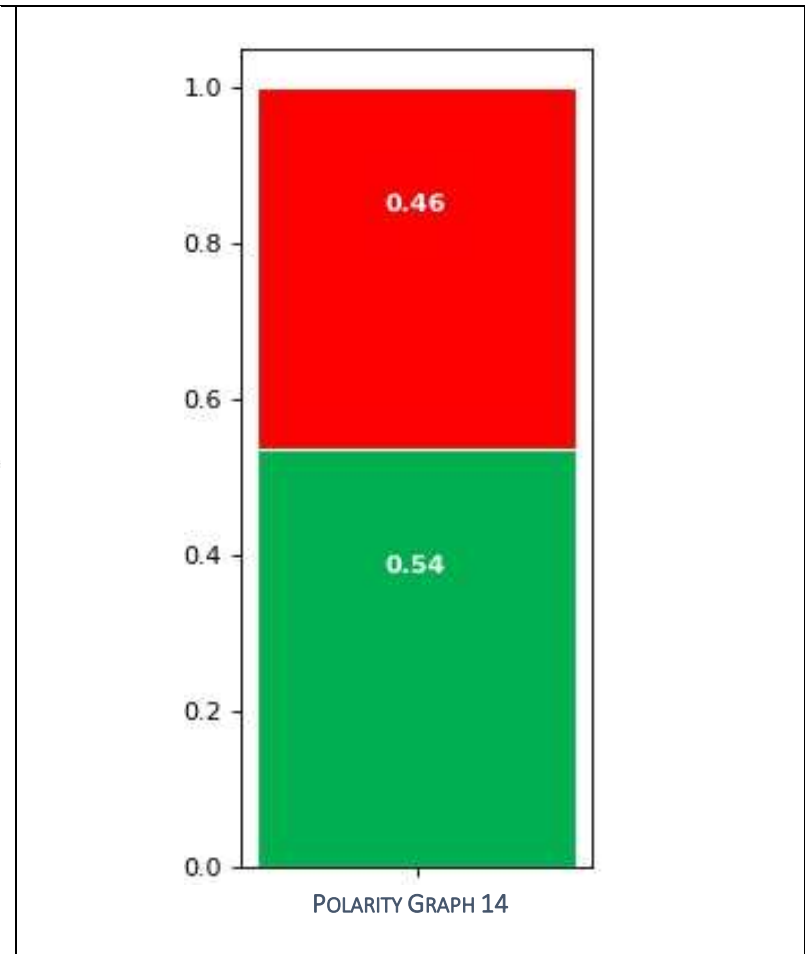
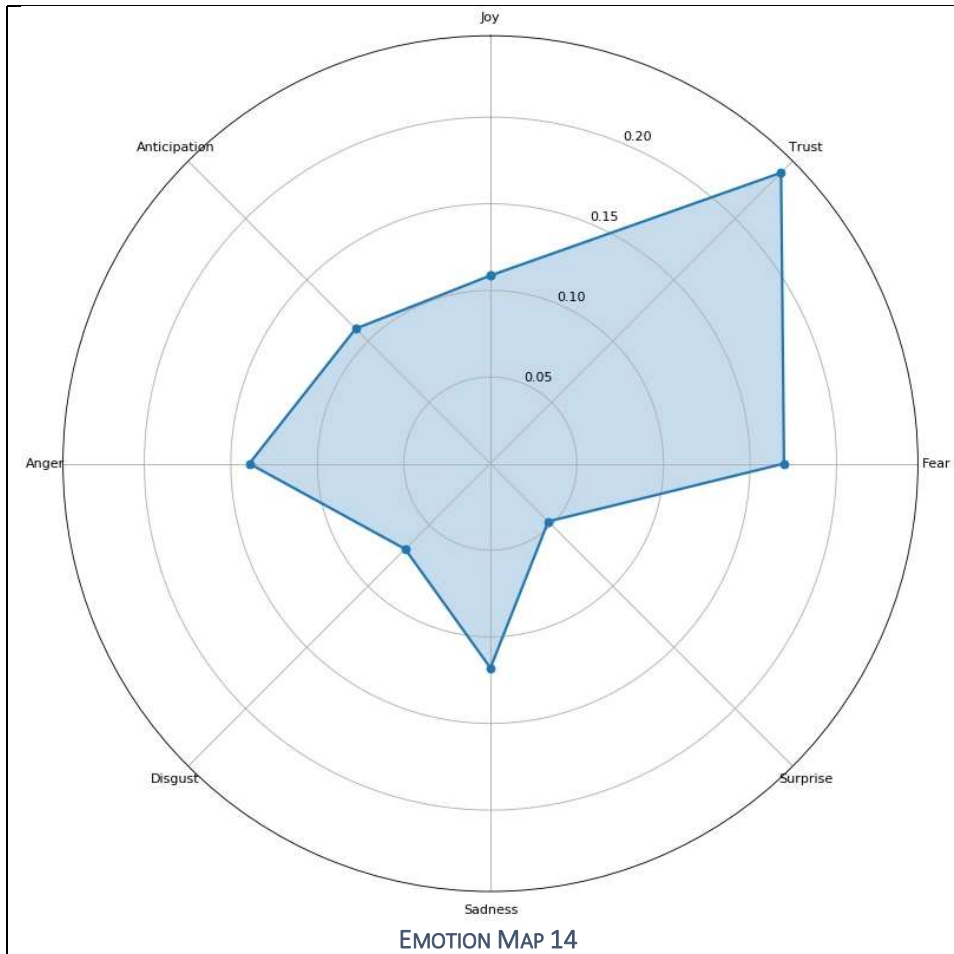


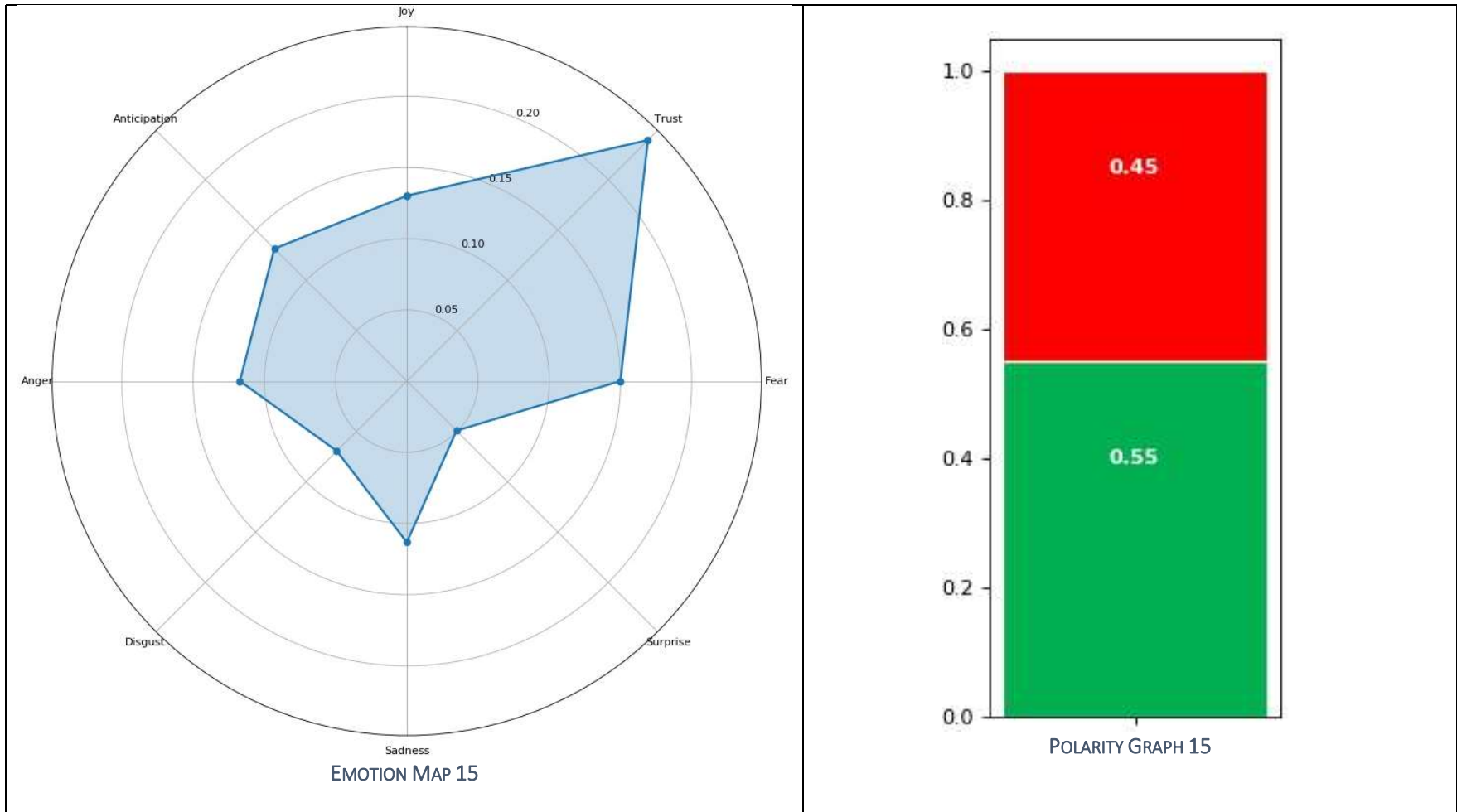
Dabiq – Issue 11





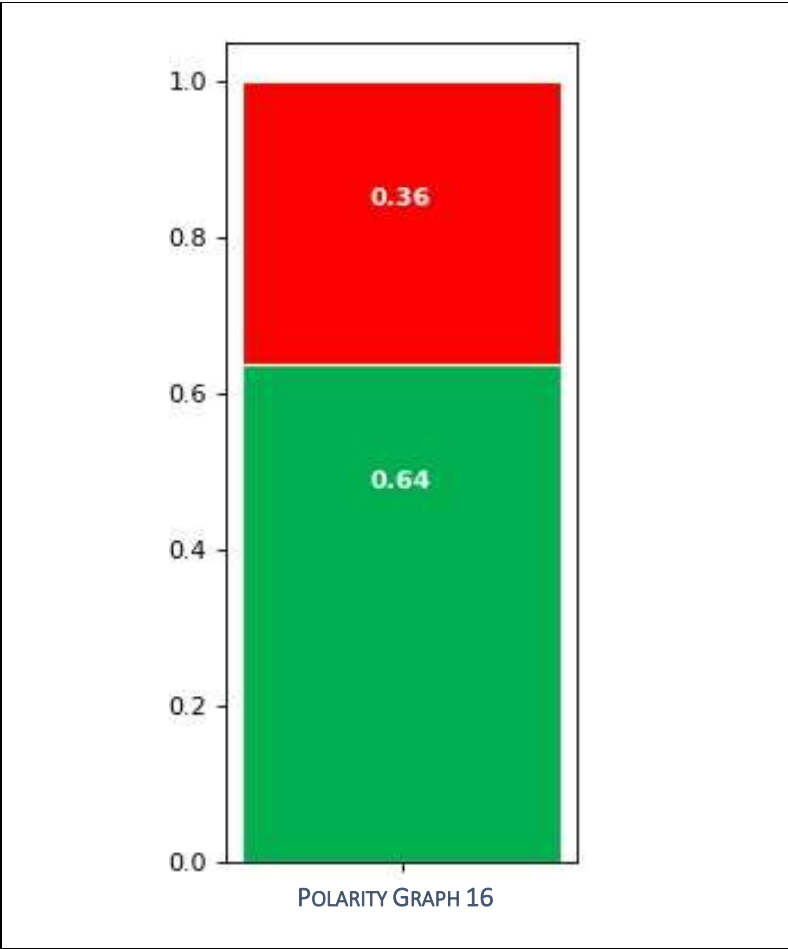
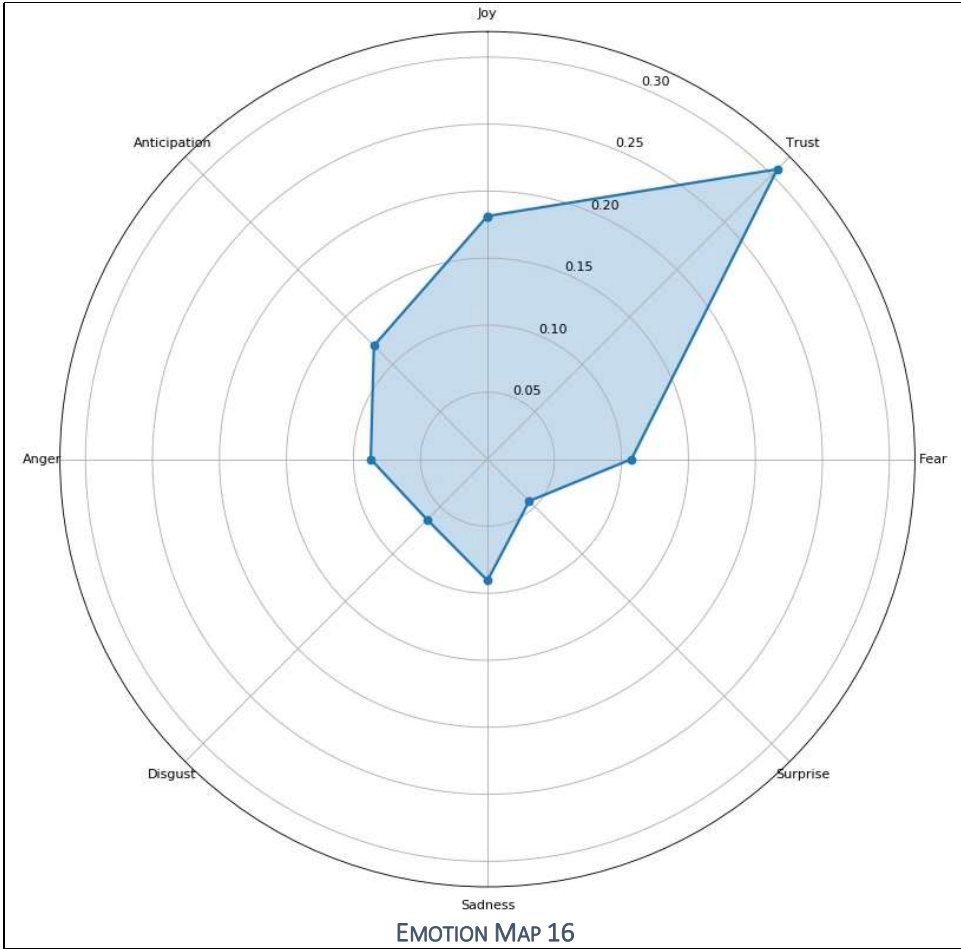


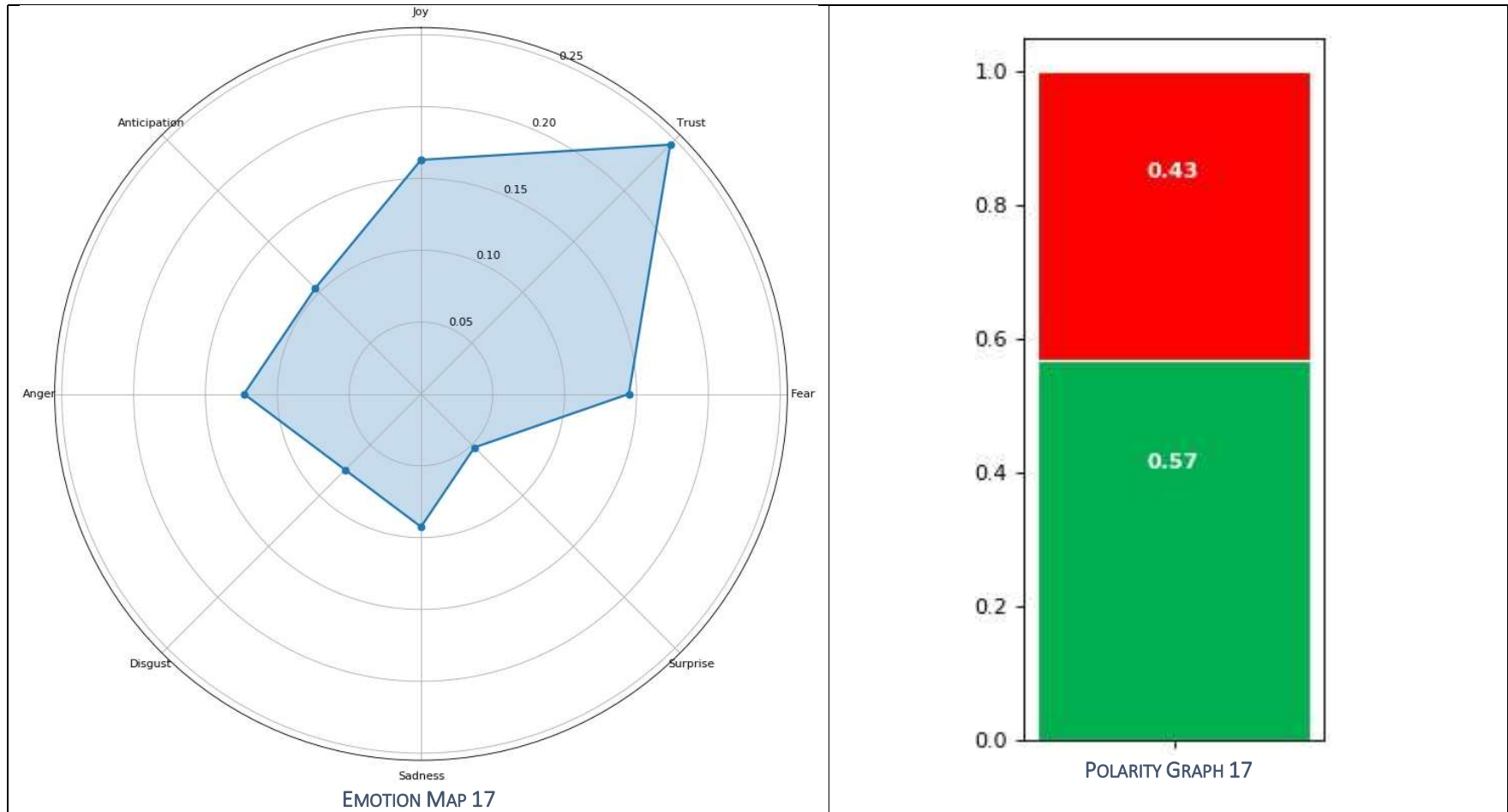


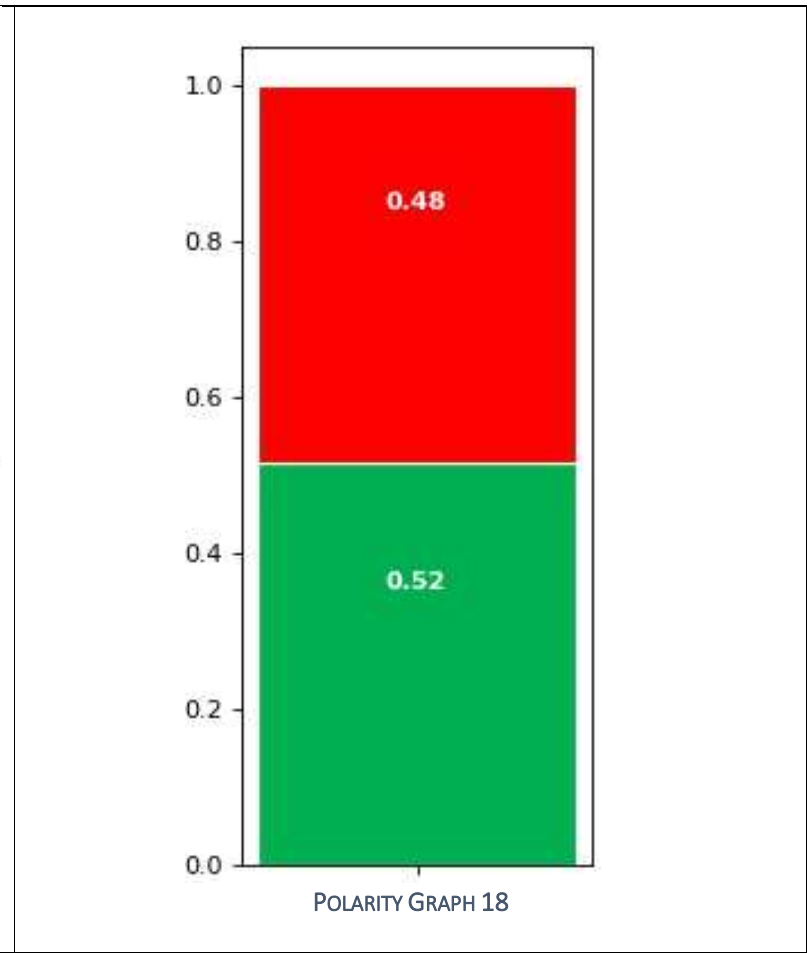
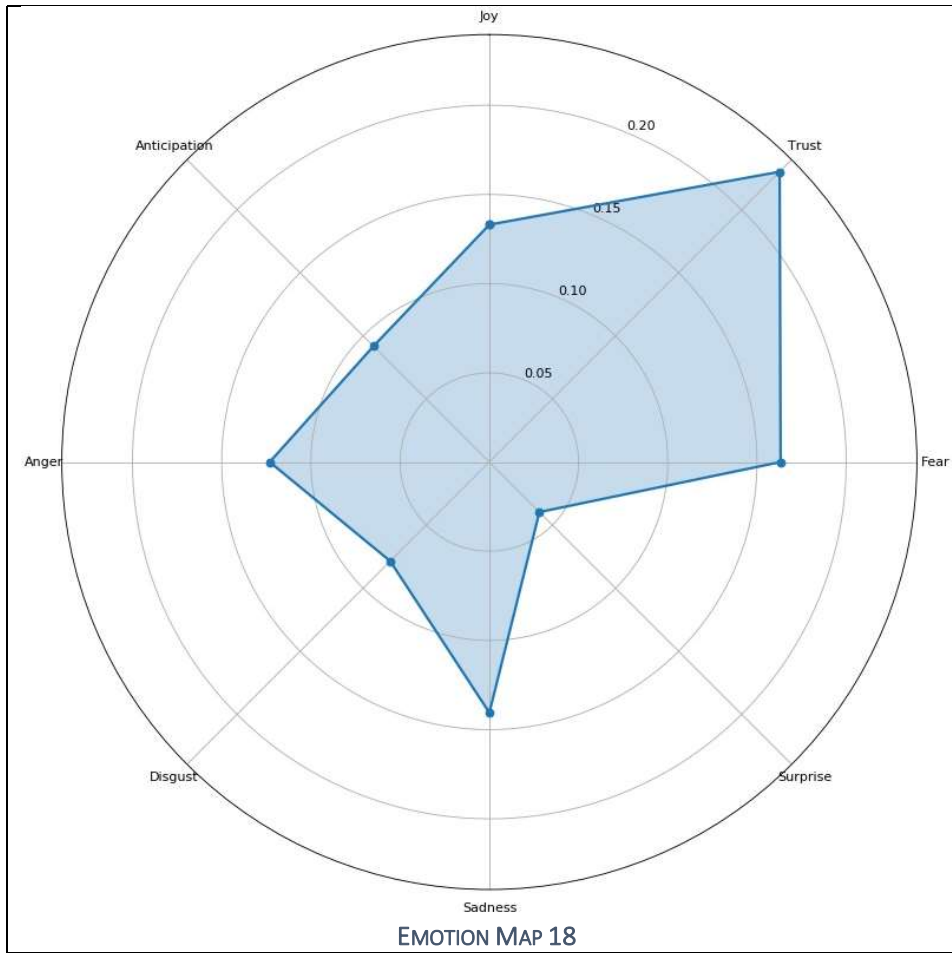


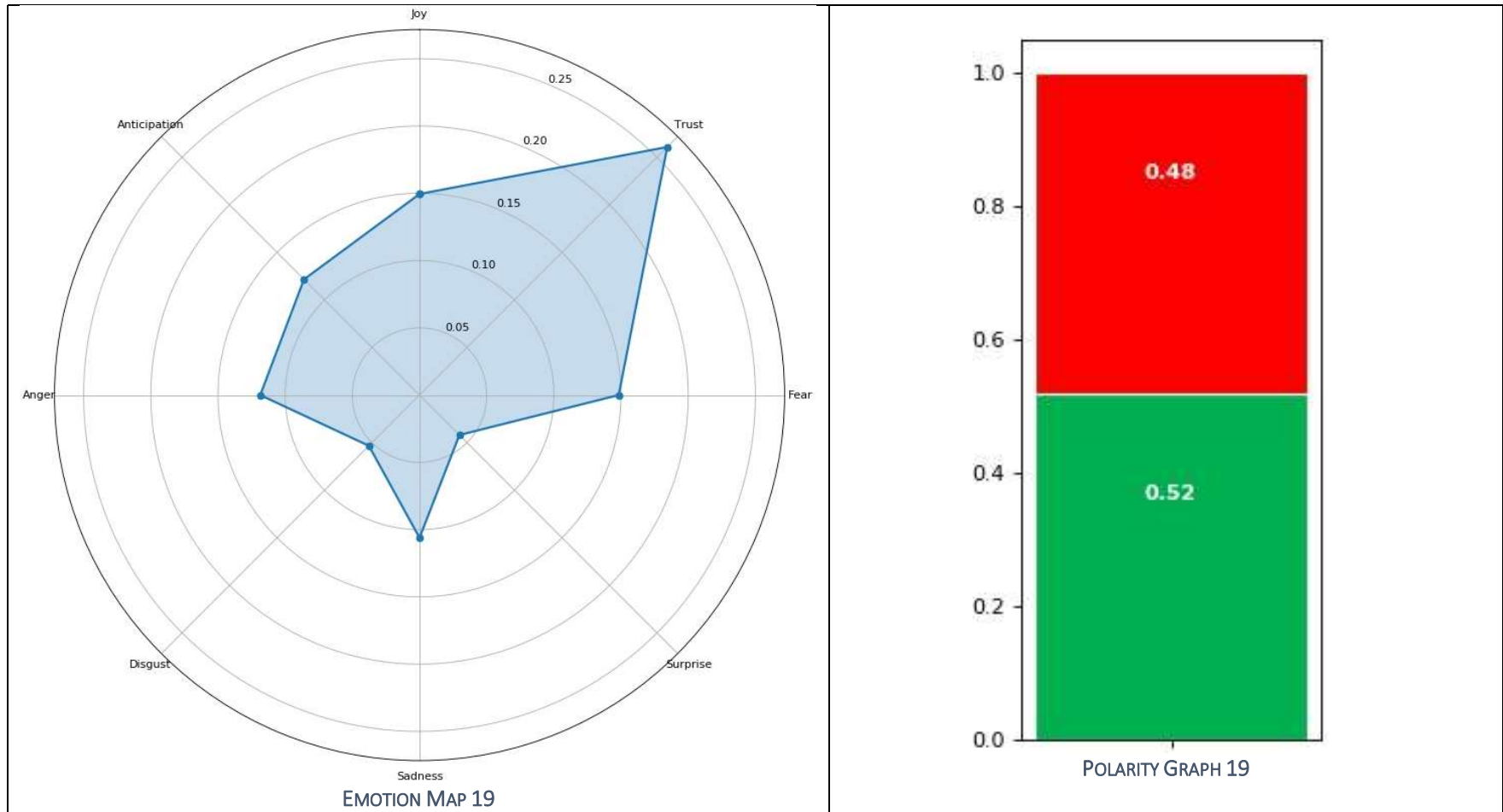
13.2 Daral-Islam

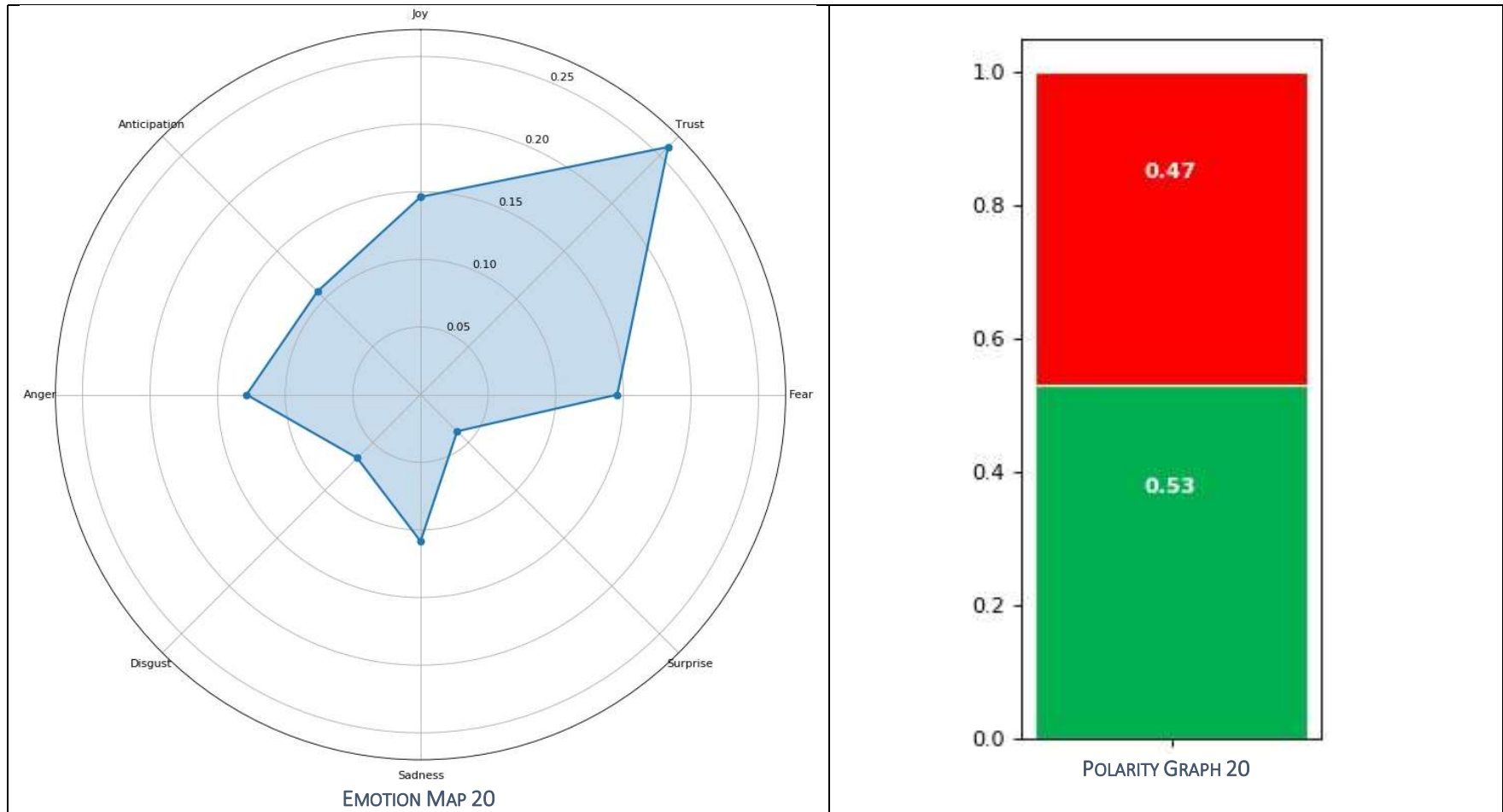
- Issue 1

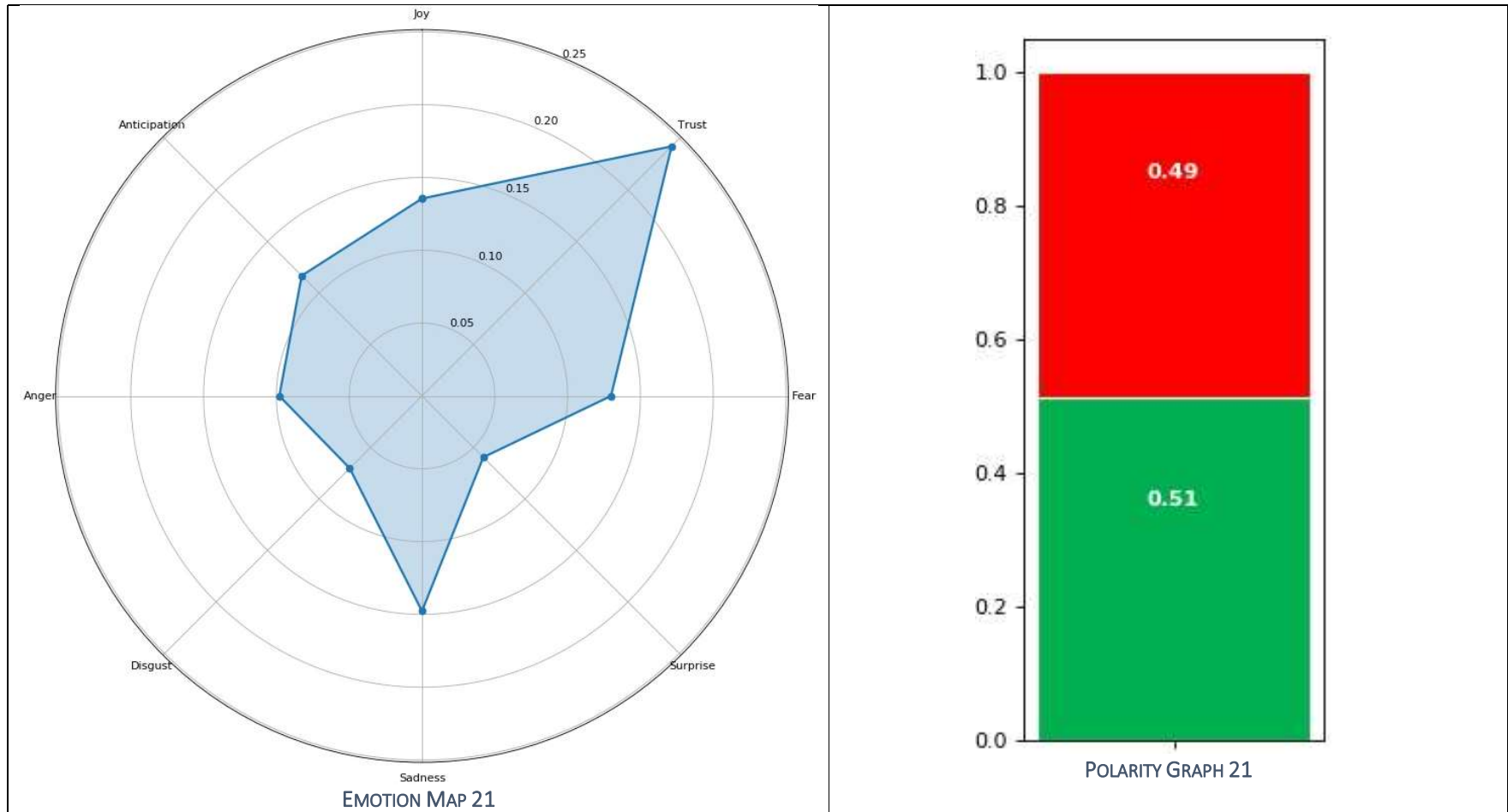


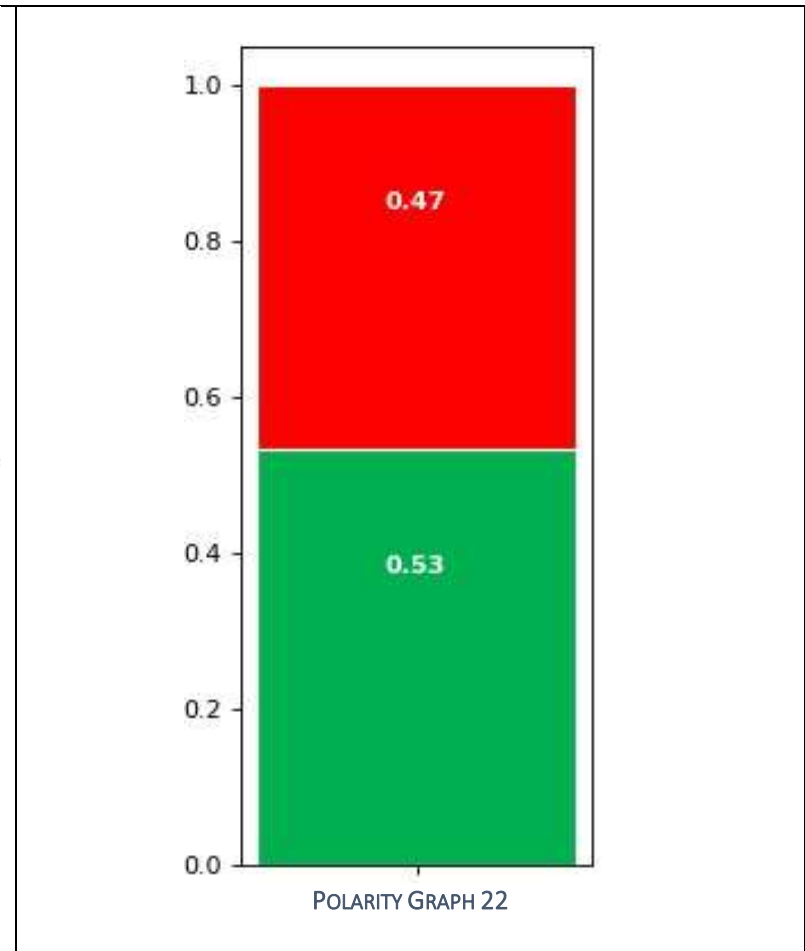
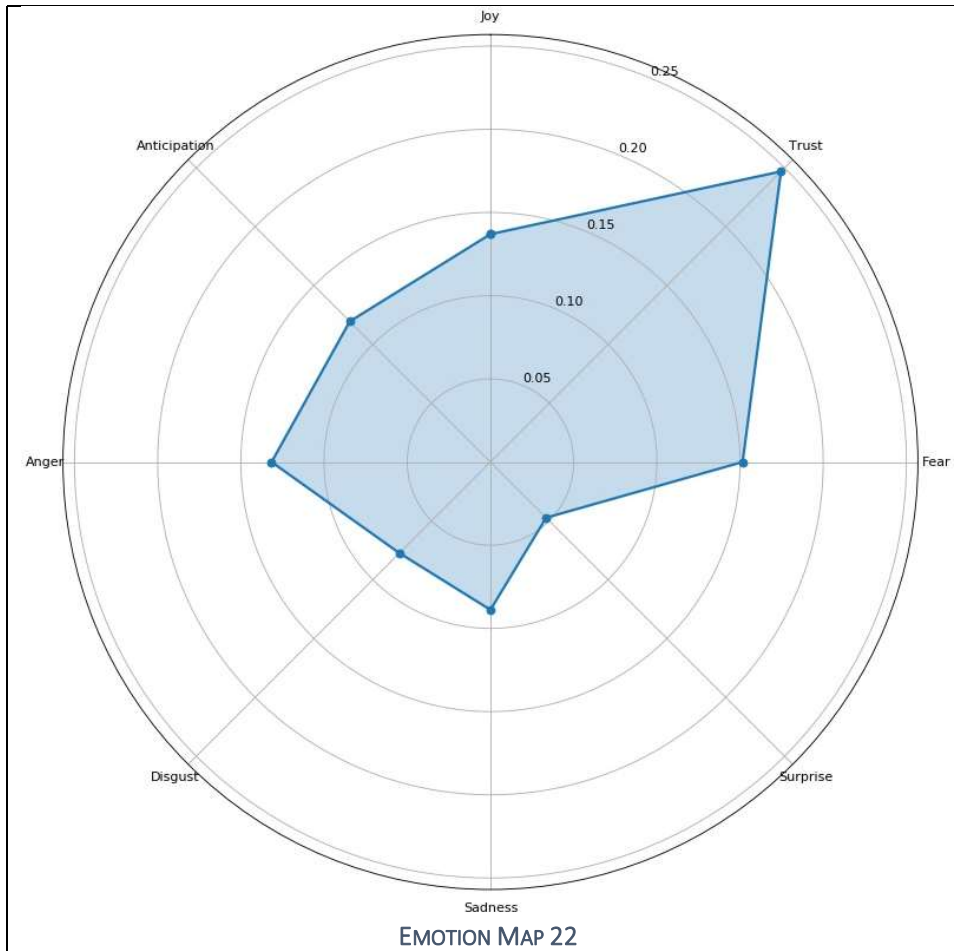


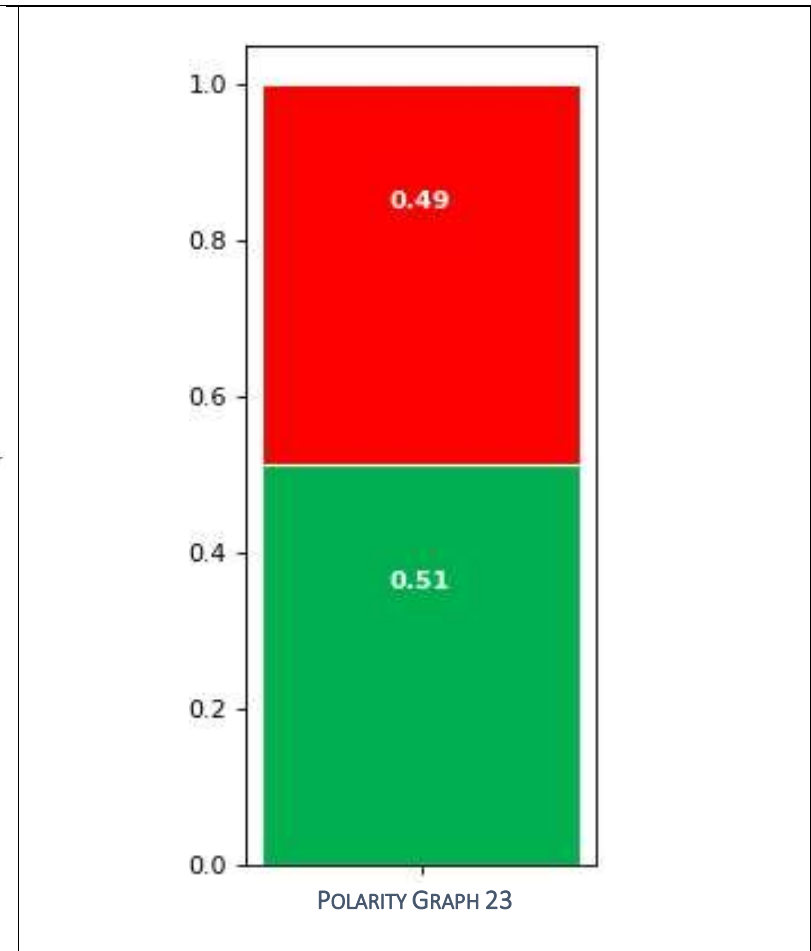
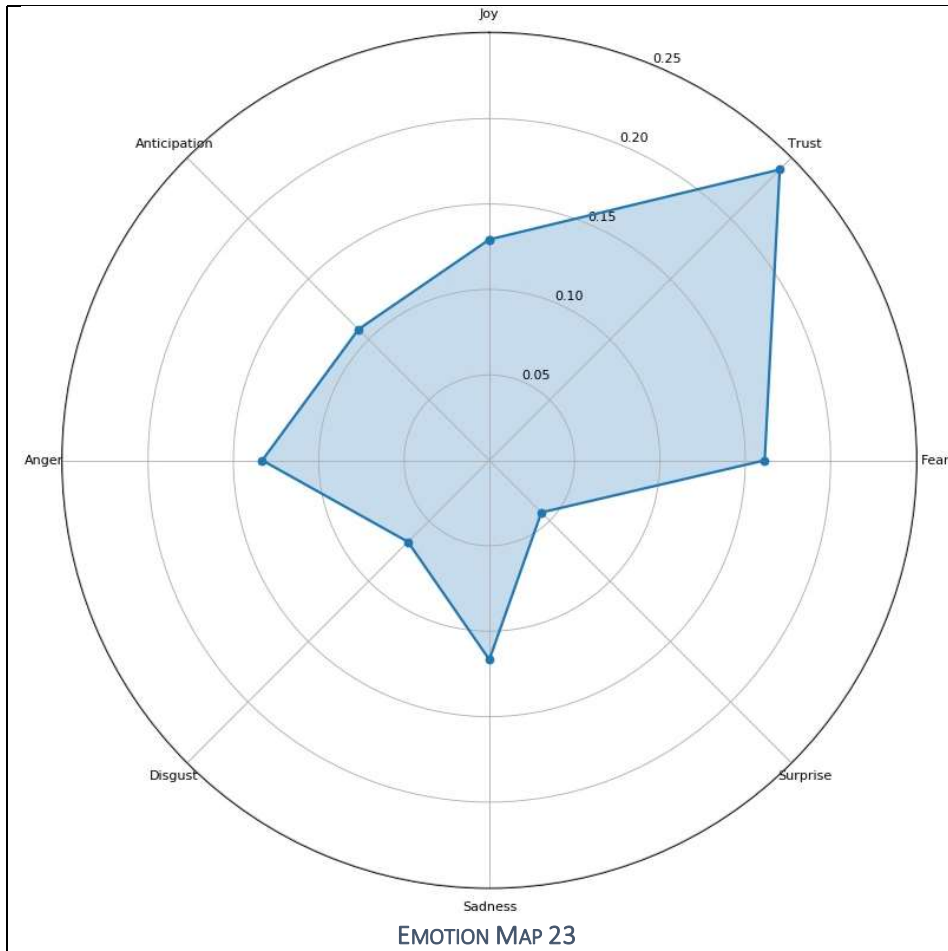


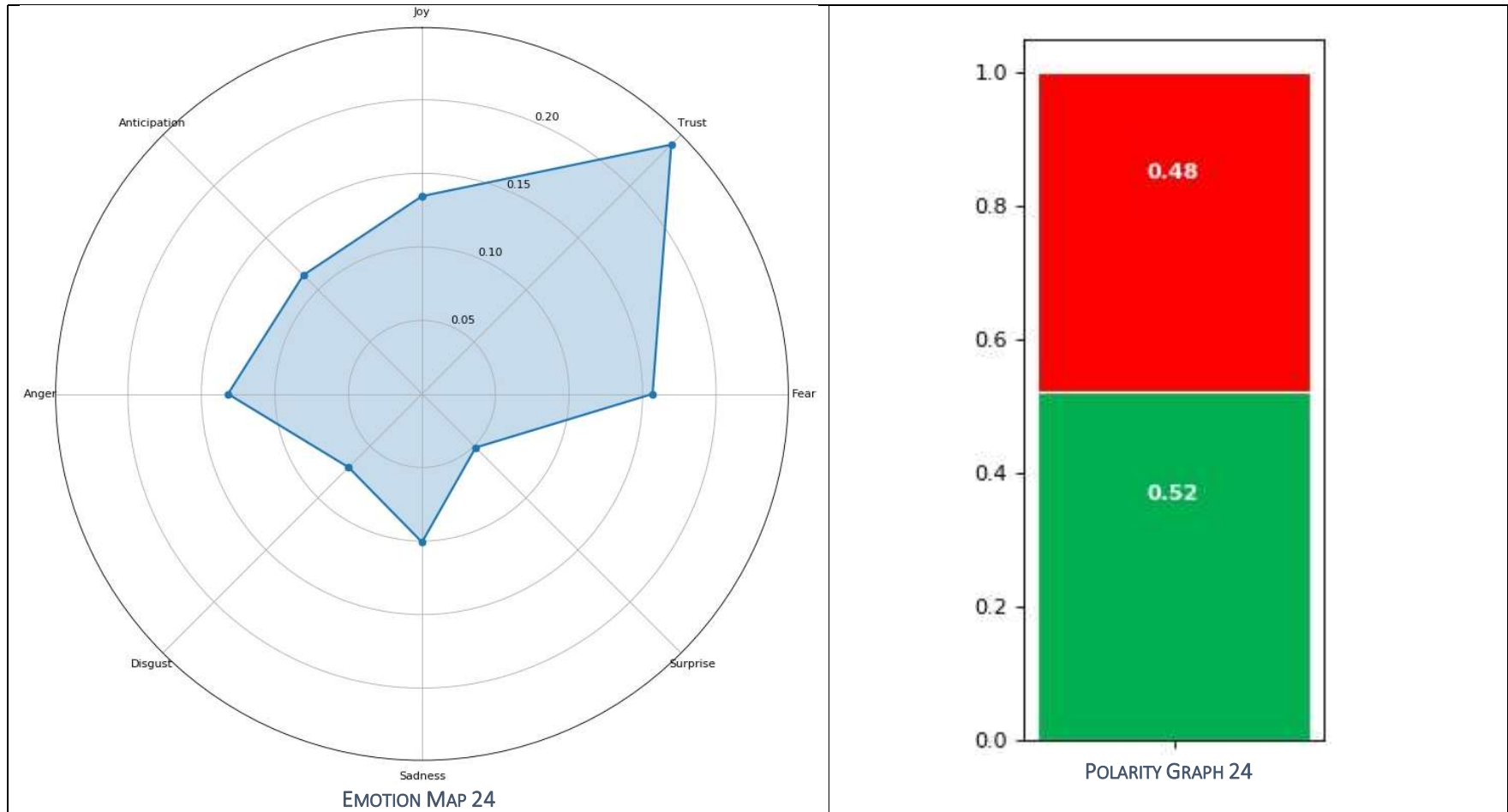


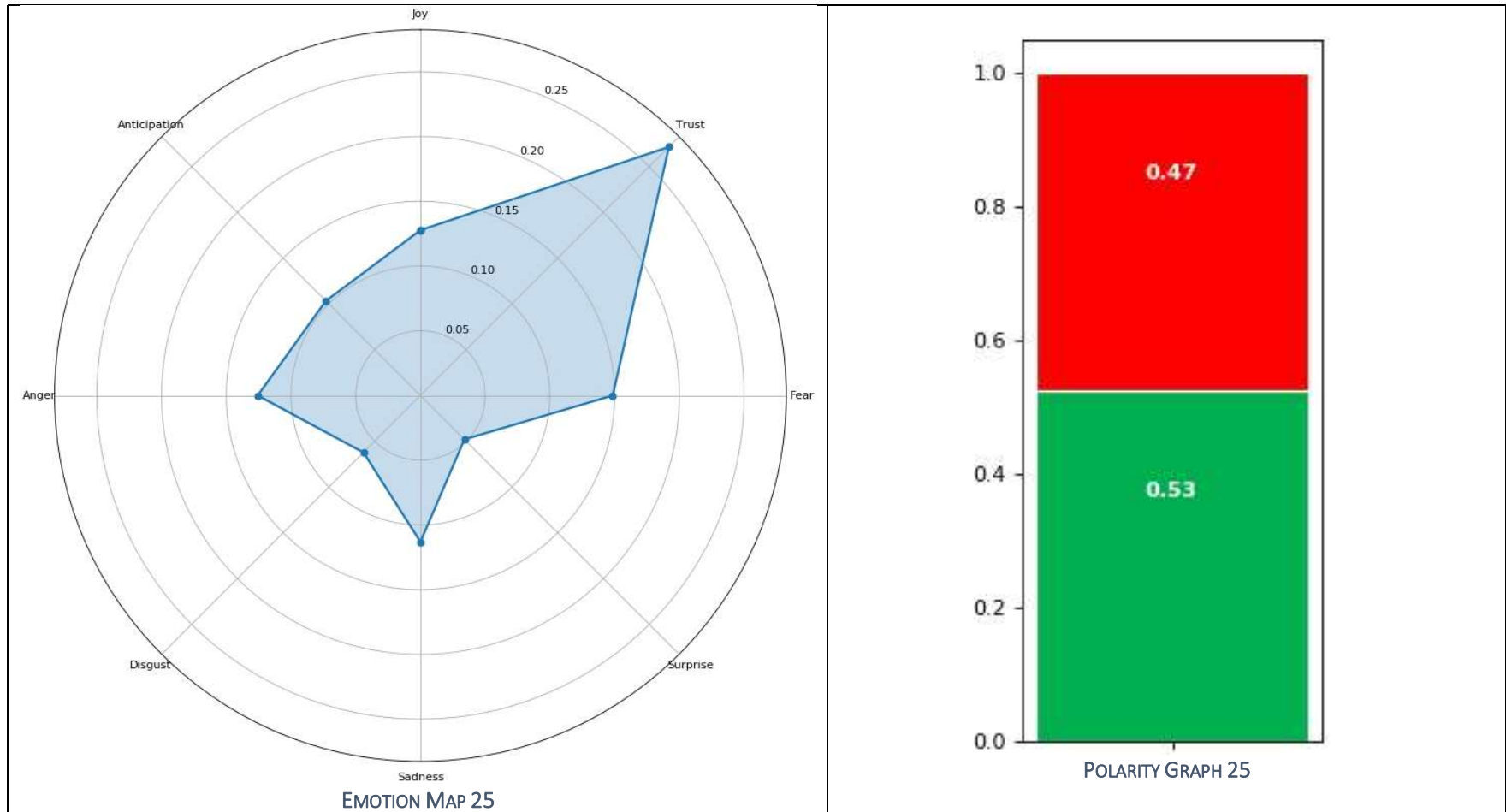






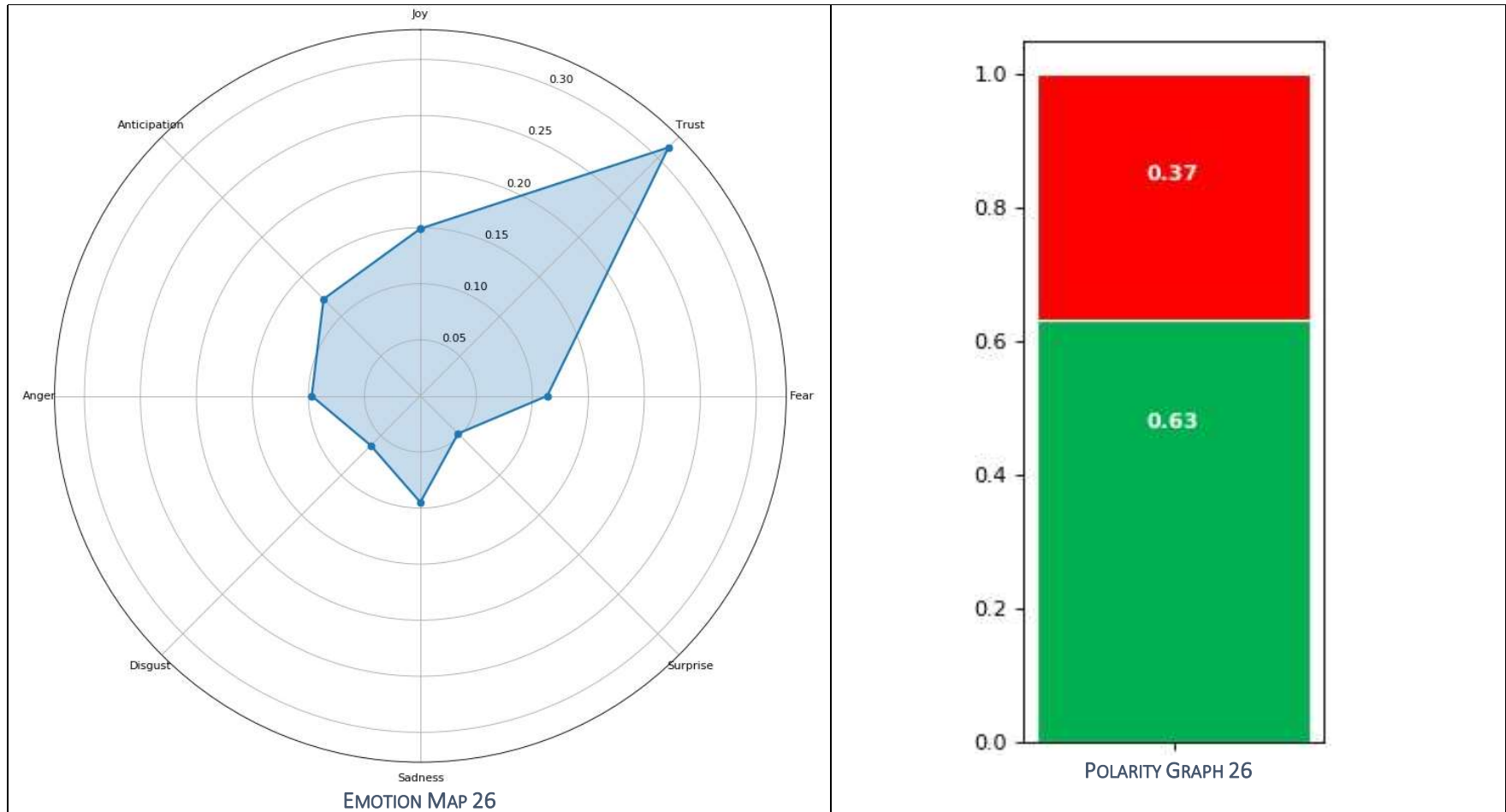


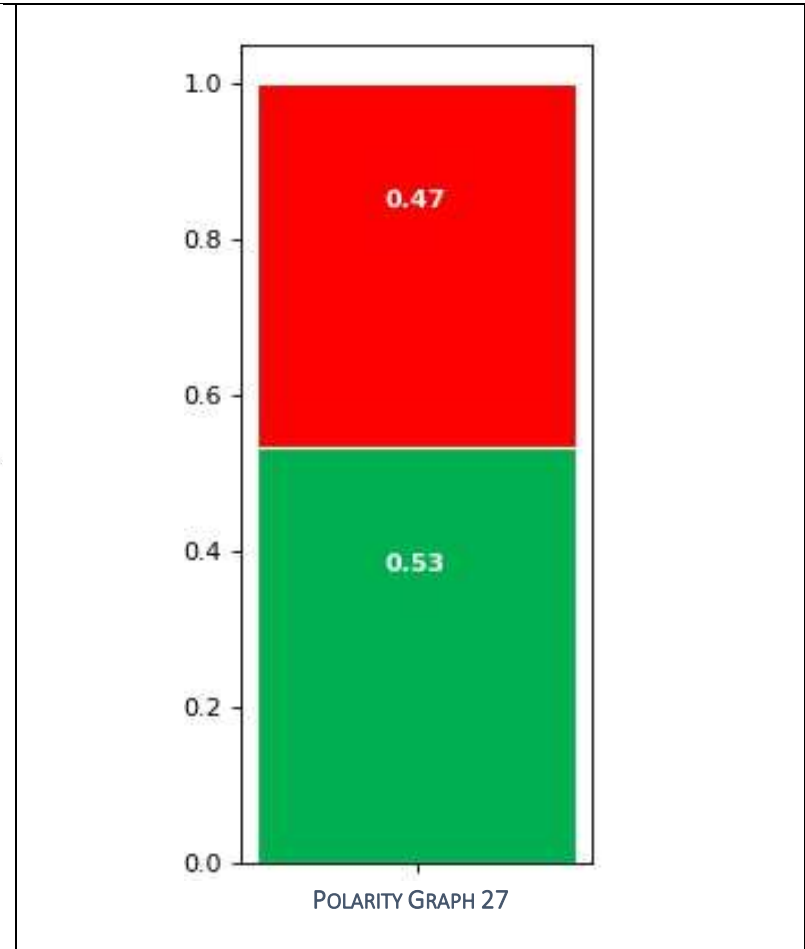
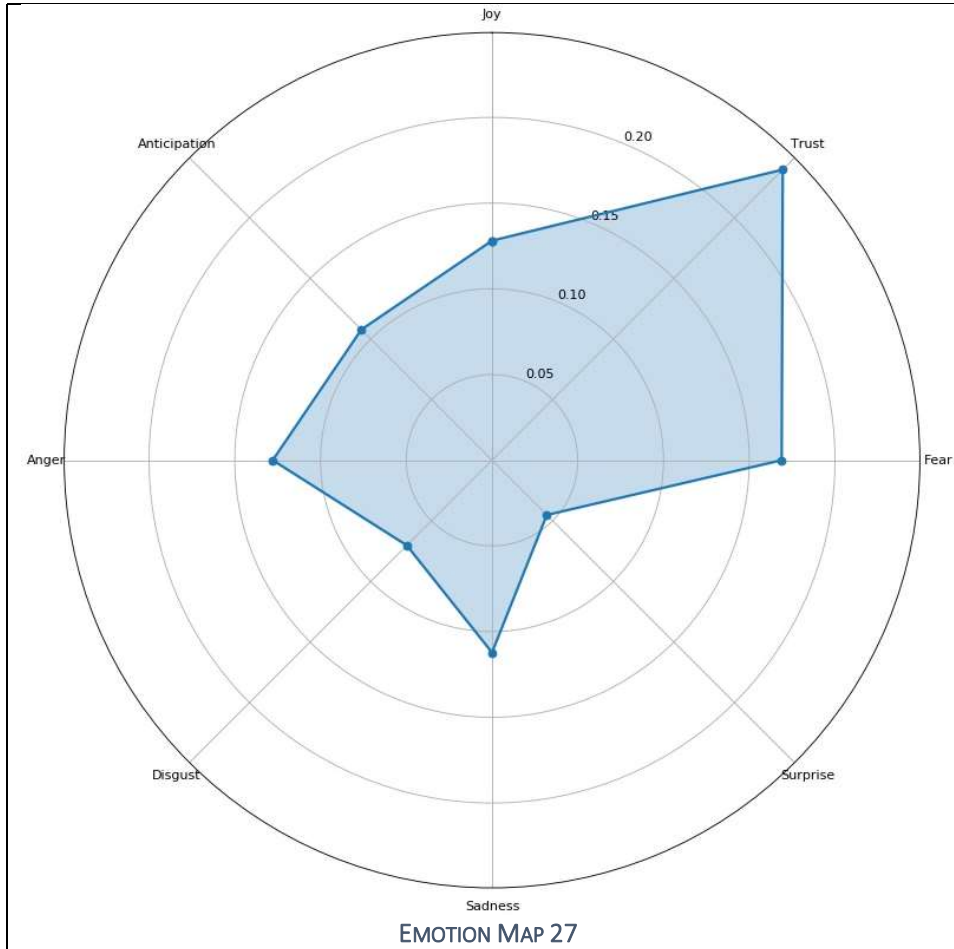


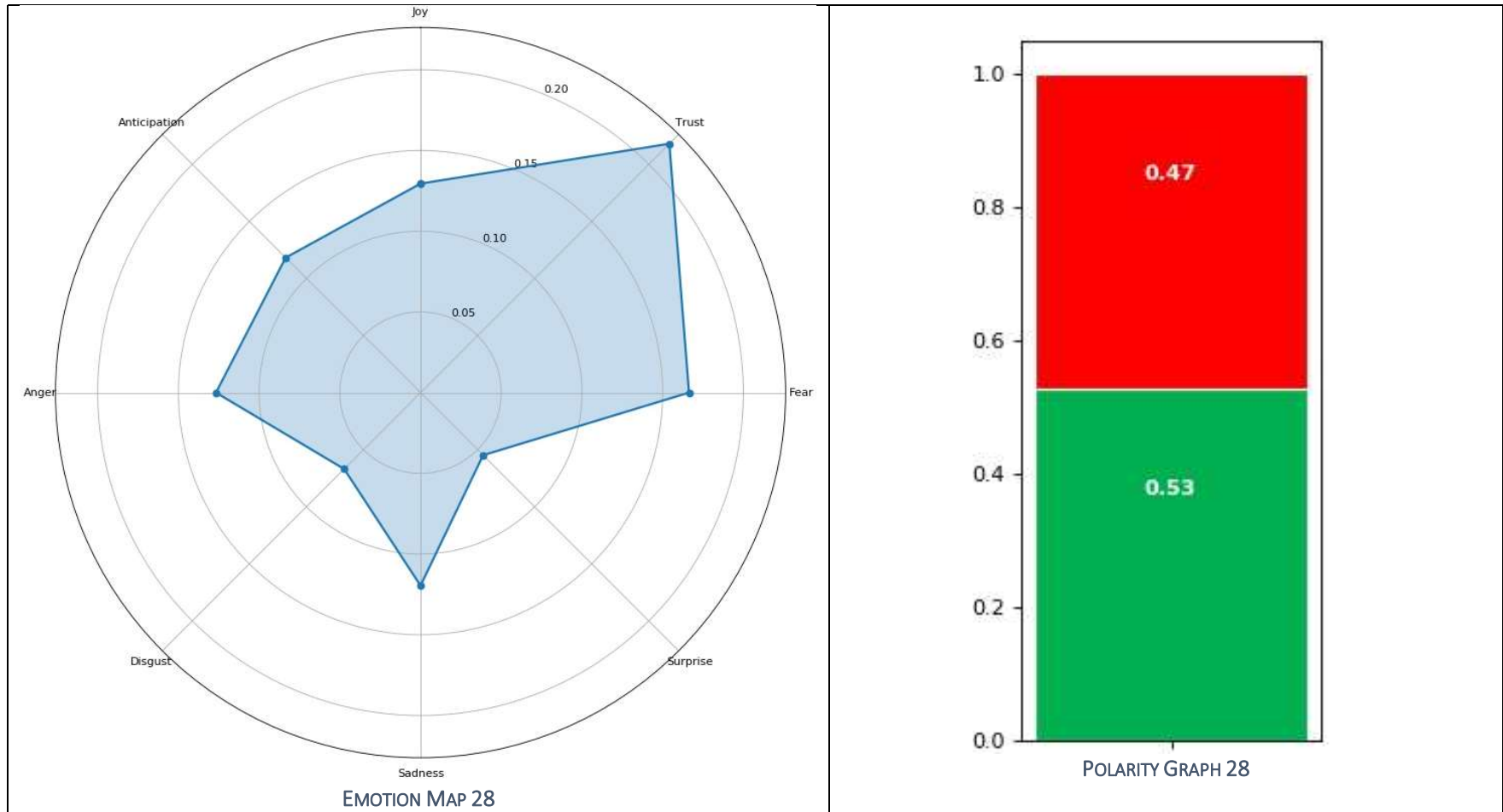


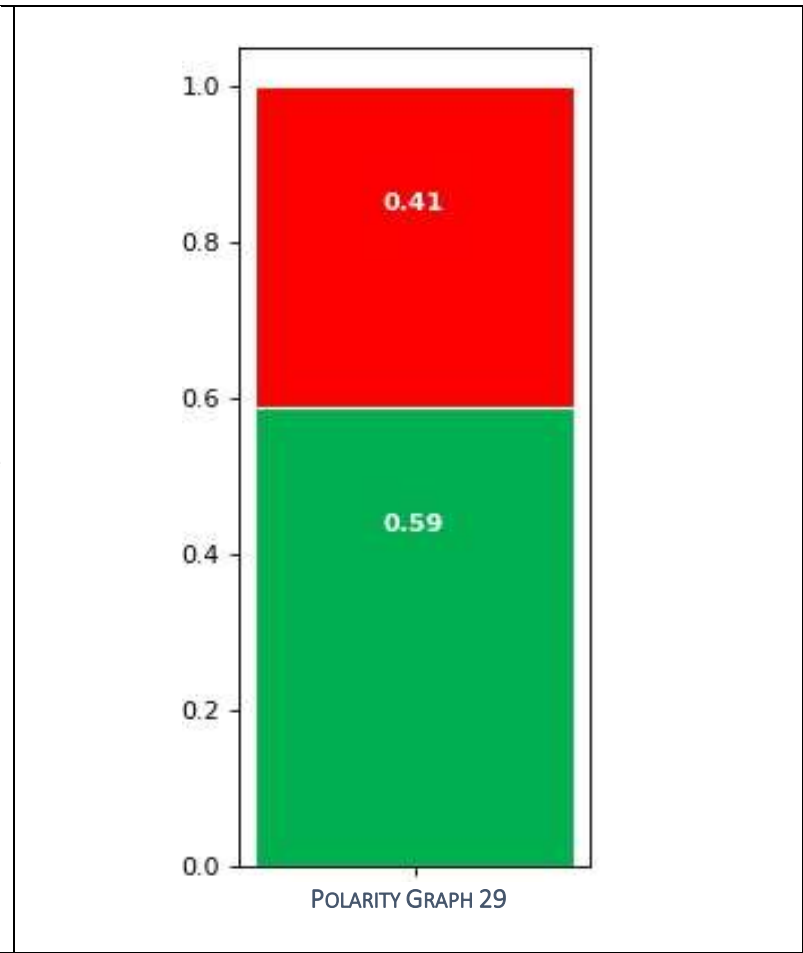
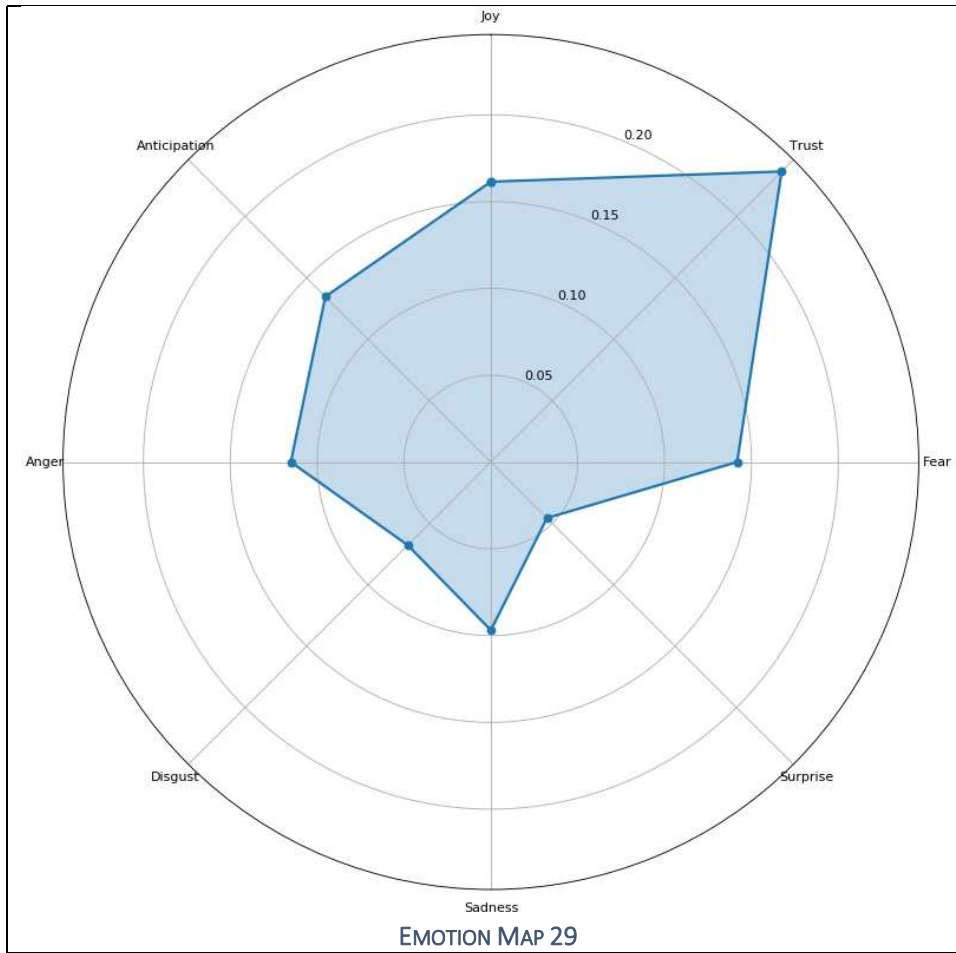
13.3Istok

– Issue 1



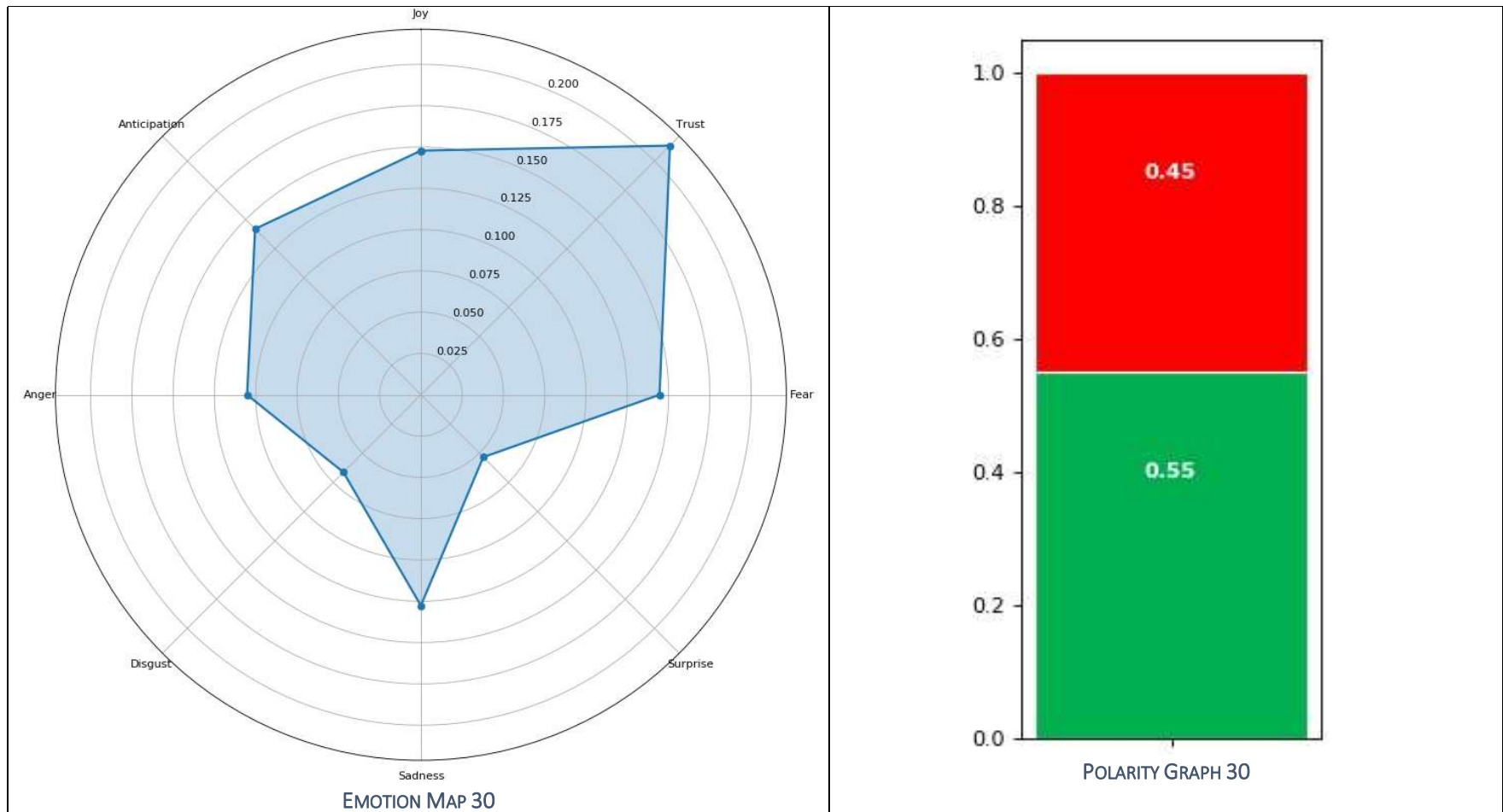


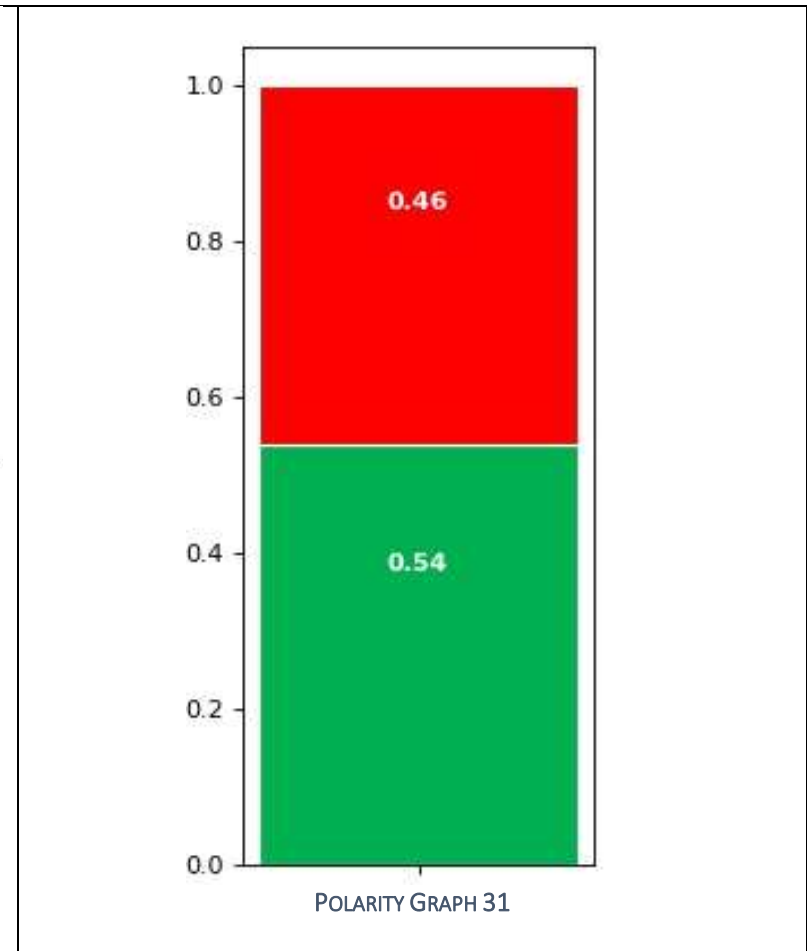
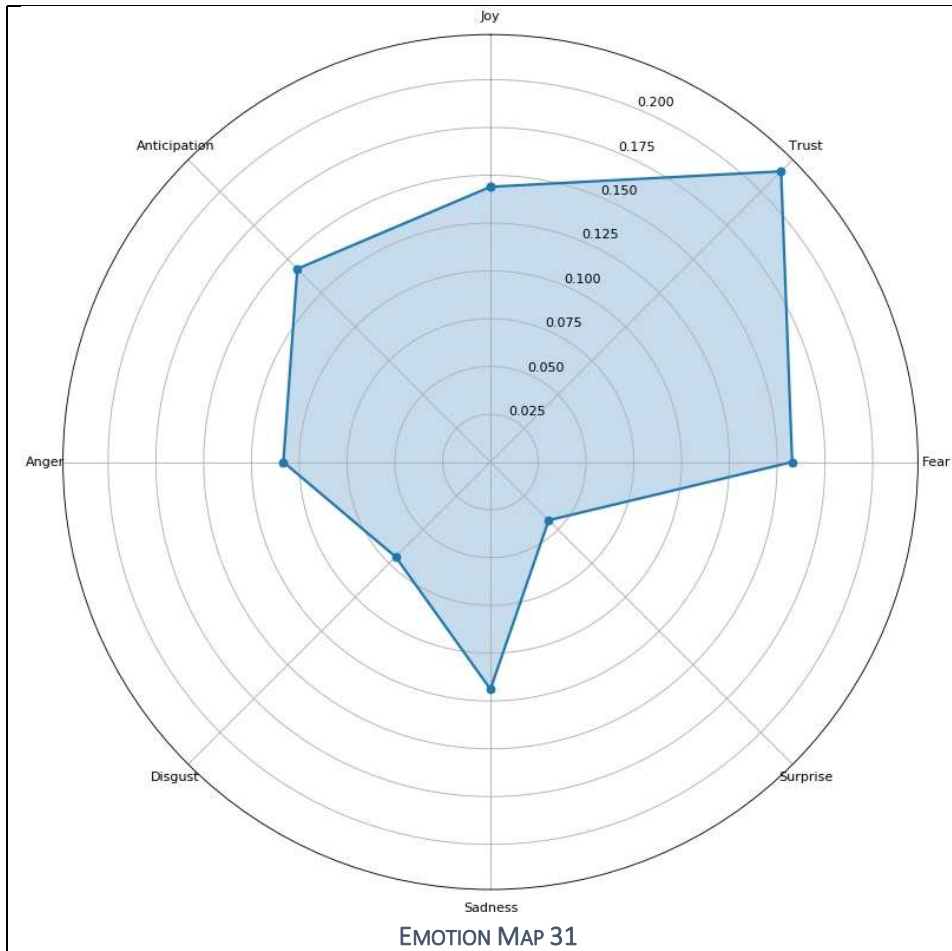


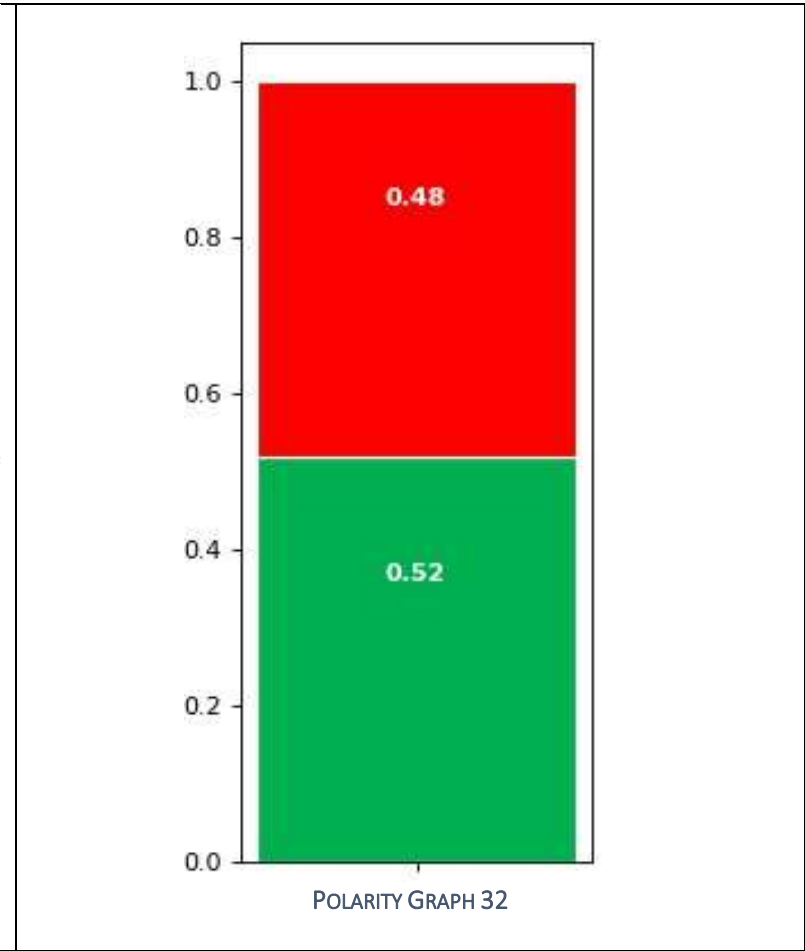
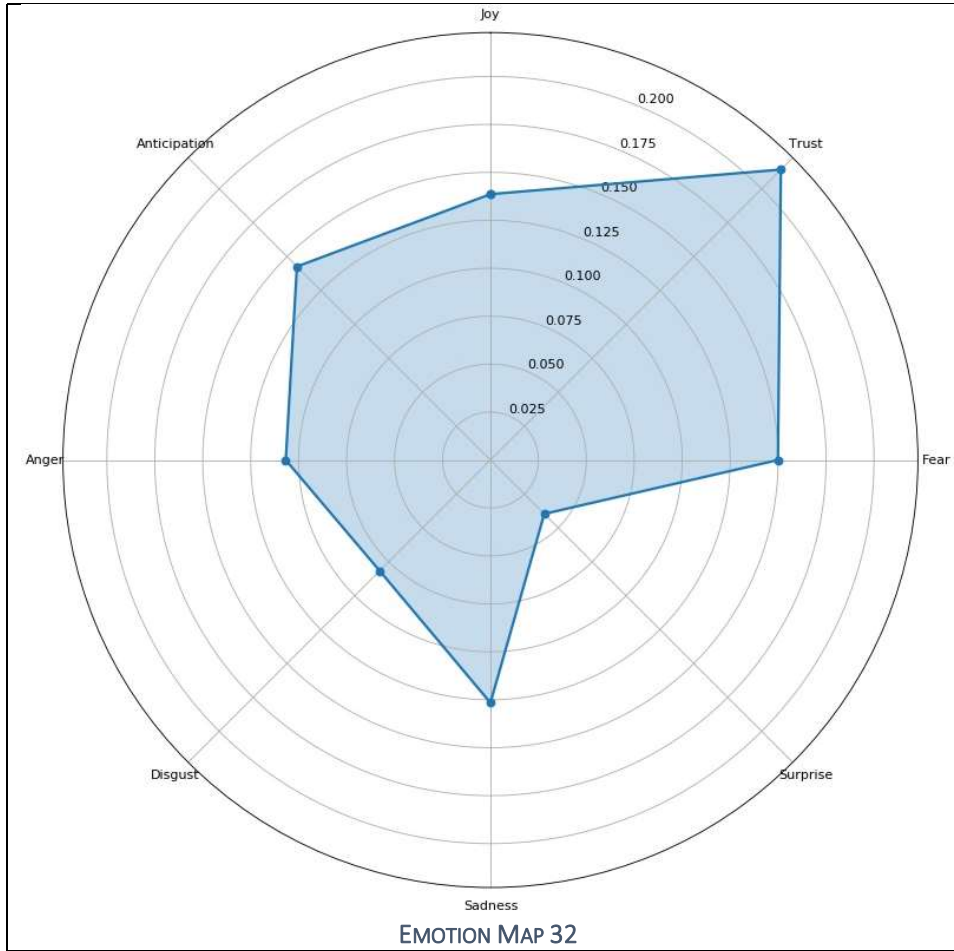


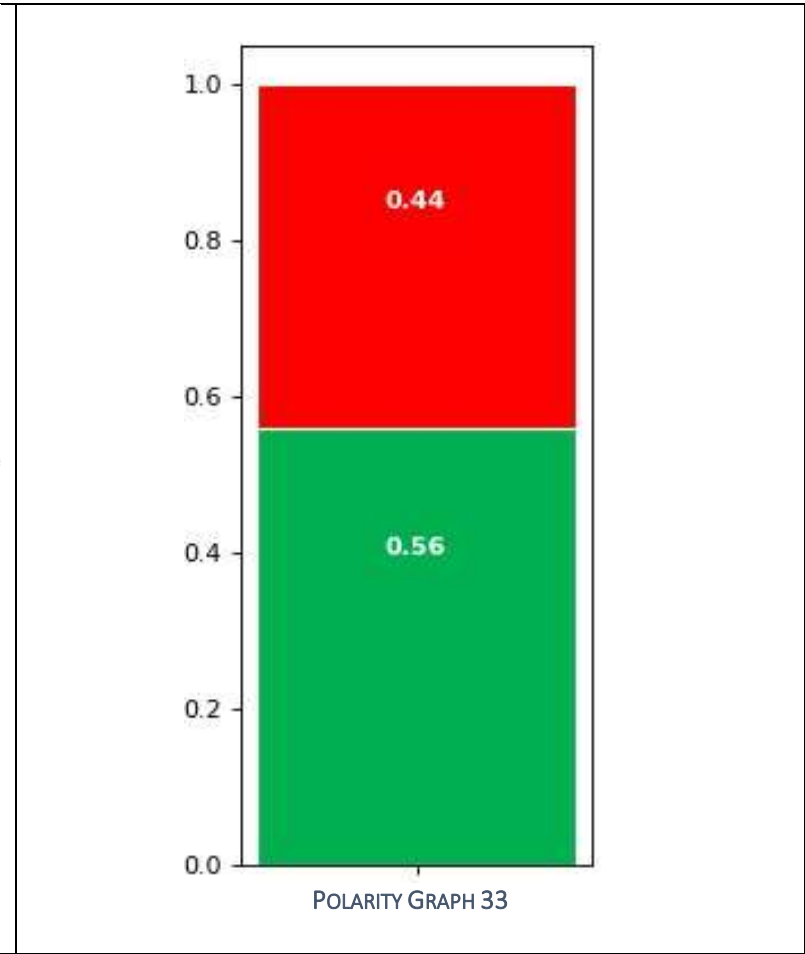
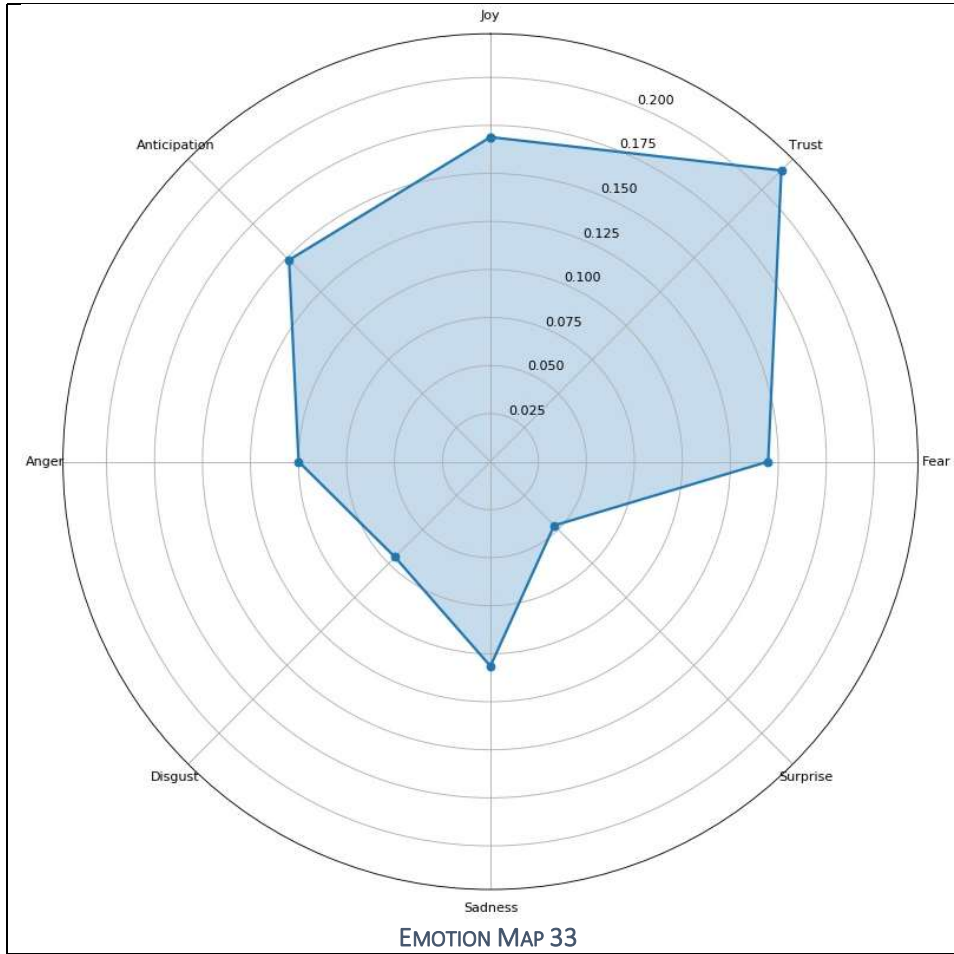
13.4Konstantiniyye

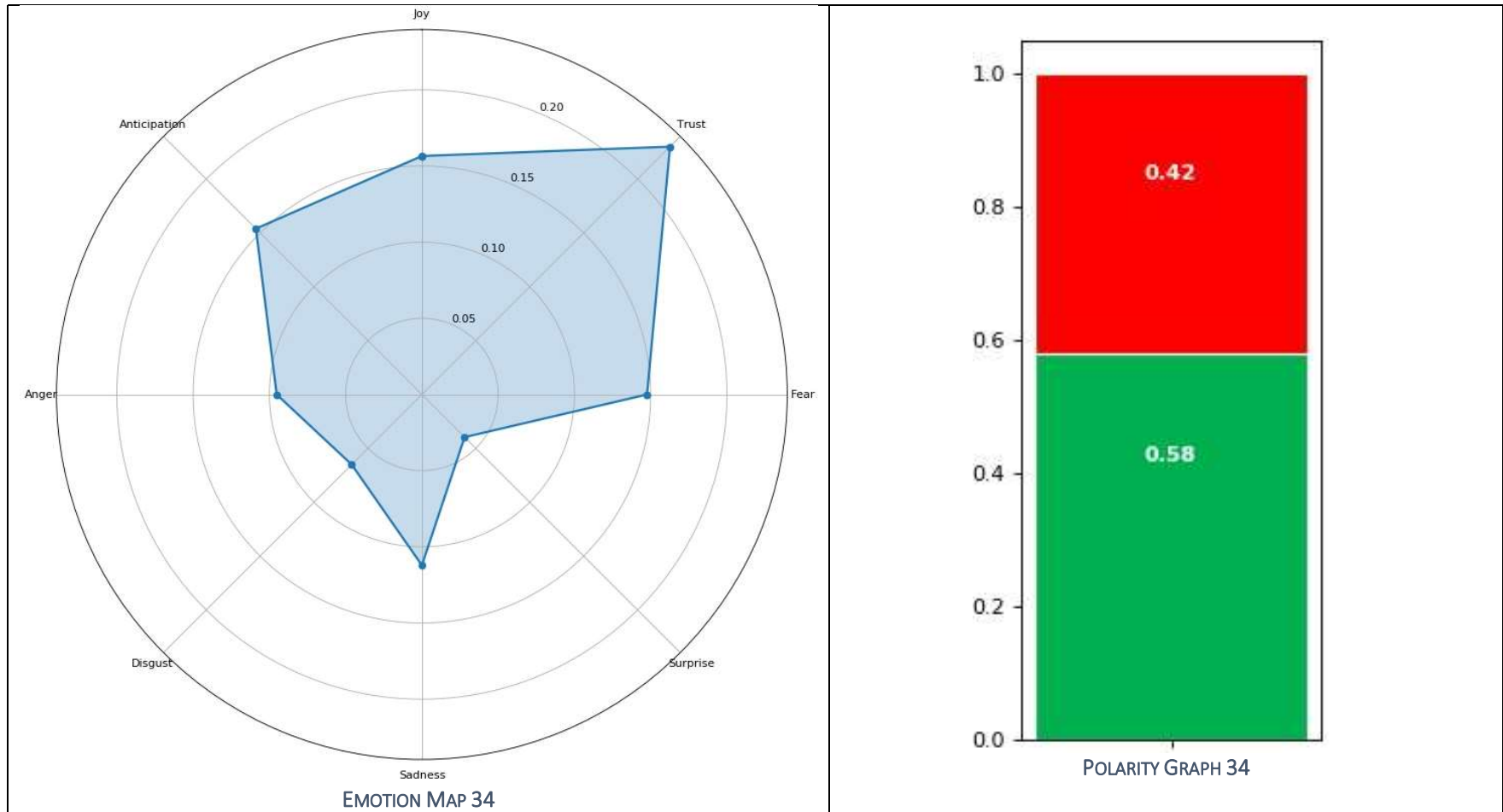
- Issue 1

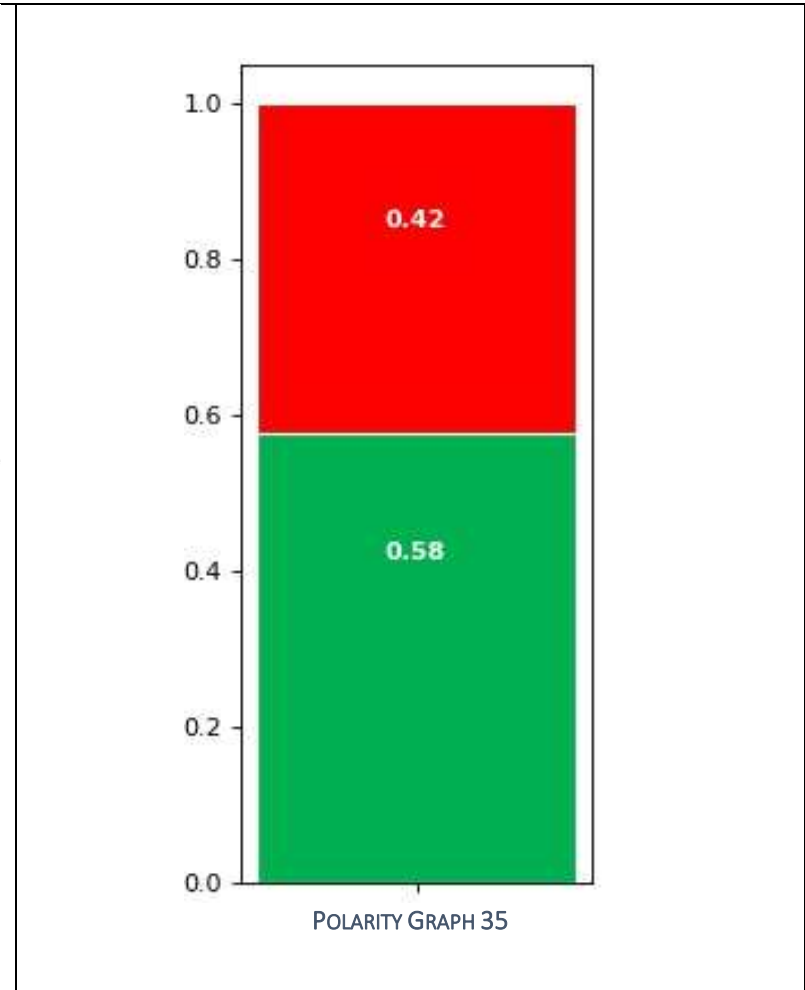
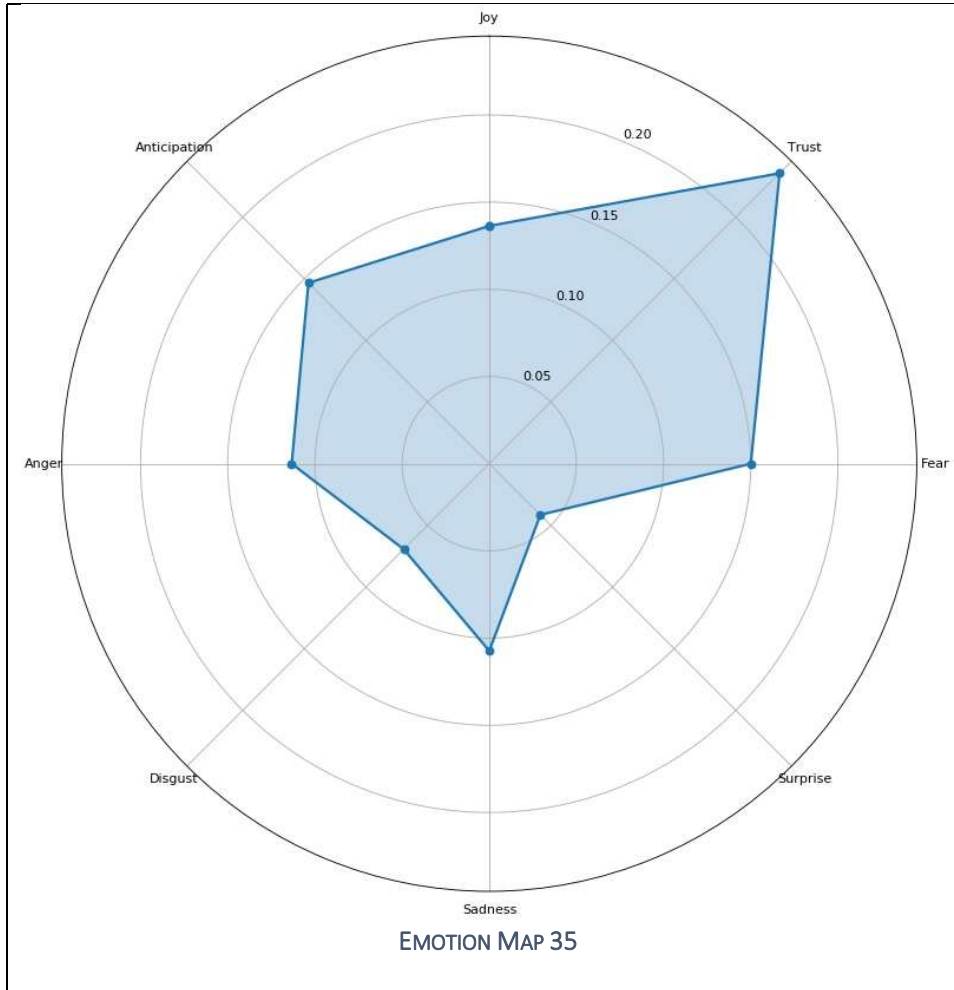


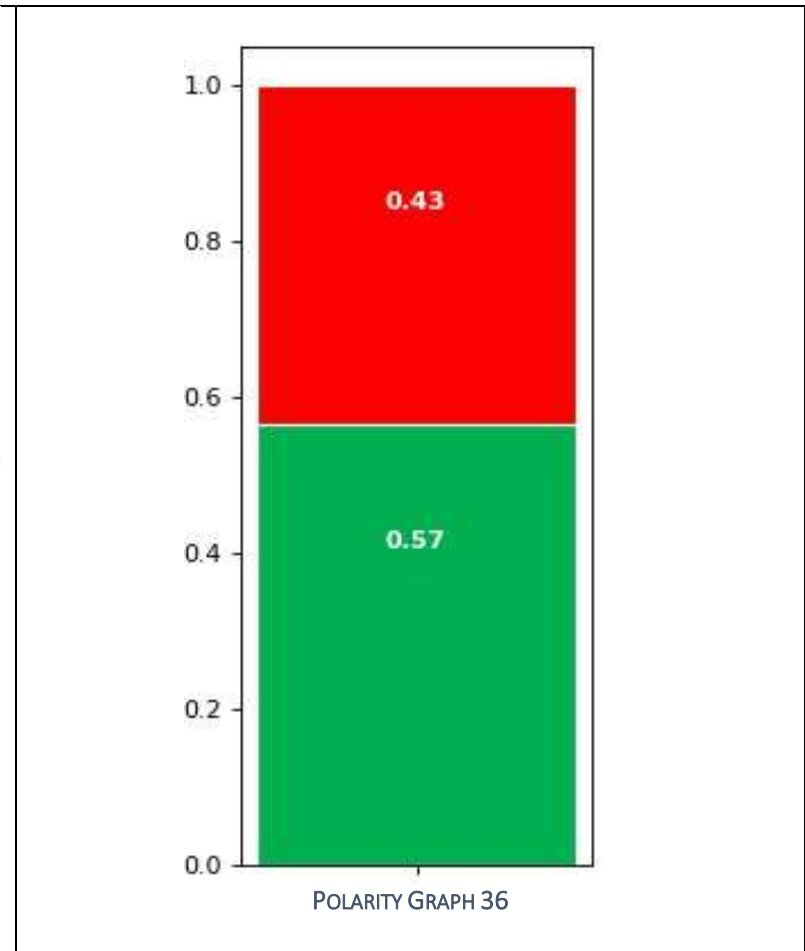
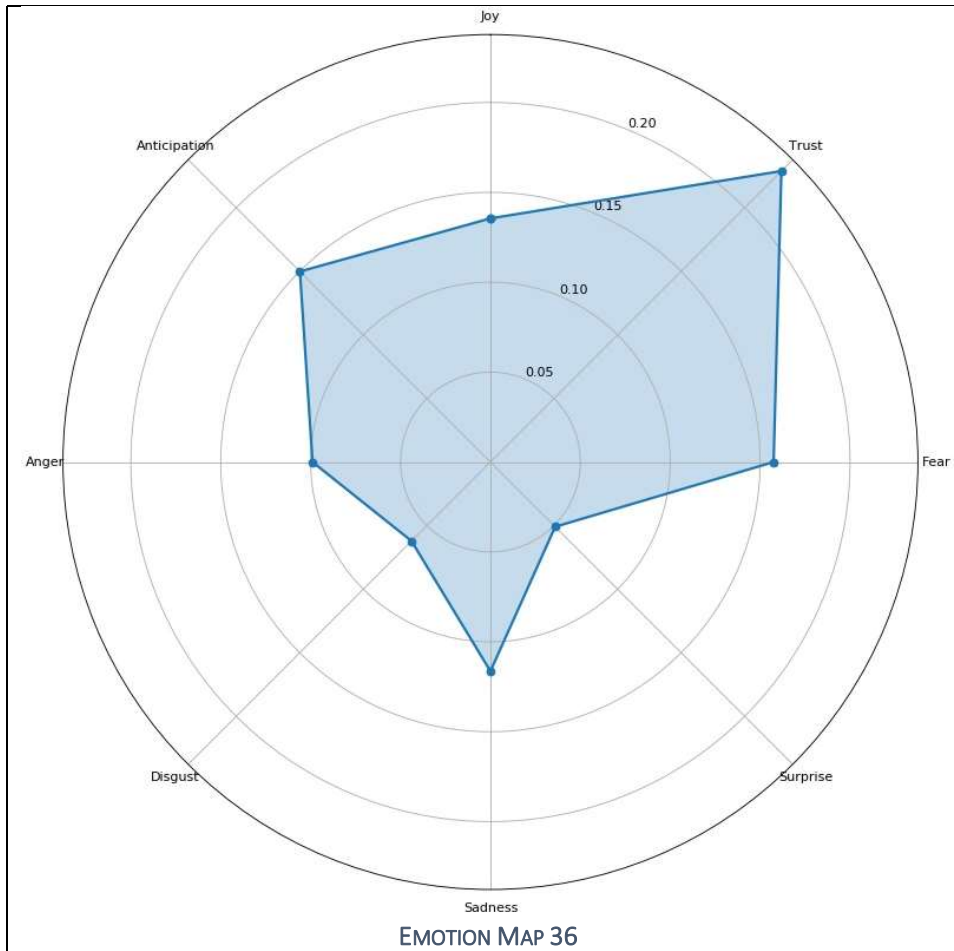






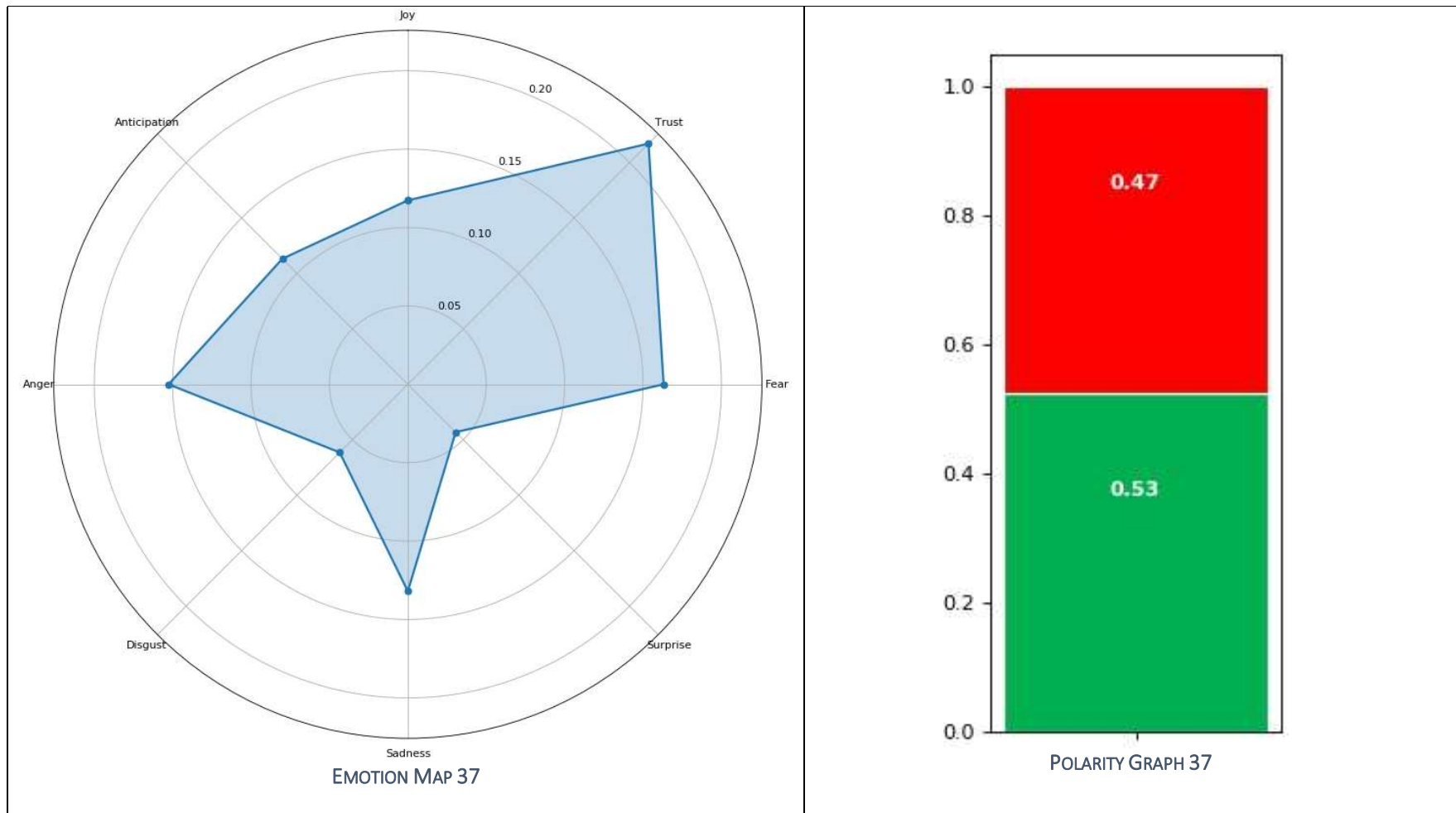


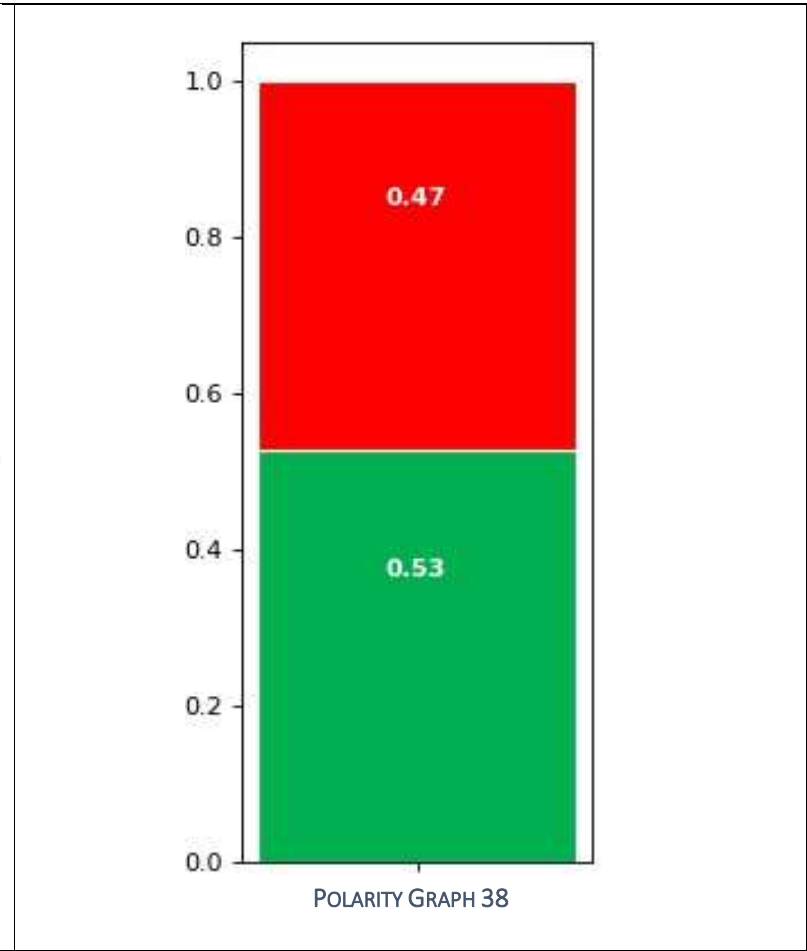
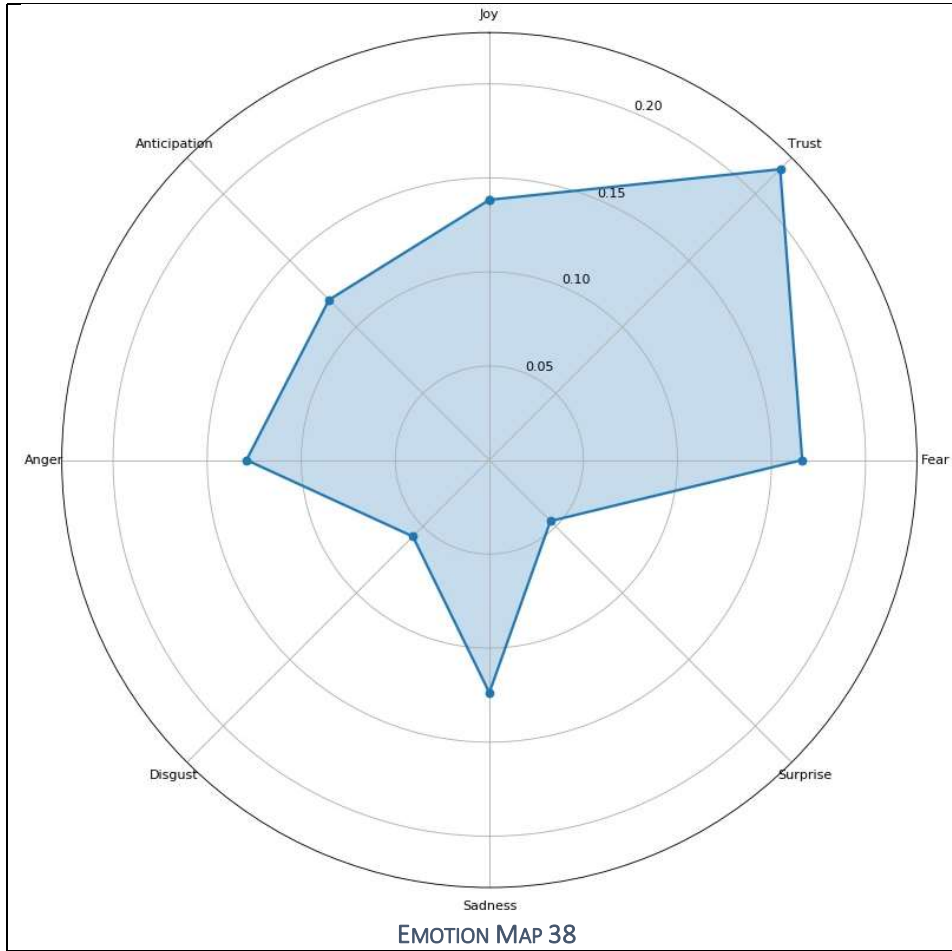


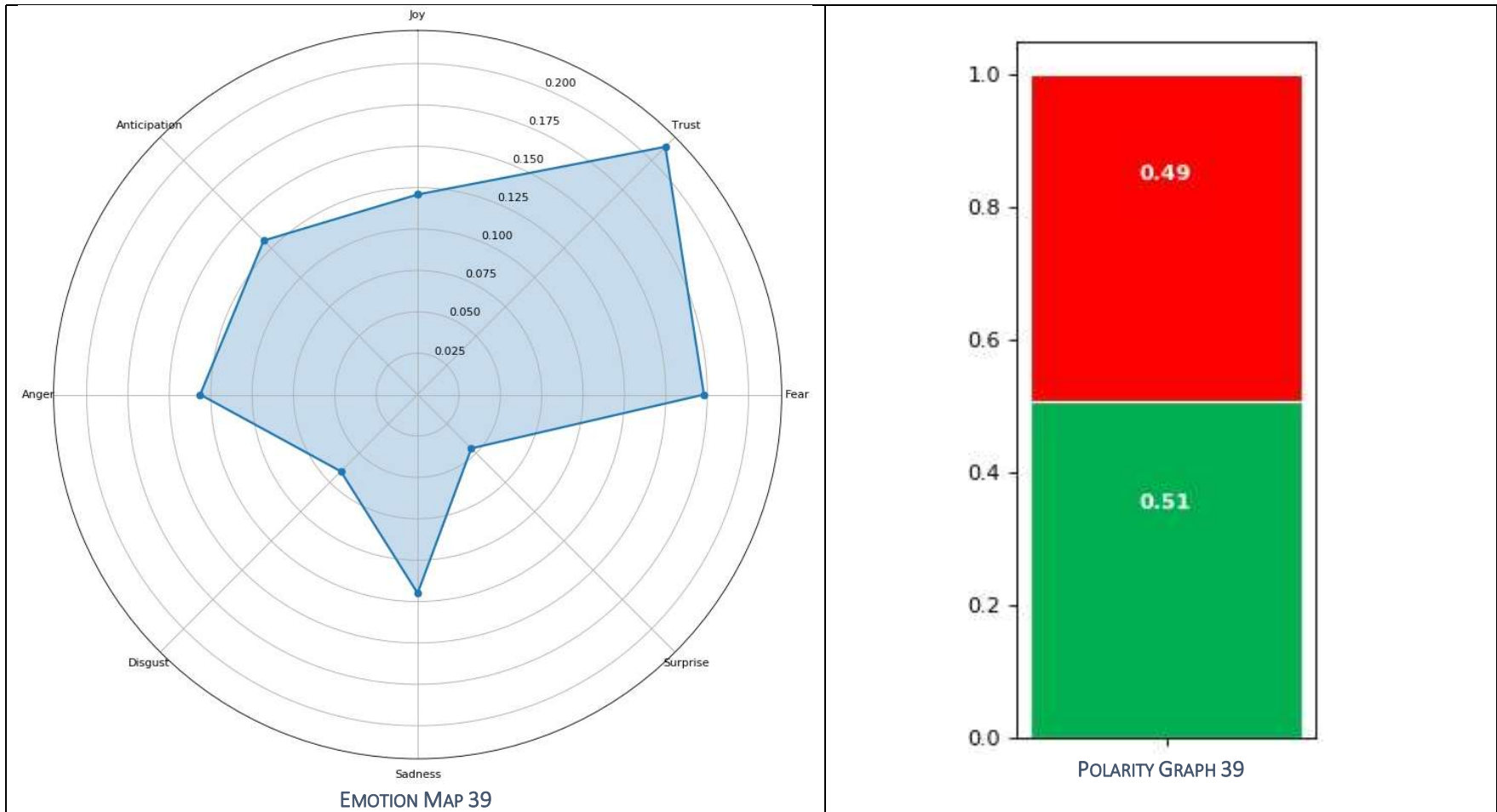


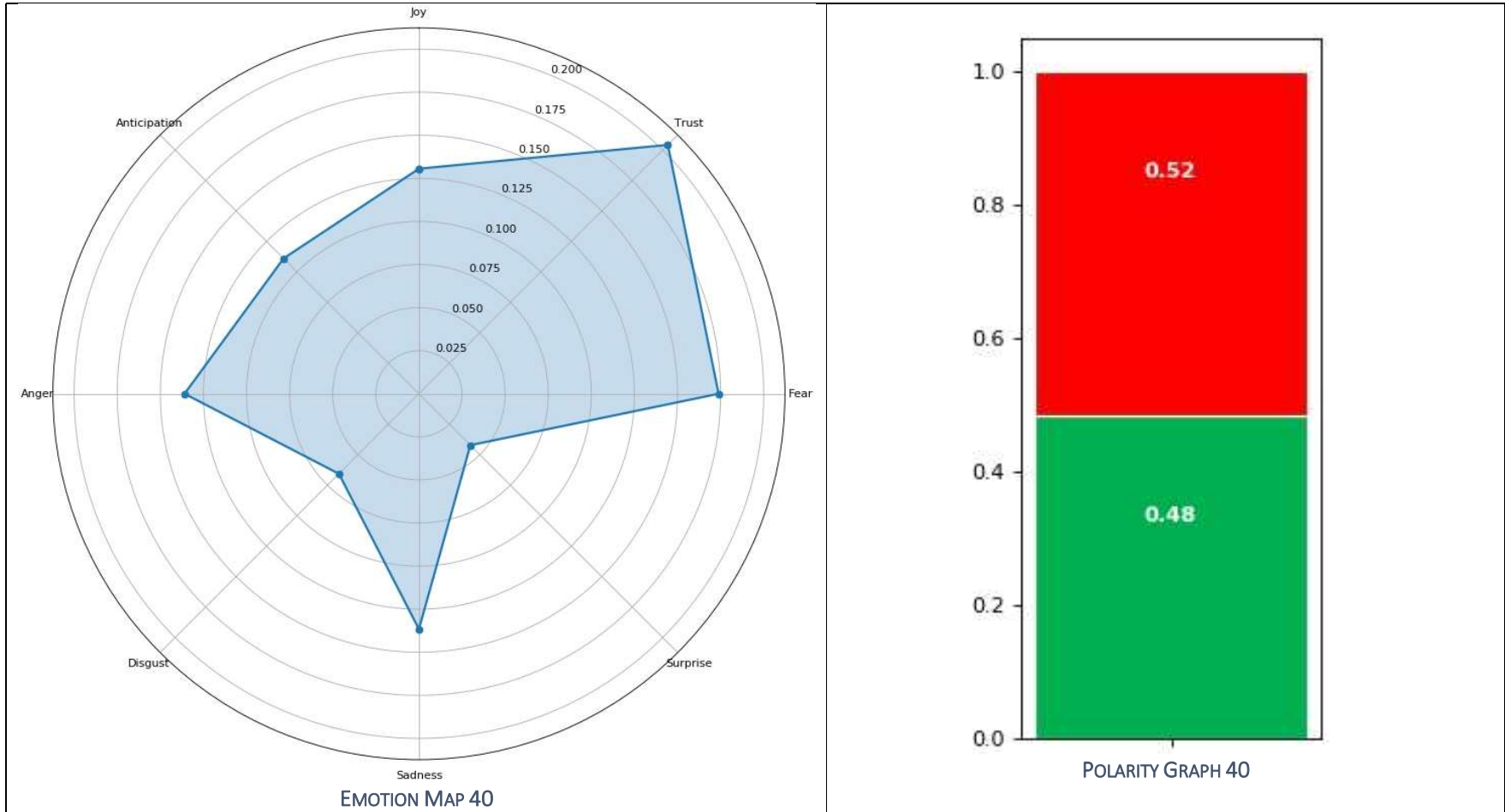
13.5Rumiyah

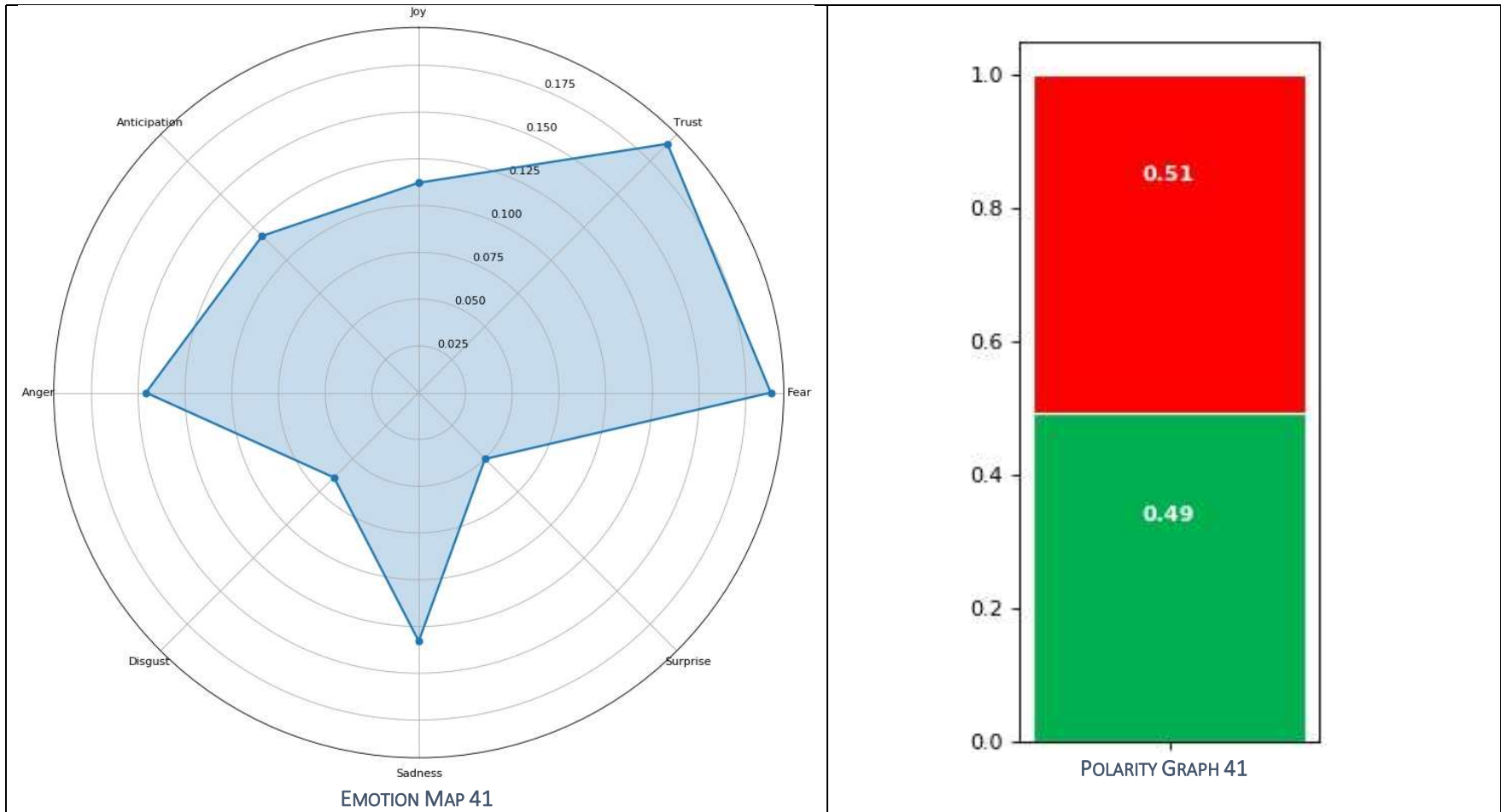
- Issue 1

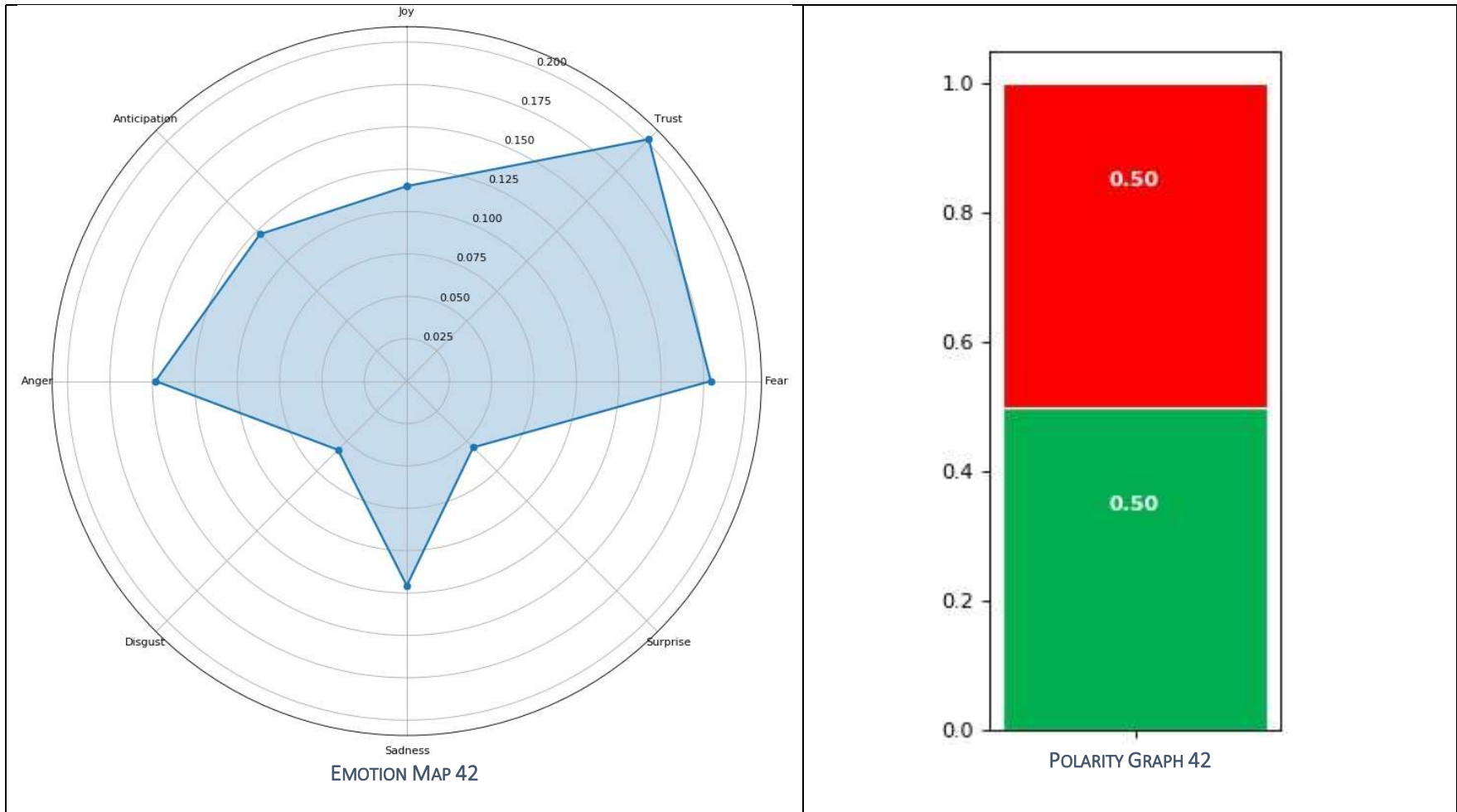


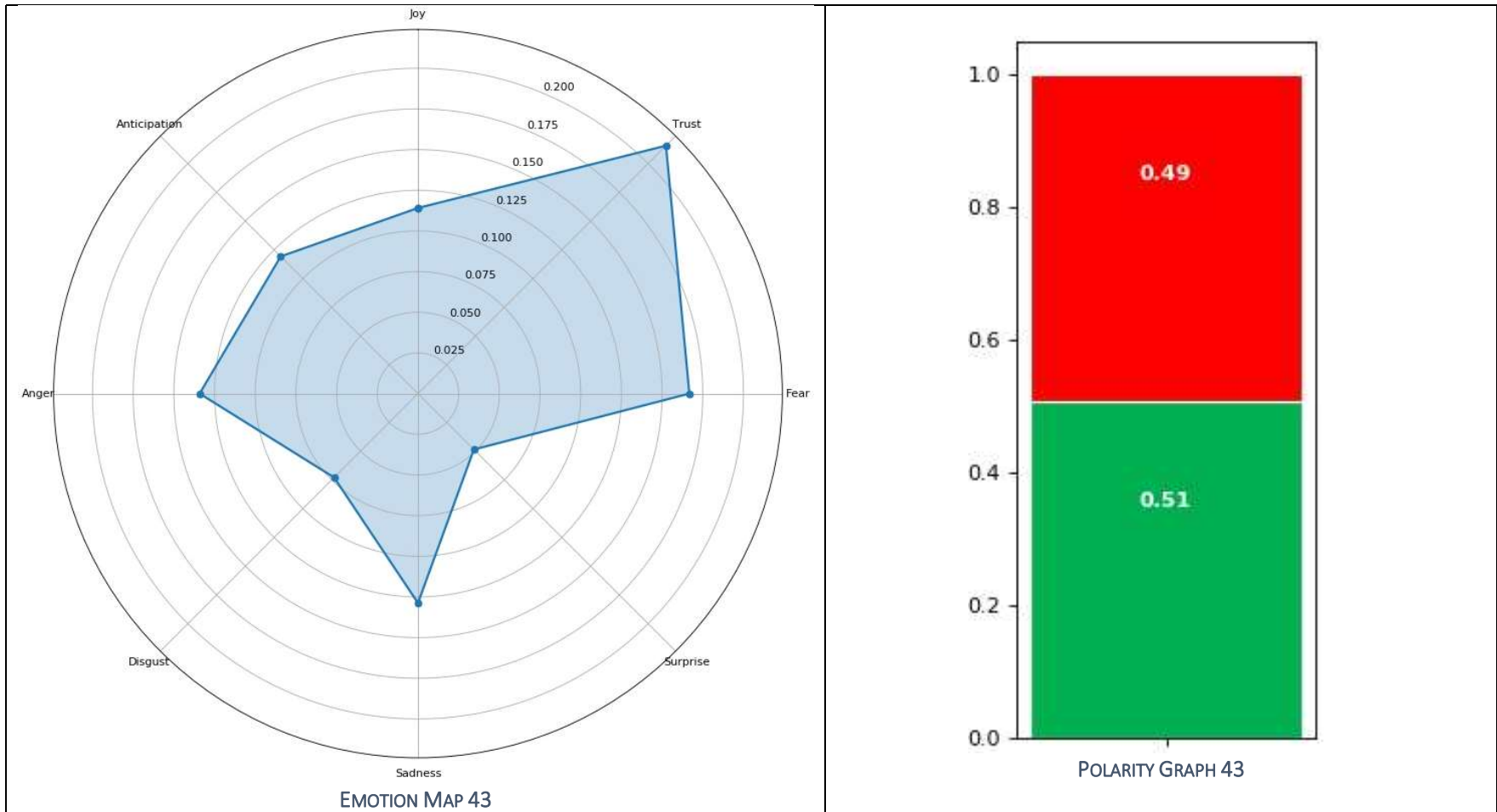


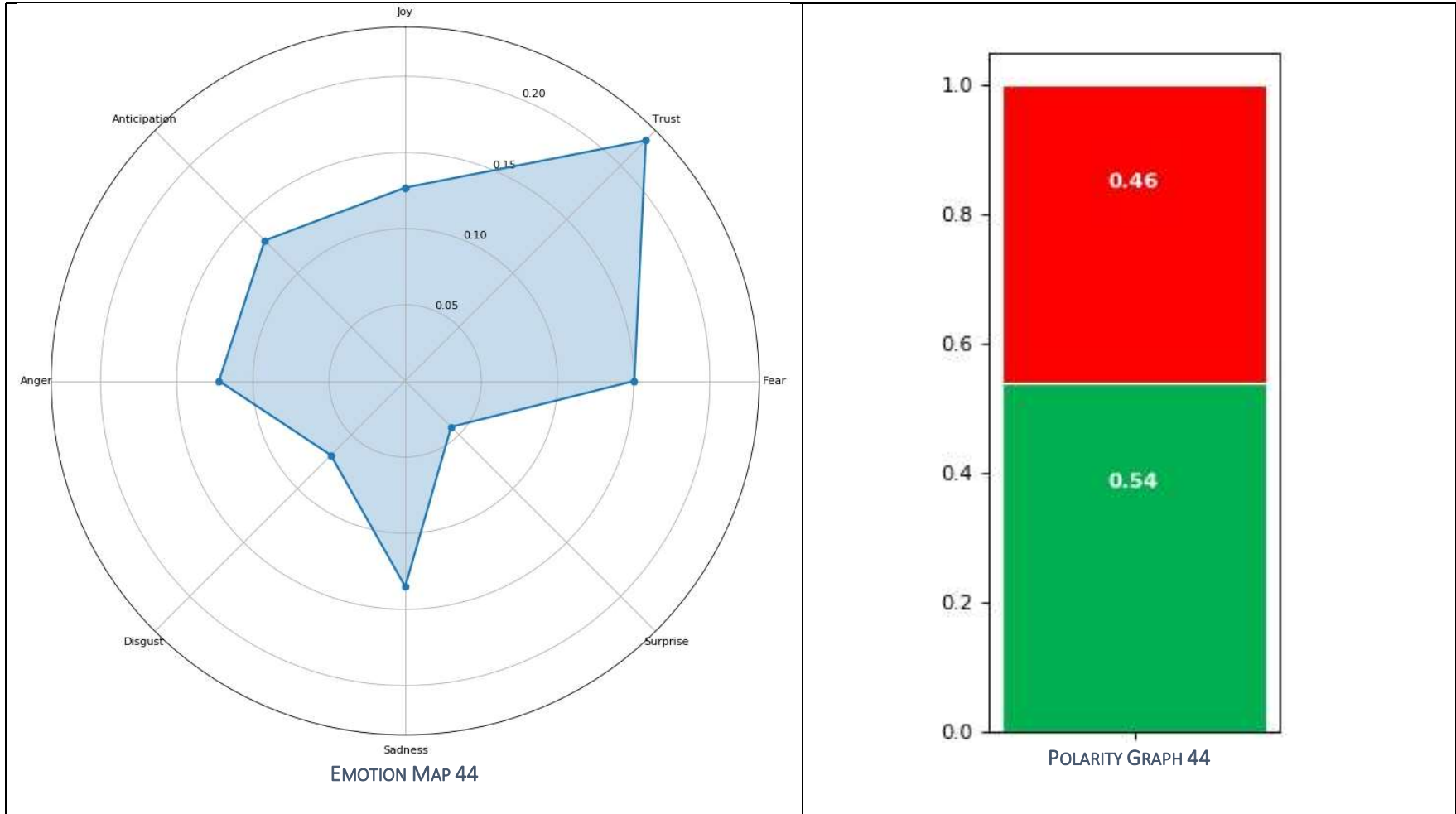


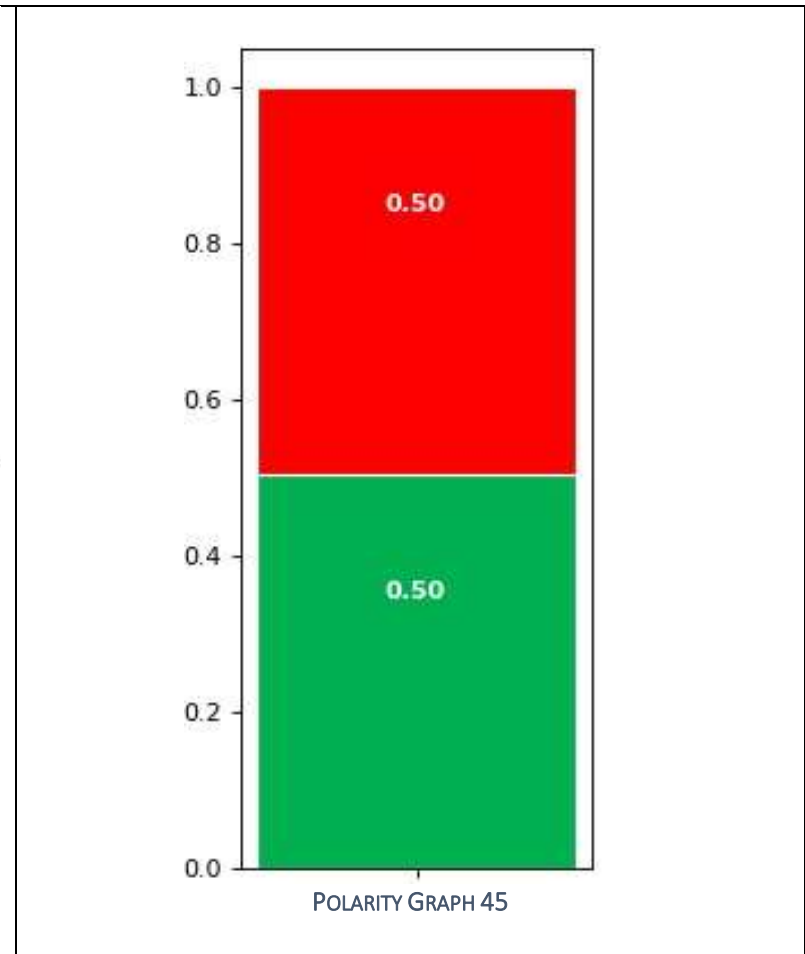
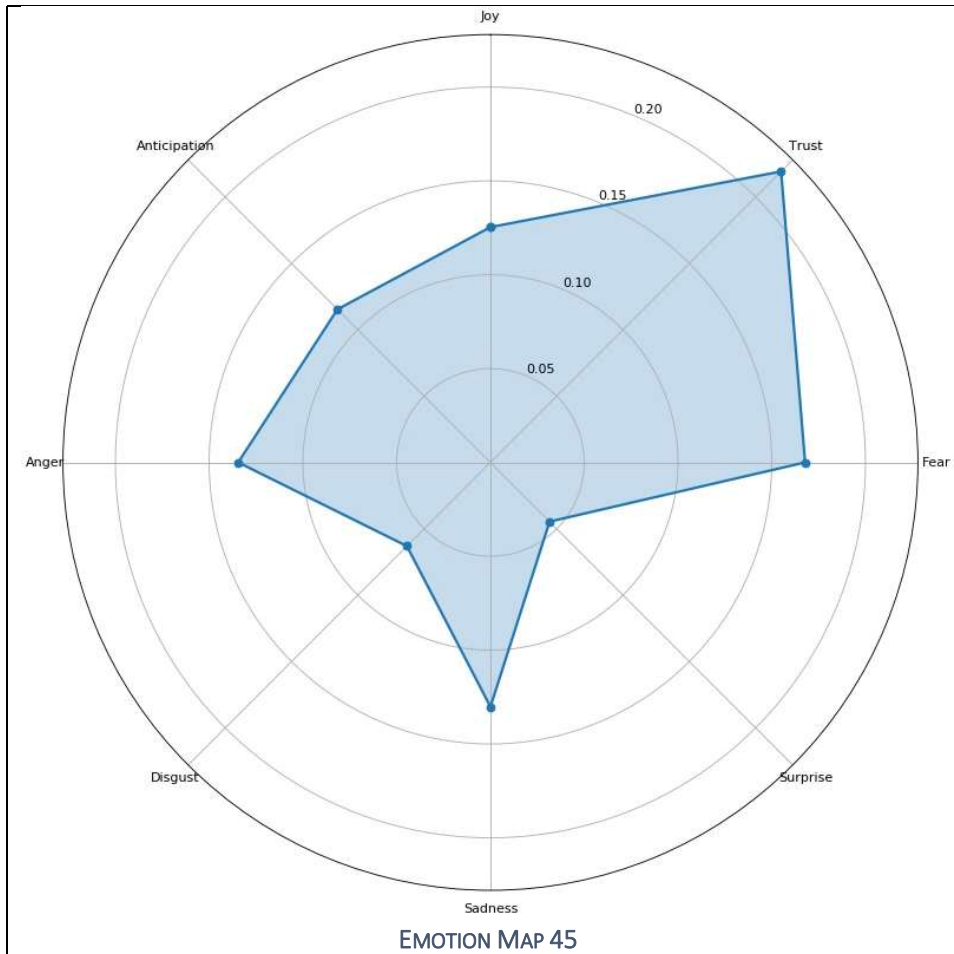


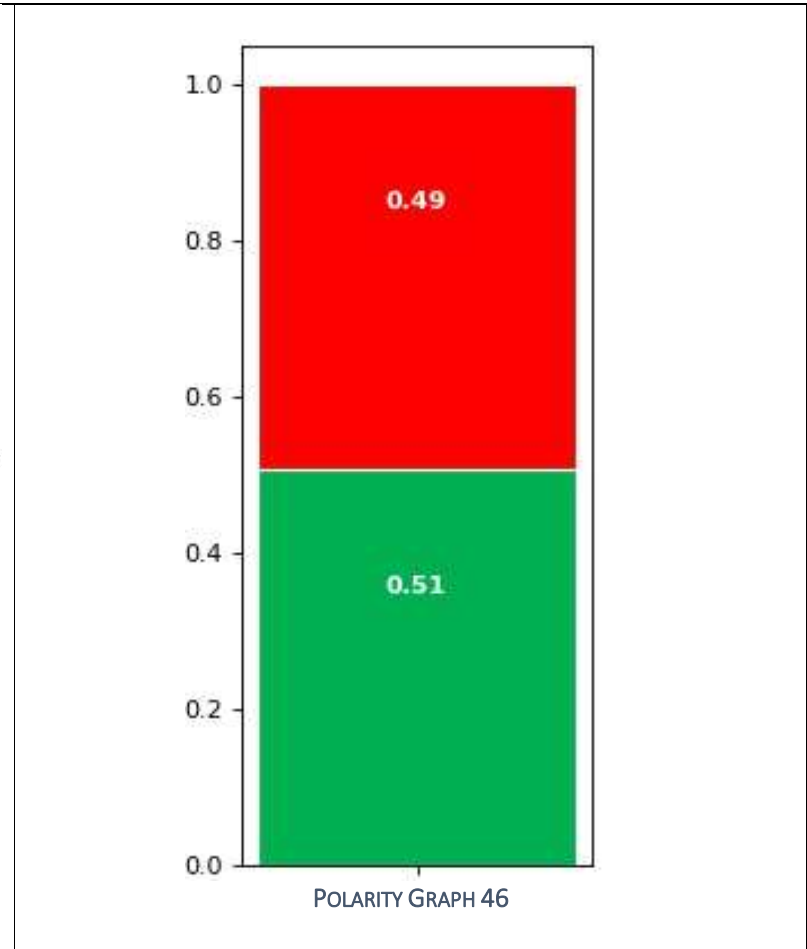
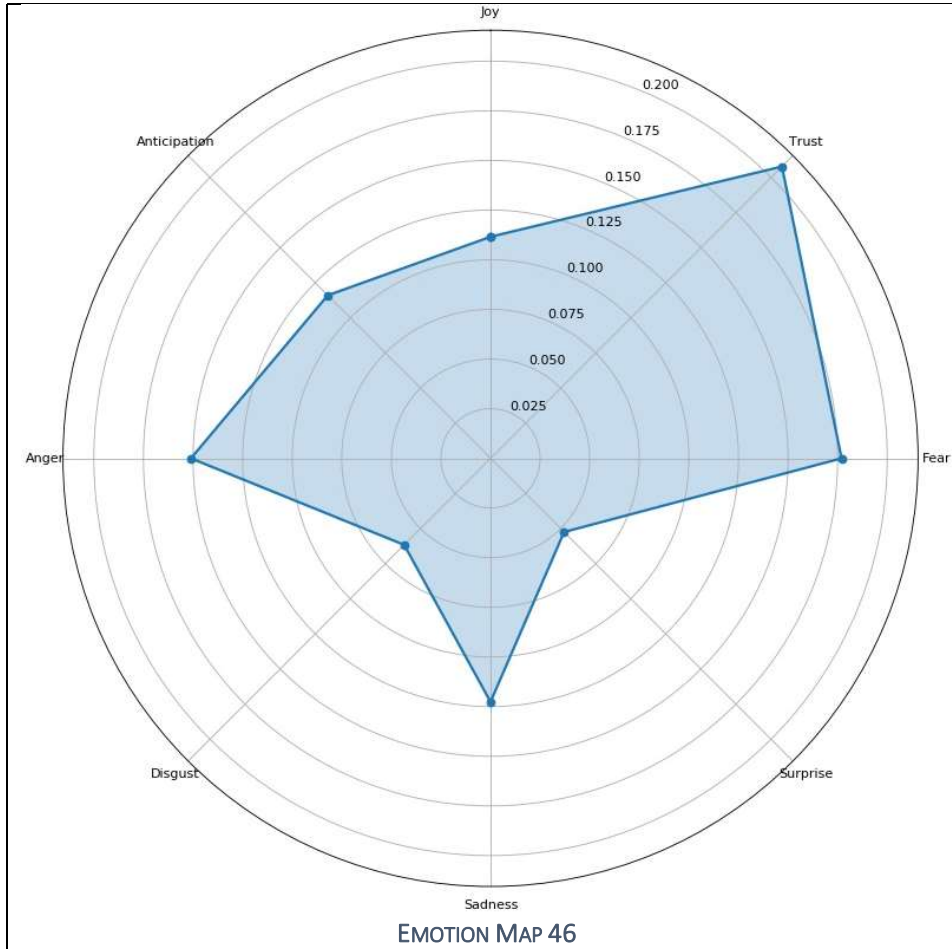


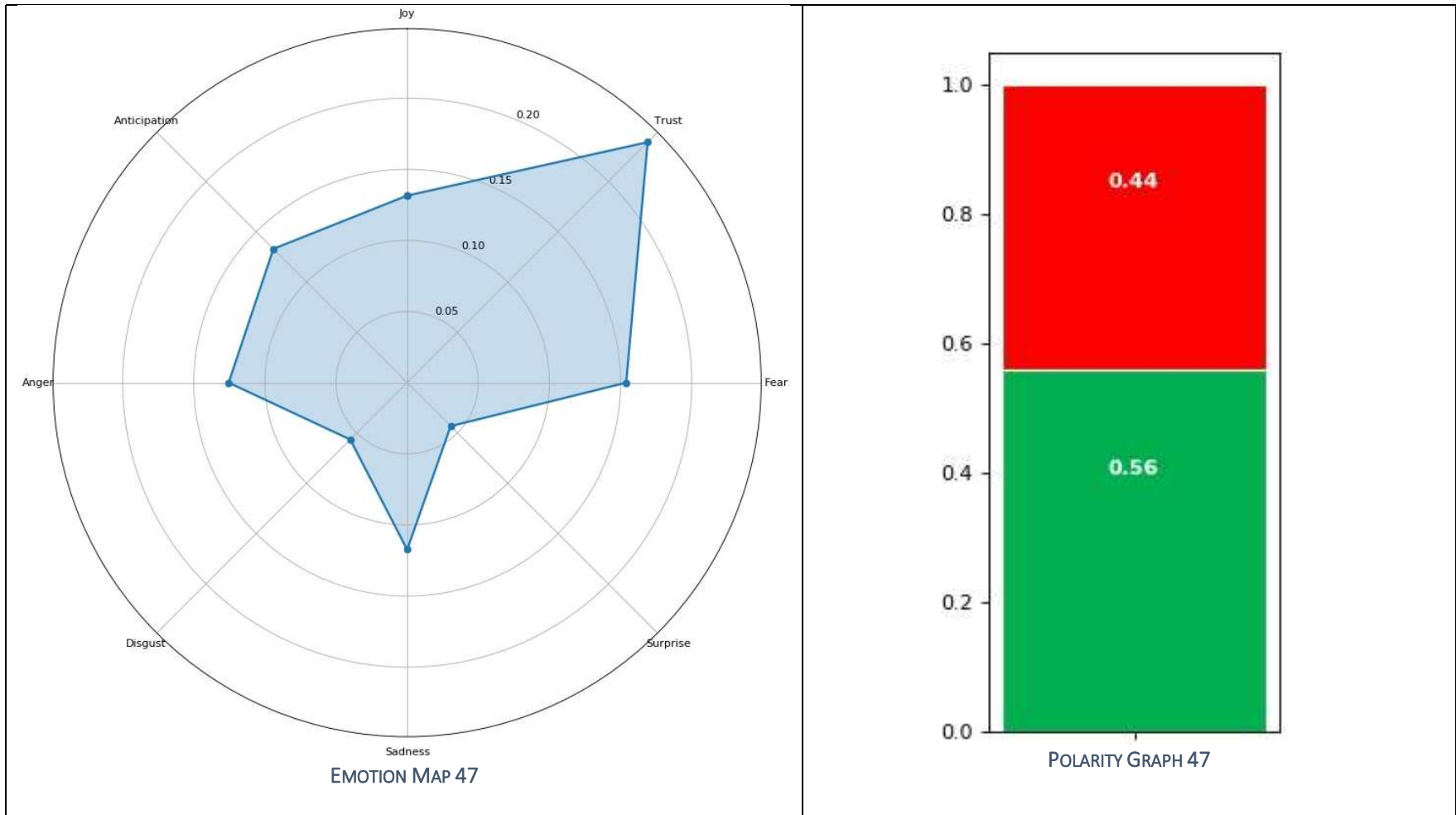


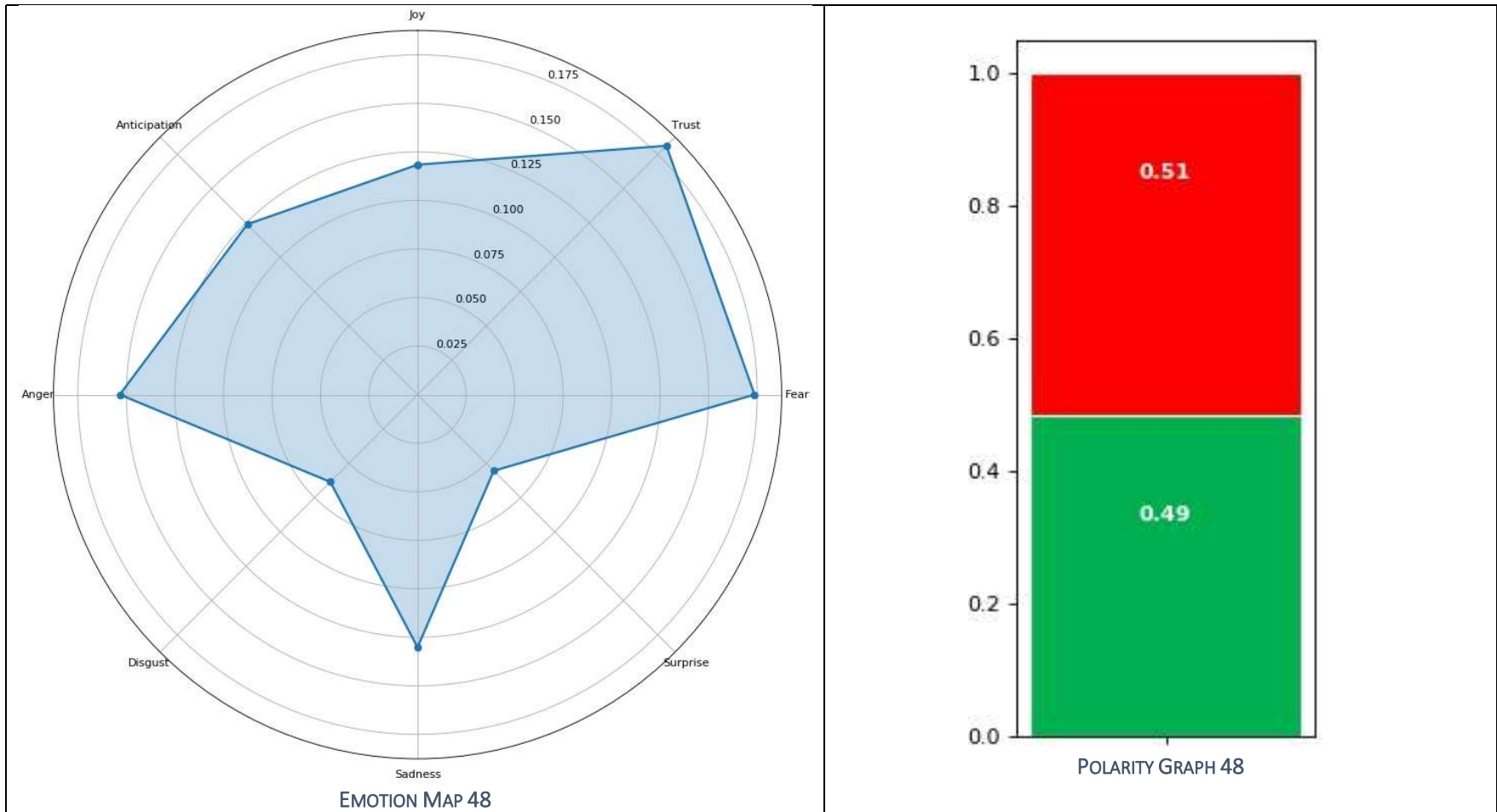


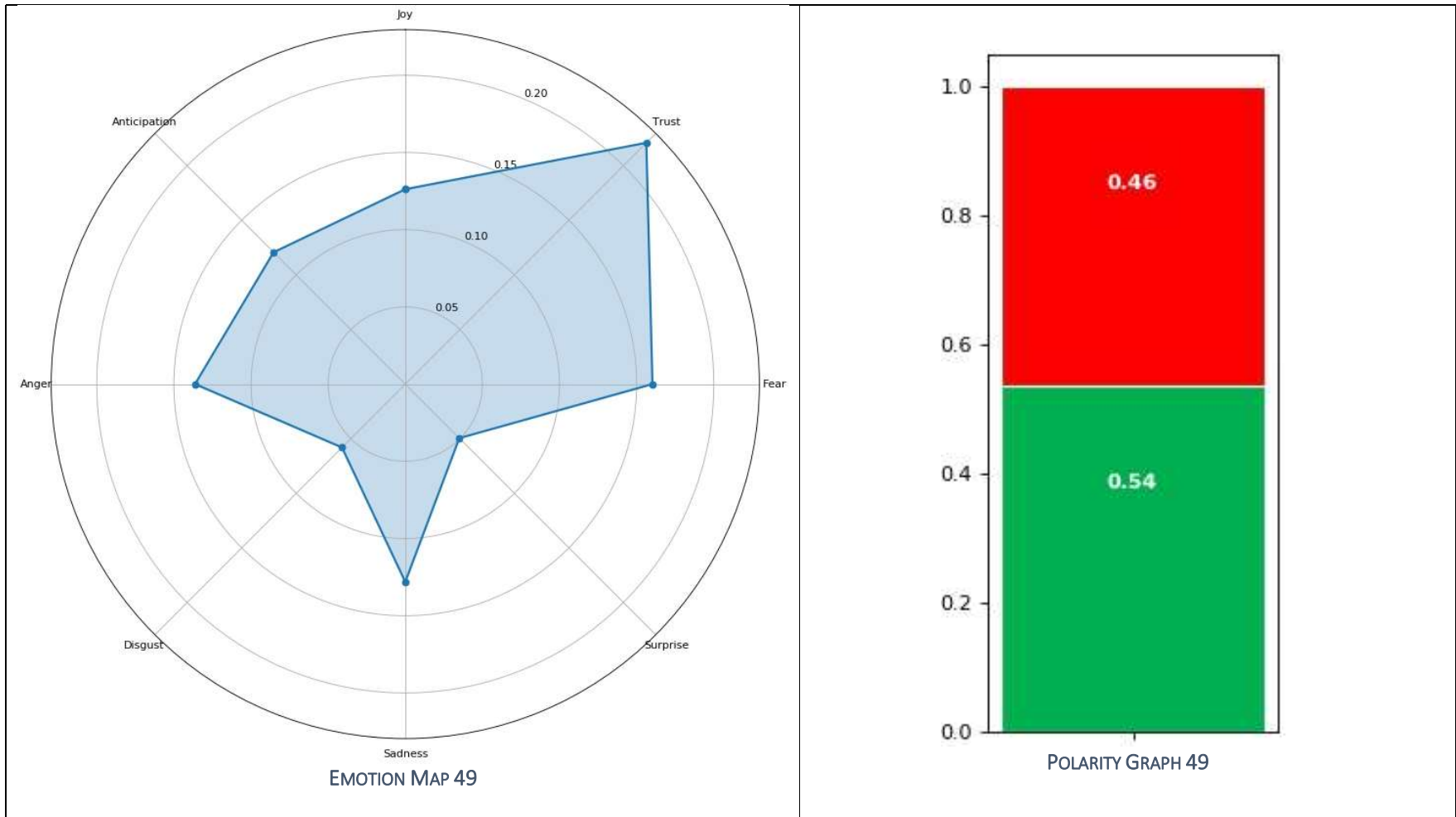












14 Παράρτημα Z – Επαναλήψεις της LDA. Προσδιοριζόμενες λέξεις ανά επανάληψη.

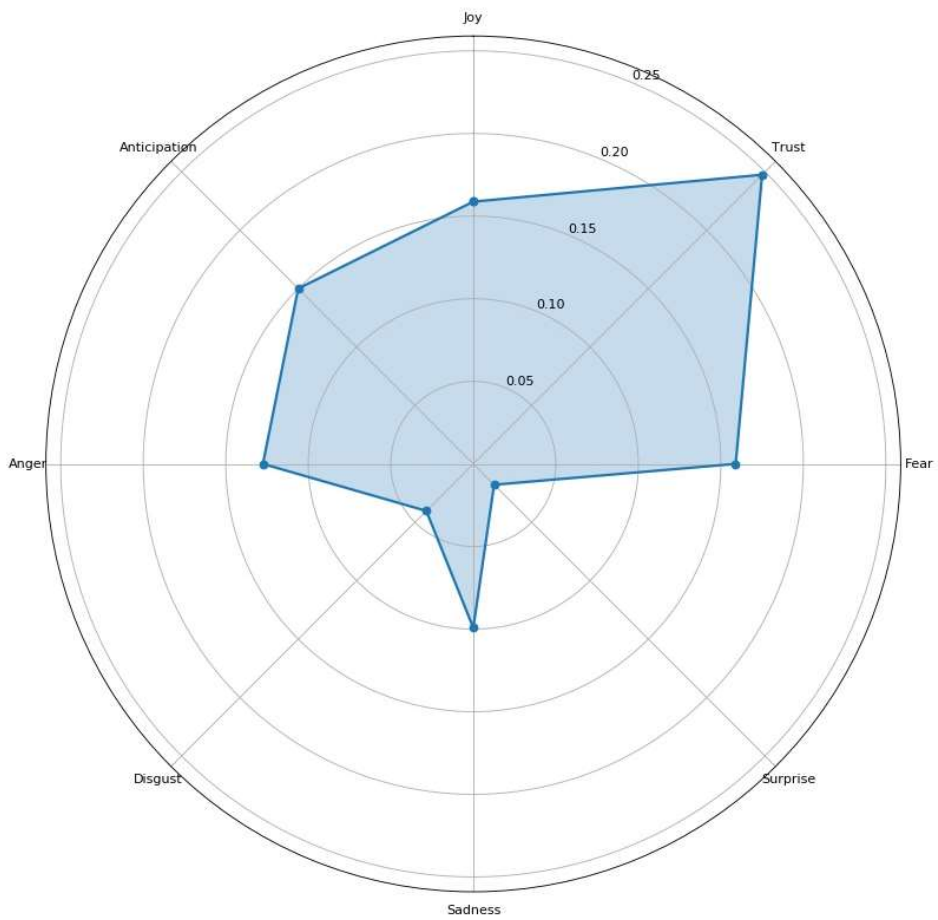
14.11^η επανάληψη της LDA από 10

TABLE 14-1. ΠΙΝΑΚΑΣ ΠΡΟΣΔΙΟΡΙΖΟΜΕΝΩΝ ΛΕΞΕΩΝ ΑΝΑ ΘΕΜΑ ΑΠΟ ΤΗΝ LDA (ΕΠΑΝΑΛΗΨΗ: 1^η ΑΠΟ 10)

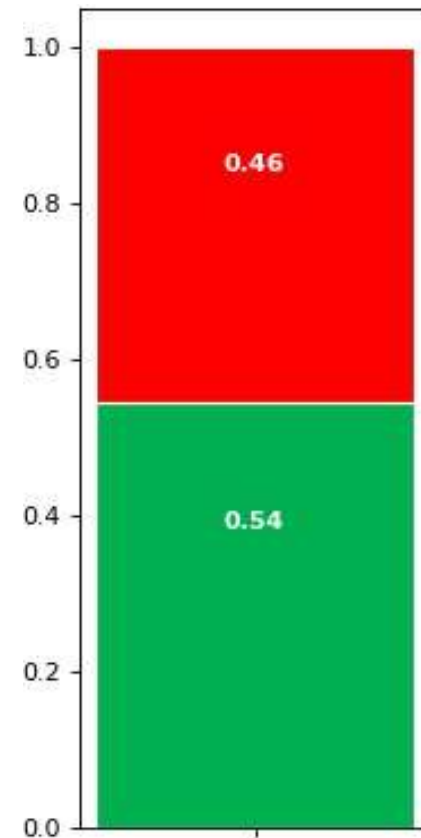
	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05	Topic # 06	Topic # 07	Topic # 08	Topic # 09	Topic # 10	Topic # 11	Topic # 12
0	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah
1	muslim	state	muslim	muslim	muslim	people	muslim	muslim	people	muslim	state	muslim
2	state	muslim	people	state	people	muslim	state	state	state	state	islam	state
3	islamic	islamic	islamic	people	state	islamic	islamic	islamic	islamic	people	people	people
4	people	people	state	east	islamic	make	people	people	muslim	islamic	give	god
5	give	islam	religion	islamic	take	state	east	born	religion	islam	time	make
6	god	east	islam	god	make	east	islam	east	give	religion	muslim	islamic
7	make	born	messenger	take	war	know	god	religion	islam	take	islamic	east
8	islam	make	time	time	brother	religion	make	islam	make	make	god	born
9	time	religion	make	soldier	religion	islam	born	make	fight	become	come	islam
10	good	messenger	know	islam	time	kill	messenger	jihad	kill	crusader	know	religion
11	know	take	soldier	prophet	jihad	report	know	take	messenger	kill	prophet	prophet
12	religion	know	kill	make	kill	god	prophet	god	know	god	religion	woman
13	born	time	day	come	messenger	take	religion	time	time	report	make	messenger
14	jihad	kill	take	give	fight	brother	give	come	take	time	day	know
15	take	god	come	born	crusader	messenger	good	day	come	soldier	kill	take
16	day	good	give	jihad	prophet	war	take	know	god	prophet	even	sure
17	fight	report	fight	know	east	time	sure	kill	khilafah	land	messenger	give
18	war	come	jihad	religion	come	khilafah	time	messenger	jihad	come	jihad	go
19	even	soldier	prophet	kill	day	among	report	son	war	day	soldier	soldier
20	soldier	prophet	east	even	born	jihad	kill	prophet	born	fight	go	war
21	messenger	sure	land	indeed	give	land	jihad	war	good	messenger	take	son
22	come	war	crusader	messenger	islam	come	war	land	day	east	call	jihad
23	law	go	brother	war	good	soldier	day	report	east	lord	indeed	time
24	crusader	day	god	good	report	fight	son	brother	land	brother	brother	come
25	go	become	enemy	brother	even	give	among	fight	battle	know	allahin	report
26	east	jihad	report	sure	except	sure	soldier	sure	prophet	go	way	fight
27	group	give	born	son	lord	born	order	become	soldier	even	war	day
28	right	order	go	among	call	like	many	even	crusader	war	good	brother
29	kill	enemy	way	become	army	prophet	khilafah	many	brother	jihad	cent	good

Πλήθος μοναδικών λέξεων: 53

Λίστα των λέξεων με τυχαία σειρά: cent, law, islamic, god, indeed, allah, group, religion, war, call, except, muslim, among, born, battle, prophet, messenger, east, way, make, khilafah, give, come, people, order, fight, become, jihad, take, islam, report, lord, enemy, kill, woman, state, even, allahin, day, soldier, right, army, crusader, son, time, brother, sure, like, good, go, know, land, many



RADAR PLOT 1 – EMOTION MAP FOR LDA ITERATION 1 OF 10



BAR PLOT 1 – POLARITY SENTIMENT FOR LDA ITERATION 1 OF 10

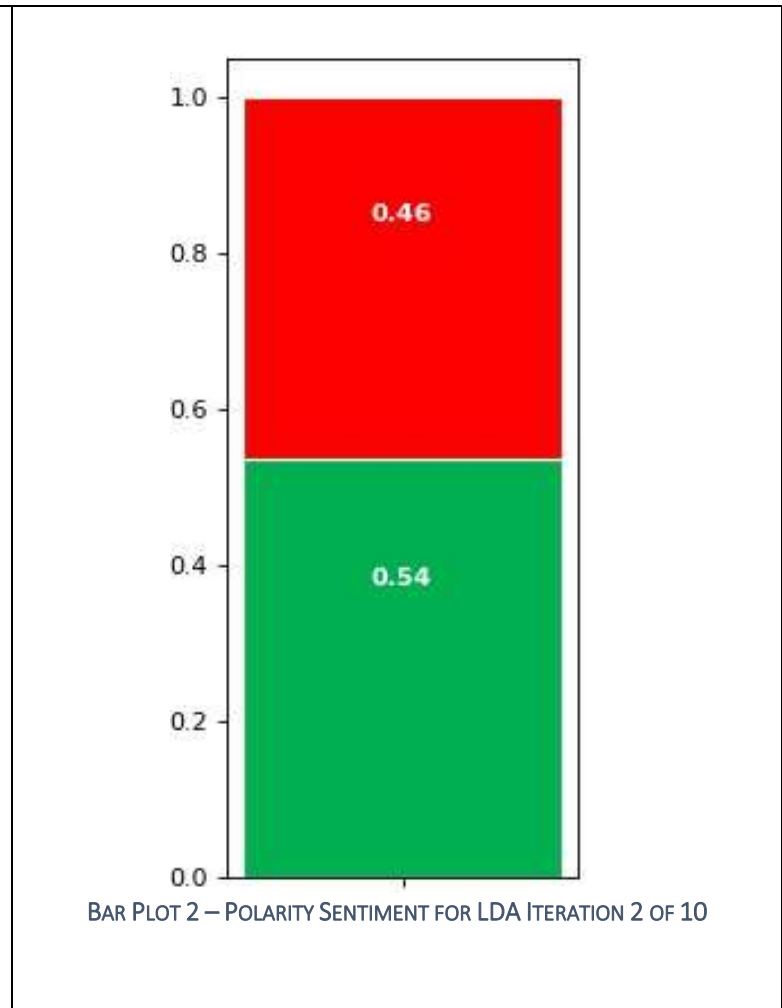
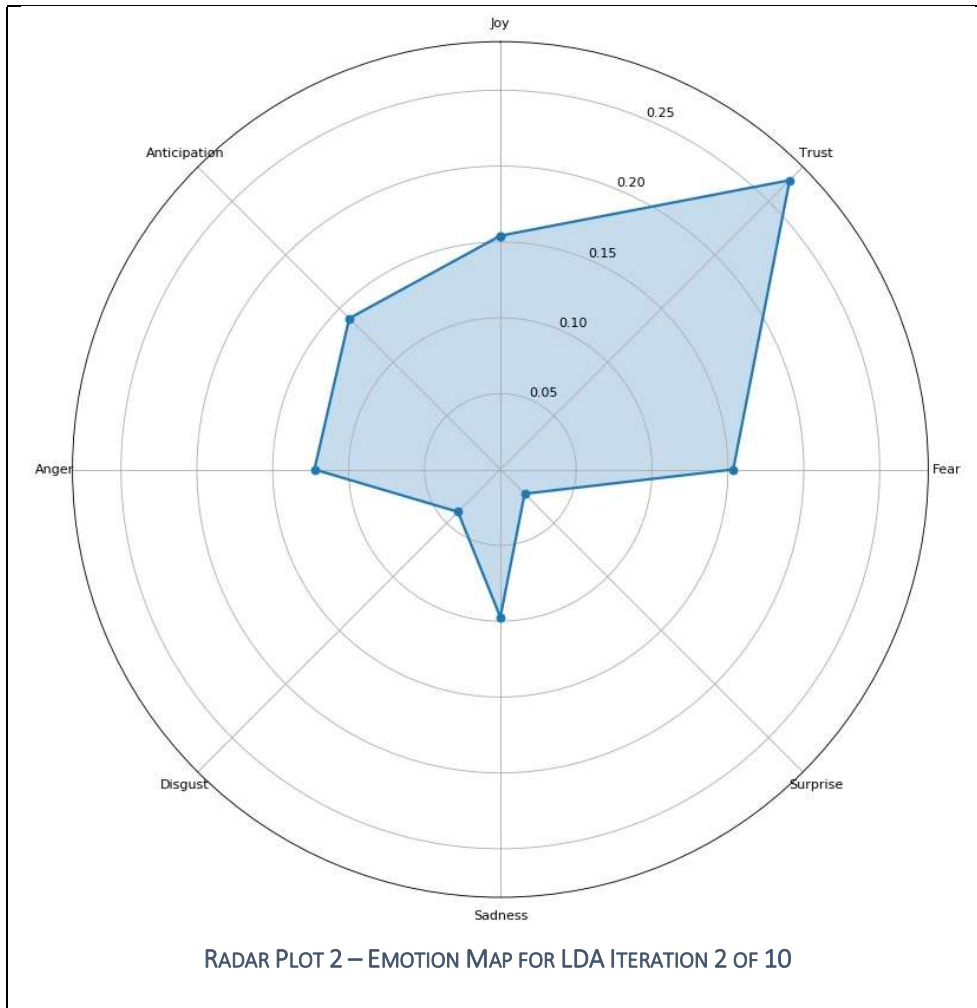
14.22^η επανάληψη της LDA από 10

TABLE 14-2. ΠΙΝΑΚΑΣ ΠΡΟΣΔΙΟΡΙΖΟΜΕΝΩΝ ΛΕΞΕΩΝ ΑΝΑ ΘΕΜΑ ΑΠΟ ΤΗΝ LDA (ΕΠΑΝΑΛΗΨΗ: 2Η ΑΠΟ 10)

	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05	Topic # 06	Topic # 07	Topic # 08	Topic # 09	Topic # 10	Topic # 11	Topic # 12
0	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah
1	state	muslim	people	muslim	people	muslim	islamic	muslim	muslim	state	muslim	muslim
2	islamic	people	muslim	people	muslim	state	state	people	east	people	people	state
3	muslim	islamic	islamic	state	state	people	people	state	people	muslim	islamic	people
4	islam	state	state	islamic	islamic	islam	make	make	islamic	islam	state	islamic
5	make	god	know	east	religion	god	muslim	islamic	god	islamic	religion	give
6	people	islam	religion	born	islam	religion	islam	fight	born	religion	time	take
7	give	make	make	make	make	time	give	religion	state	born	kill	religion
8	take	east	come	god	take	islamic	religion	islam	son	east	soldier	make
9	religion	kill	take	islam	prophet	take	brother	take	make	know	islam	god
10	god	messenger	report	know	time	know	come	come	islam	time	make	messenger
11	know	know	good	messenger	messenger	make	crusader	jihad	sure	make	east	jihad
12	come	time	soldier	give	kill	jihad	kill	know	time	war	take	soldier
13	time	religion	messenger	time	jihad	born	take	east	know	messenger	messenger	even
14	soldier	good	fight	sure	day	kill	know	day	take	god	give	islam
15	messenger	born	east	prophet	land	east	jihad	messenger	give	day	prophet	time
16	prophet	give	time	come	god	day	time	go	religion	give	jihad	kill
17	way	war	prophet	good	crusader	brother	become	time	report	prophet	born	report
18	war	come	kill	kill	soldier	come	god	even	plus	jihad	day	land
19	become	fight	born	soldier	enemy	messenger	soldier	prophet	good	land	good	brother
20	east	soldier	war	take	east	give	land	war	kill	kill	khilafah	come
21	many	sure	go	group	give	war	indeed	crusader	day	brother	god	among
22	fight	prophet	islam	religion	know	fight	among	good	brother	report	come	know
23	kill	jihad	among	jihad	come	crusader	day	report	jihad	come	report	fight
24	born	land	jihad	even	report	prophet	battle	born	fact	sure	sure	east
25	report	report	even	son	woman	khilafah	messenger	battle	honor	even	even	prophet
26	lord	take	god	brother	khilafah	even	woman	give	messenger	crusader	war	war
27	son	even	crusader	war	war	report	go	sure	become	good	among	order
28	brother	day	day	become	brother	except	fight	land	call	son	go	day
29	order	brother	become	land	indeed	go	place	order	come	take	group	indeed

Πλήθος μοναδικών λέξεων: 51

Λίστα μοναδικών λέξεων σε τυχαία σειρά: islamic, god, indeed, allah, group, religion, war, call, except, muslim, among, born, battle, prophet, messenger, east, way, make, khilafah, give, come, people, order, become, fight, jihad, honor, take, islam, report, lord, enemy, kill, woman, state, even, day, soldier, crusader, son, time, brother, place, fact, plus, sure, good, go, know, land, many



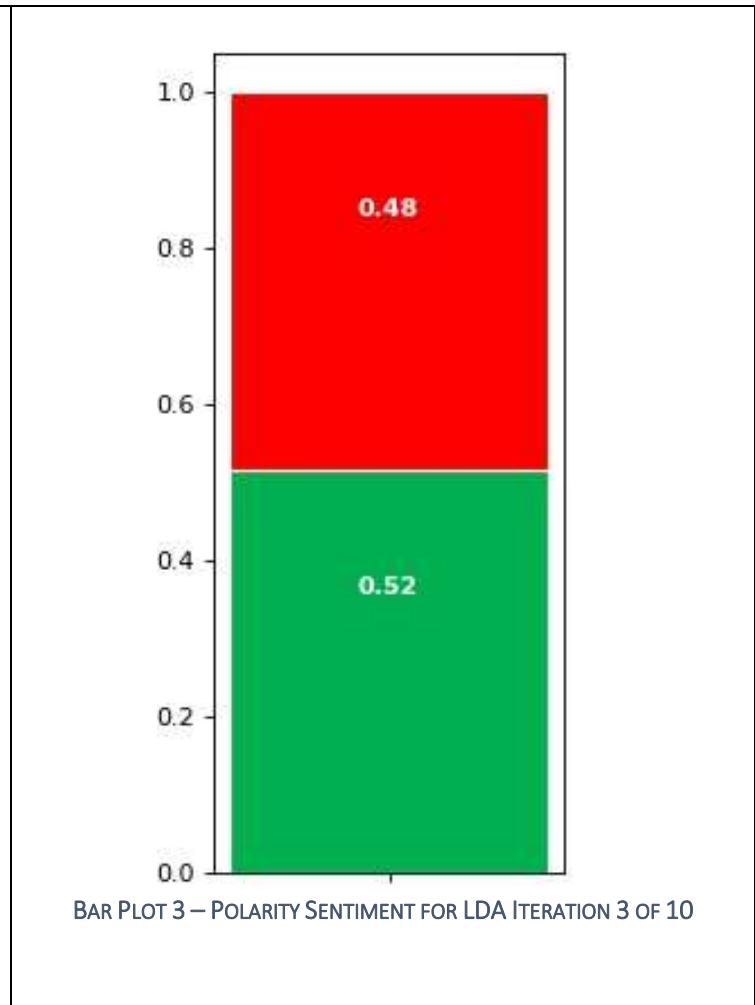
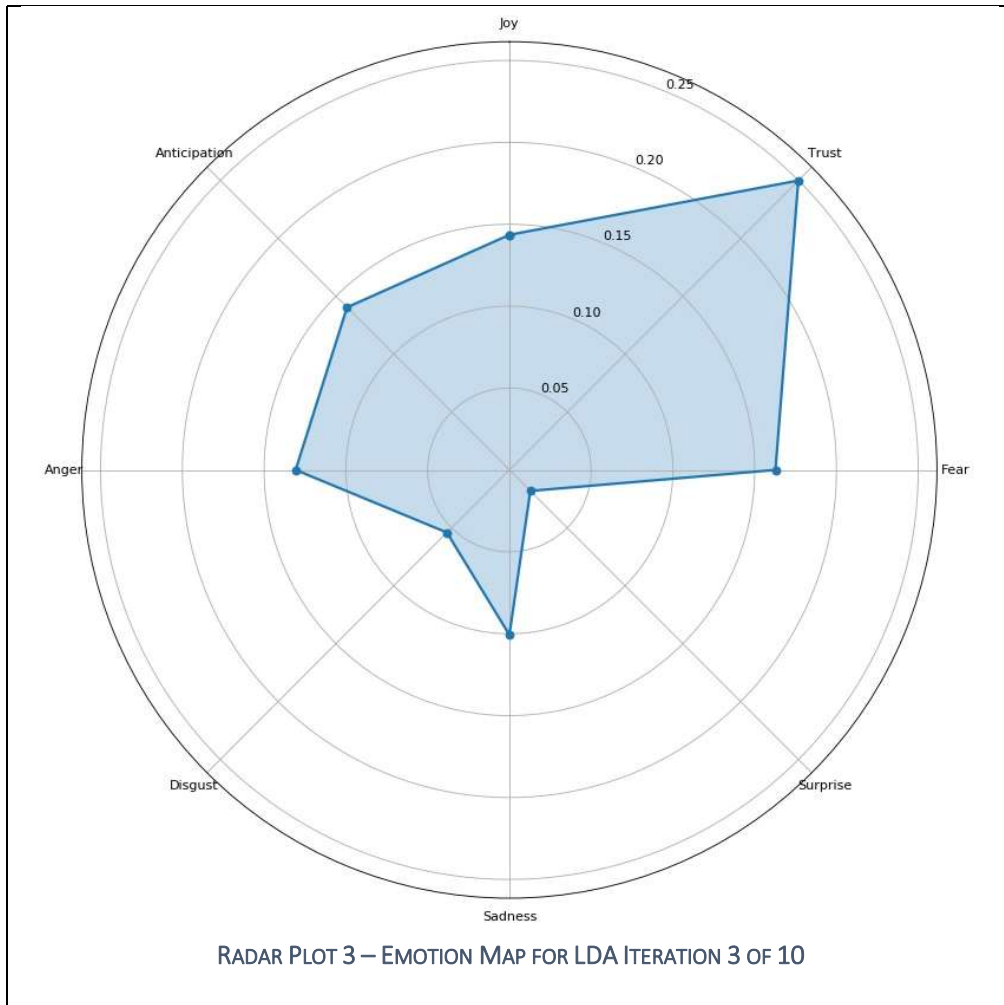
14.33^η επανάληψη της LDA από 10

TABLE 14-3. ΠΙΝΑΚΑΣ ΠΡΟΣΔΙΟΡΙΖΟΜΕΝΩΝ ΛΕΞΕΩΝ ΑΝΑ ΘΕΜΑ ΑΠΟ ΤΗΝ LDA (ΕΠΑΝΑΛΗΨΗ: 3^Η ΑΠΟ 10)

	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05	Topic # 06	Topic # 07	Topic # 08	Topic # 09	Topic # 10	Topic # 11	Topic # 12
0	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah
1	state	people	muslim	muslim	muslim	state	muslim	state	state	state	muslim	muslim
2	muslim	muslim	people	people	people	people	state	muslim	muslim	muslim	people	east
3	people	state	state	islamic	islamic	muslim	people	islamic	give	islamic	east	state
4	islamic	make	islamic	make	state	islamic	islamic	people	islam	people	islamic	islamic
5	make	god	religion	born	god	make	east	islam	islamic	islam	born	people
6	islam	islam	make	east	time	religion	religion	know	god	religion	state	born
7	jihad	east	come	state	islam	islam	god	religion	know	make	time	god
8	time	islamic	prophet	sure	religion	soldier	time	make	people	take	make	religion
9	good	give	islam	war	messenger	messenger	know	crusader	east	messenger	messenger	jihad
10	give	time	time	take	make	jihad	make	kill	come	know	islam	sure
11	take	religion	give	know	east	kill	islam	take	make	come	god	islam
12	religion	take	fight	god	know	fight	born	jihad	religion	report	sure	kill
13	prophet	messenger	kill	religion	give	come	kill	land	born	give	take	take
14	brother	know	day	give	prophet	know	take	brother	time	prophet	know	give
15	crusader	born	land	report	war	take	soldier	fight	take	day	good	time
16	messenger	kill	god	islam	soldier	report	prophet	time	go	fight	prophet	know
17	soldier	soldier	jihad	kill	take	day	day	god	prophet	brother	religion	messenger
18	day	jihad	messenger	time	fight	give	jihad	messenger	good	time	son	good
19	god	come	know	good	born	time	come	day	war	soldier	kill	crusader
20	see	lord	take	come	come	khilafah	give	prophet	brother	kill	go	report
21	born	good	become	son	kill	prophet	sure	war	cent	indeed	group	group
22	kill	woman	even	day	jihad	war	report	good	jihad	god	come	son
23	know	even	born	soldier	even	battle	war	khilafah	day	even	soldier	come
24	enemy	war	soldier	messenger	report	land	son	enemy	even	among	give	war
25	become	report	enemy	even	go	become	among	become	land	call	among	order
26	lord	order	report	jihad	brother	even	many	soldier	son	war	jihad	make
27	come	way	brother	go	land	crusader	fight	come	kill	order	war	soldier
28	even	son	lord	fact	see	brother	force	indeed	way	enemy	day	become
29	khilafah	fight	place	woman	call	born	messenger	order	sure	mujahidin	brother	towards

Πλήθος μοναδικών λέξεων: 53

Λίστα μοναδικών λέξεων σε τυχαία σειρά: cent, islamic, god, indeed, allah, call, religion, war, mujahidin, group, muslim, battle, born, among, prophet, messenger, east, way, make, khilafah, give, come, people, order, become, fight, jihad, see, take, islam, report, lord, enemy, kill, woman, state, even, day, towards, soldier, crusader, son, time, brother, place, fact, sure, force, good, go, know, land, many



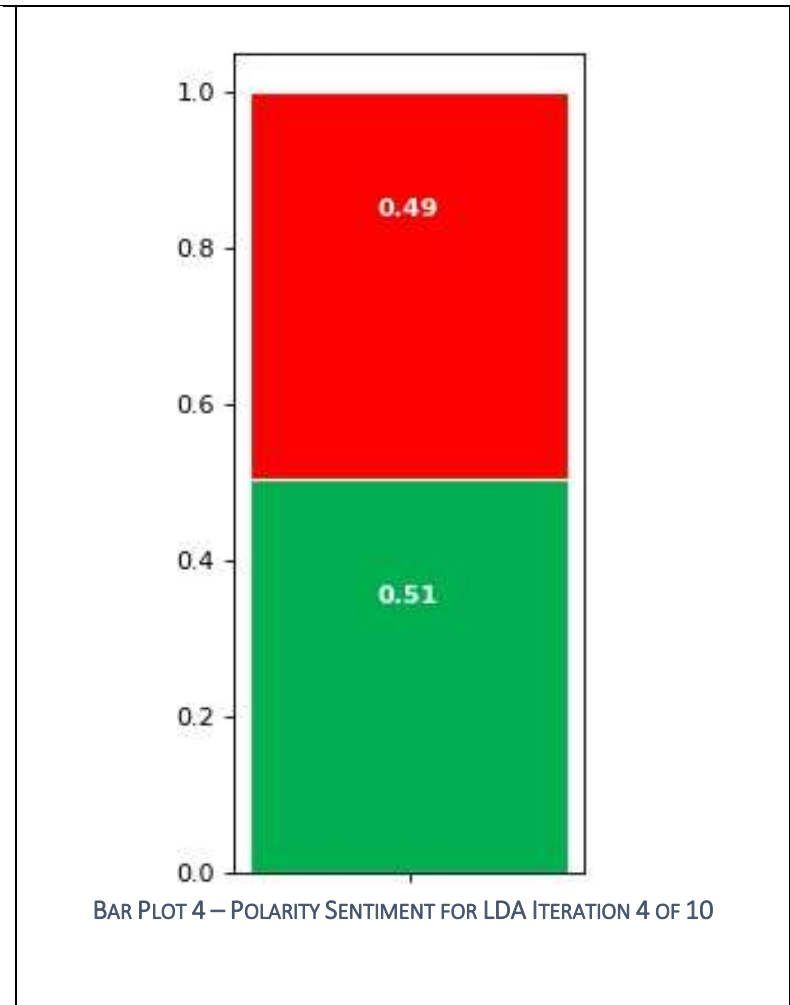
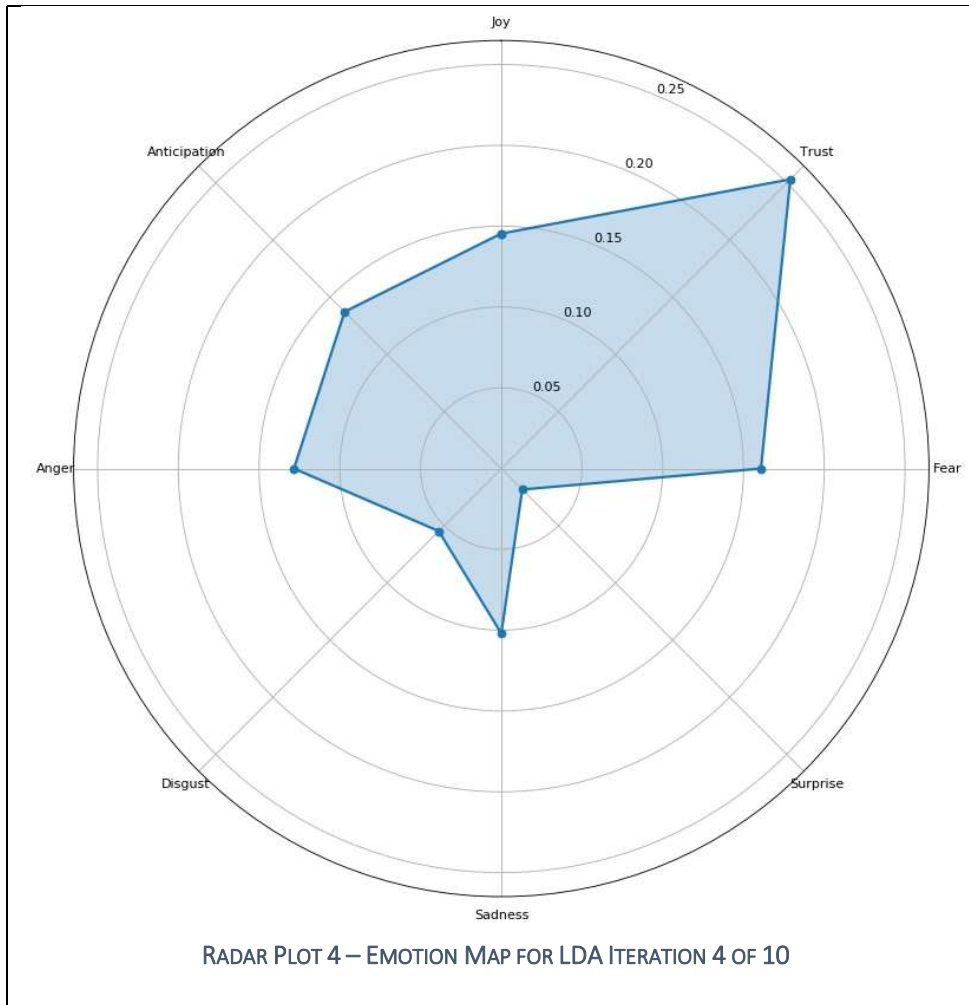
14.44ⁿ επανάληψη της LDA από 10

TABLE 14-4. ΠΙΝΑΚΑΣ ΠΡΟΣΔΙΟΡΙΖΟΜΕΝΩΝ ΛΕΞΕΩΝ ΑΝΑ ΘΕΜΑ ΑΠΟ ΤΗΝ LDA (ΕΠΑΝΑΛΗΨΗ: 4^η ΑΠΟ 10)

	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05	Topic # 06	Topic # 07	Topic # 08	Topic # 09	Topic # 10	Topic # 11	Topic # 12
0	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah
1	people	muslim	muslim	muslim	state	people	muslim	state	islamic	state	people	muslim
2	islamic	state	people	god	muslim	muslim	people	muslim	people	muslim	muslim	people
3	muslim	islamic	islamic	people	people	state	state	islamic	state	give	state	state
4	state	people	religion	east	islamic	religion	islamic	people	muslim	islamic	islam	islamic
5	god	religion	state	state	islam	islamic	east	messenger	make	people	religion	islam
6	make	make	islam	islamic	make	islam	take	islam	messenger	islam	islamic	born
7	religion	take	time	make	east	make	make	take	religion	religion	prophet	god
8	islam	islam	east	born	time	know	islam	time	come	time	make	make
9	messenger	come	make	sure	take	messenger	kill	war	islam	make	kill	know
10	know	know	god	islam	kill	god	report	prophet	know	know	take	religion
11	give	give	born	religion	born	soldier	born	religion	kill	come	know	time
12	jihad	time	soldier	time	god	give	war	know	take	god	crusader	prophet
13	take	kill	jihad	son	give	time	time	jihad	time	soldier	jihad	report
14	report	soldier	day	day	go	east	messenger	east	prophet	take	east	east
15	kill	fight	kill	know	religion	take	god	crusader	day	born	soldier	come
16	soldier	report	war	give	know	go	know	born	give	messenger	time	sure
17	east	prophet	brother	good	good	prophet	day	khilafah	born	kill	come	day
18	day	messenger	give	fight	messenger	crusader	good	fight	jihad	good	god	messenger
19	time	become	know	fact	jihad	war	religion	make	enemy	east	give	good
20	good	land	take	come	come	brother	prophet	day	among	jihad	war	go
21	brother	even	report	war	even	land	come	land	god	brother	become	jihad
22	fight	god	land	take	prophet	come	give	mujahidin	many	war	brother	soldier
23	come	jihad	come	jihad	son	fight	soldier	enemy	soldier	land	fight	among
24	woman	among	crusader	go	land	good	call	come	east	day	go	give
25	khilafah	brother	fight	kill	brother	jihad	sure	kill	crusader	even	messenger	even
26	even	east	messenger	see	fight	place	indeed	soldier	lord	khilafah	khilafah	war
27	enemy	war	become	messenger	sure	enemy	among	good	become	fight	day	take
28	crusader	lord	prophet	group	war	indeed	order	give	good	enemy	sure	woman
29	war	battle	among	cent	day	become	jihad	god	mujahidin	order	battle	kill

Πλήθος μοναδικών λέξεων: 50

Λίστα μοναδικών λέξεων σε τυχαία σειρά: cent, islamic, god, indeed, allah, group, religion, war, call, mujahidin, muslim, among, battle, born, prophet, messenger, east, make, khilafah, give, come, people, order, fight, become, jihad, see, take, islam, report, lord, enemy, kill, woman, state, even, day, soldier, crusader, son, time, brother, place, fact, sure, good, go, know, land, many



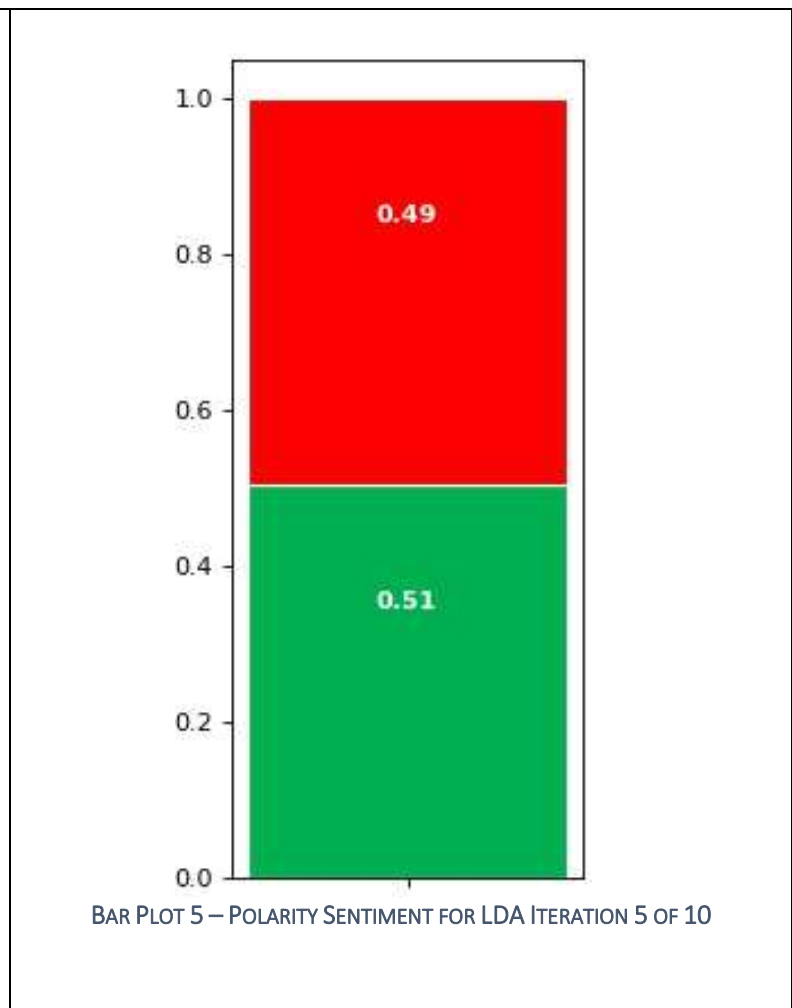
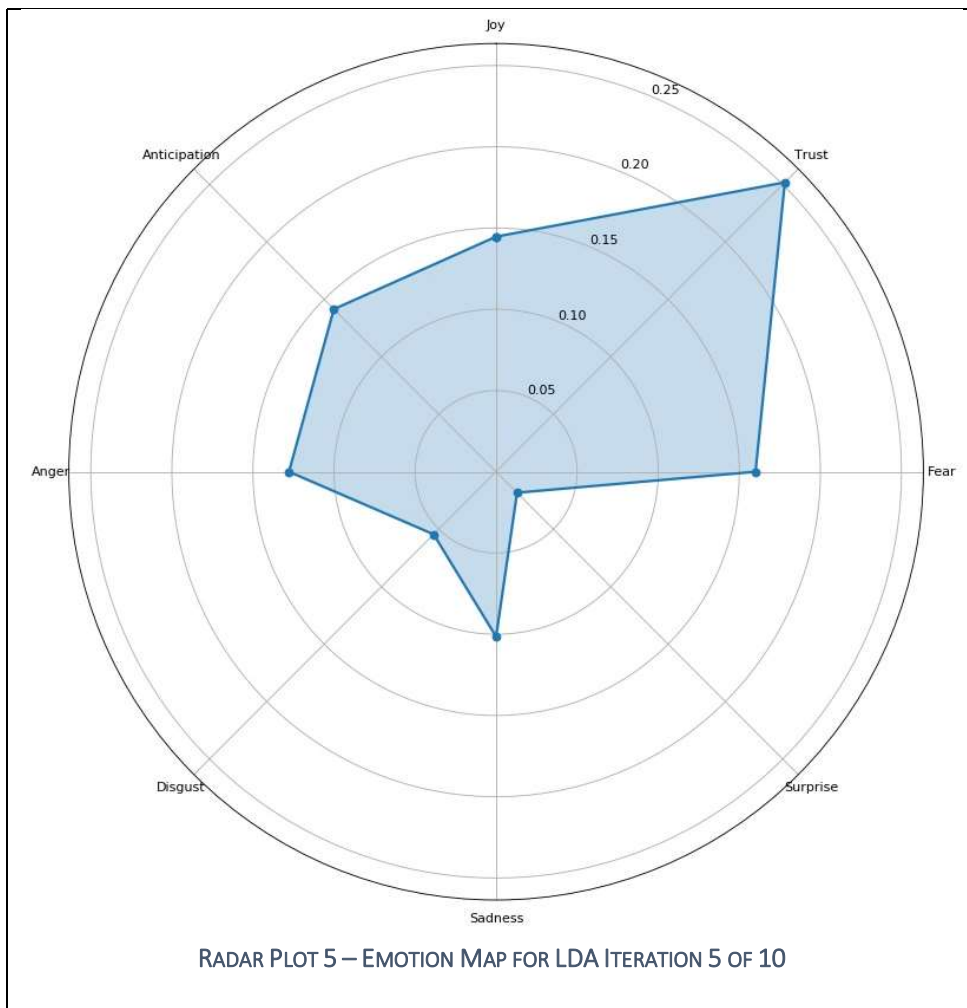
14.55ⁿ επανάληψη της LDA από 10

TABLE 14-5. ΠΙΝΑΚΑΣ ΠΡΟΣΔΙΟΡΙΖΟΜΕΝΩΝ ΛΕΞΕΩΝ ΑΝΑ ΘΕΜΑ ΑΠΟ ΤΗΝ LDA (ΕΠΑΝΑΛΗΨΗ: 5^H ΑΠΟ 10)

	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05	Topic # 06	Topic # 07	Topic # 08	Topic # 09	Topic # 10	Topic # 11	Topic # 12
0	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah
1	state	muslim	state	muslim	people	people	muslim	muslim	east	people	people	muslim
2	people	islamic	people	state	muslim	state	islamic	people	born	state	muslim	islamic
3	islamic	god	muslim	people	islamic	muslim	people	state	muslim	muslim	state	state
4	muslim	state	islamic	islamic	state	islam	state	east	state	islamic	islamic	people
5	religion	people	kill	islam	islam	east	make	god	islamic	take	religion	time
6	make	islam	religion	make	make	know	born	islamic	god	religion	islam	make
7	islam	make	make	religion	religion	give	religion	make	people	islam	make	take
8	war	religion	islam	soldier	god	messenger	time	jihad	sure	messenger	messenger	religion
9	kill	give	east	know	know	god	come	messenger	make	make	take	east
10	know	take	born	give	give	come	kill	time	islam	time	even	messenger
11	brother	know	time	prophet	brother	prophet	islam	take	time	jihad	soldier	islam
12	born	east	messenger	report	war	make	east	report	know	come	kill	give
13	messenger	messenger	know	jihad	east	islamic	know	fight	good	know	jihad	come
14	soldier	jihad	give	messenger	messenger	day	jihad	religion	son	prophet	time	kill
15	time	time	prophet	time	time	born	take	born	go	give	come	born
16	son	soldier	go	take	jihad	take	prophet	among	religion	fight	prophet	prophet
17	day	even	war	come	soldier	become	god	islam	brother	soldier	give	even
18	fight	come	god	land	day	good	give	war	take	crusader	god	know
19	jihad	war	day	east	report	time	good	sure	kill	war	know	day
20	take	report	brother	day	come	religion	fight	kill	come	day	crusader	god
21	prophet	land	fight	among	born	fight	messenger	good	war	land	land	soldier
22	east	day	take	god	fight	kill	sure	give	day	even	fight	enemy
23	give	kill	sure	even	take	indeed	report	brother	group	born	report	war
24	good	prophet	land	good	even	battle	crusader	know	plus	good	day	jihad
25	among	woman	crusader	crusader	khilafah	brother	soldier	day	among	god	enemy	order
26	god	place	even	kill	kill	land	war	son	fight	son	war	become
27	report	brother	become	fight	way	enemy	brother	go	fact	kill	east	report
28	mujahidin	enemy	enemy	lord	prophet	report	order	prophet	become	indeed	order	go
29	come	call	soldier	enemy	woman	among	among	many	saddle	see	indeed	good

Πλήθος μοναδικών λέξεων: 52

Λίστα μοναδικών λέξεων σε τυχαία σειρά: islamic, god, indeed, allah, mujahidin, religion, war, call, group, muslim, among, born, battle, prophet, messenger, east, way, make, khilafah, give, come, people, order, fight, become, jihad, see, saddle, take, islam, report, lord, enemy, kill, woman, state, even, day, soldier, crusader, son, time, brother, place, fact, plus, sure, good, go, know, land, many



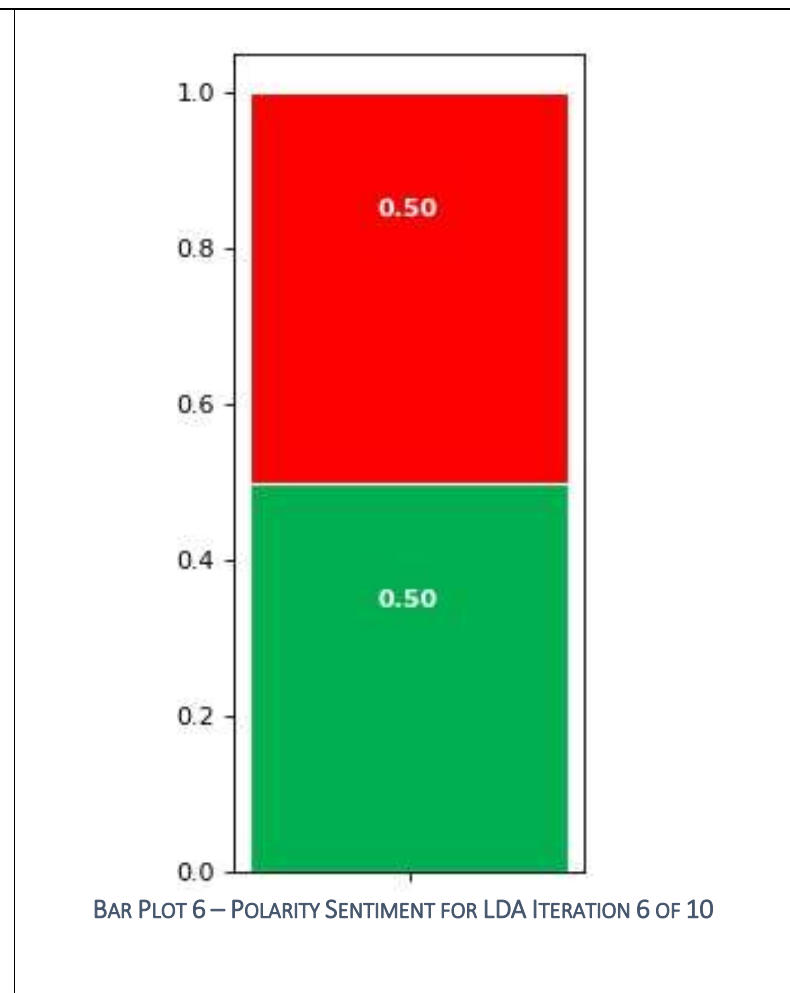
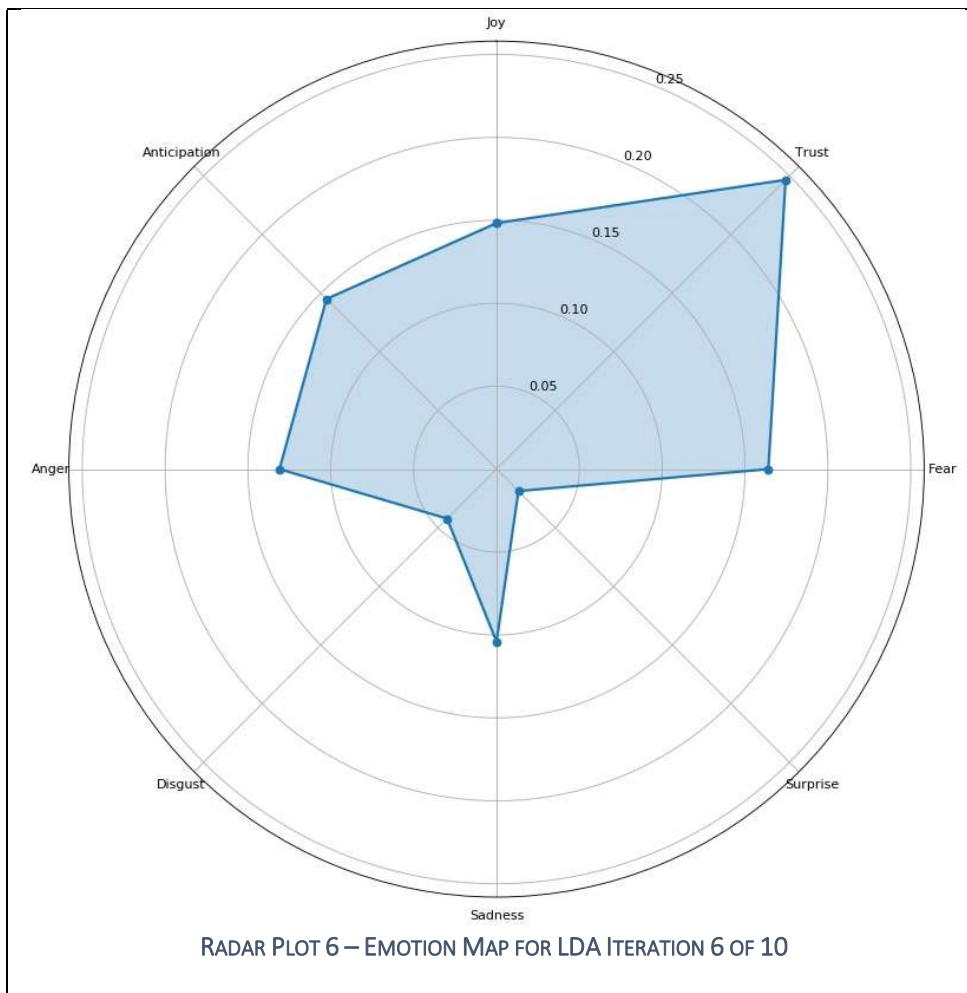
14.66ⁿ επανάληψη της LDA από 10

TABLE 14-6. ΠΙΝΑΚΑΣ ΠΡΟΣΔΙΟΡΙΖΟΜΕΝΩΝ ΛΕΞΕΩΝ ΑΝΑ ΘΕΜΑ ΑΠΟ ΤΗΝ LDA (ΕΠΑΝΑΛΗΨΗ: 6^Η ΑΠΟ 10)

	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05	Topic # 06	Topic # 07	Topic # 08	Topic # 09	Topic # 10	Topic # 11	Topic # 12
0	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah
1	state	islamic	muslim	muslim	state	muslim	people	state	muslim	islamic	muslim	muslim
2	people	muslim	people	people	muslim	people	muslim	muslim	islamic	people	state	people
3	muslim	people	state	state	islam	islamic	islamic	islamic	people	state	islamic	islamic
4	east	make	religion	islamic	people	state	state	people	state	muslim	people	state
5	islam	state	islam	take	god	time	east	make	prophet	god	make	god
6	islamic	religion	islamic	east	born	religion	make	give	east	islam	islam	religion
7	born	islam	take	islam	east	make	religion	religion	islam	make	time	come
8	make	god	make	religion	islamic	islam	messenger	islam	time	east	born	islam
9	time	give	jihad	jihad	time	give	kill	time	messenger	born	take	messenger
10	god	born	know	kill	give	know	fight	know	born	give	religion	make
11	know	fight	day	brother	make	prophet	soldier	come	take	time	messenger	east
12	religion	prophet	kill	make	religion	fight	god	take	know	religion	know	take
13	take	messenger	give	know	war	god	come	messenger	kill	messenger	prophet	soldier
14	soldier	kill	come	war	know	take	report	kill	religion	report	god	kill
15	son	come	crusader	messenger	fight	come	jihad	soldier	make	know	brother	war
16	prophet	soldier	messenger	report	good	jihad	time	day	give	jihad	report	time
17	war	know	khilafah	prophet	take	messenger	born	god	god	take	come	know
18	good	east	soldier	born	sure	many	know	prophet	come	war	sure	khilafah
19	go	good	god	land	kill	good	islam	jihad	day	cent	day	good
20	come	war	report	group	day	battle	brother	brother	land	son	east	report
21	sure	take	become	good	jihad	even	woman	even	jihad	among	kill	brother
22	become	day	among	time	soldier	born	even	many	become	good	many	prophet
23	give	among	even	crusader	prophet	woman	land	land	crusader	go	even	jihad
24	fight	jihad	indeed	son	messenger	land	take	born	even	fight	order	give
25	messenger	sure	time	khilafah	report	place	war	war	soldier	day	give	enemy
26	even	report	call	god	lord	group	good	sure	report	khilafah	enemy	day
27	land	even	prophet	day	come	crusader	day	indeed	brother	crusader	jihad	call
28	day	brother	mujahidin	come	go	brother	sure	fight	good	become	believer	even
29	kill	way	fight	go	among	war	son	become	war	enemy	land	become

Πλήθος μοναδικών λέξεων: 50

Λίστα μοναδικών λέξεων με τυχαία σειρά: cent, islamic, god, indeed, allah, call, religion, war, mujahidin, group, muslim, among, born, battle, prophet, messenger, east, way, make, khilafah, come, give, people, order, become, fight, jihad, believer, take, islam, report, lord, enemy, kill, woman, state, even, day, soldier, crusader, son, time, brother, place, sure, good, go, know, land, many



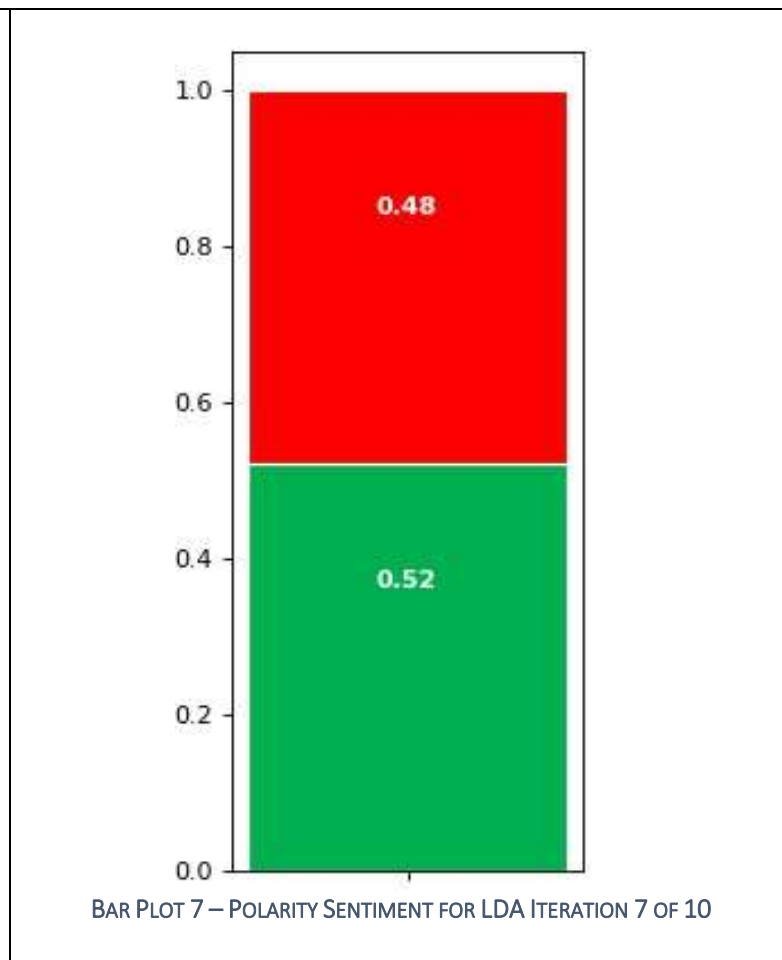
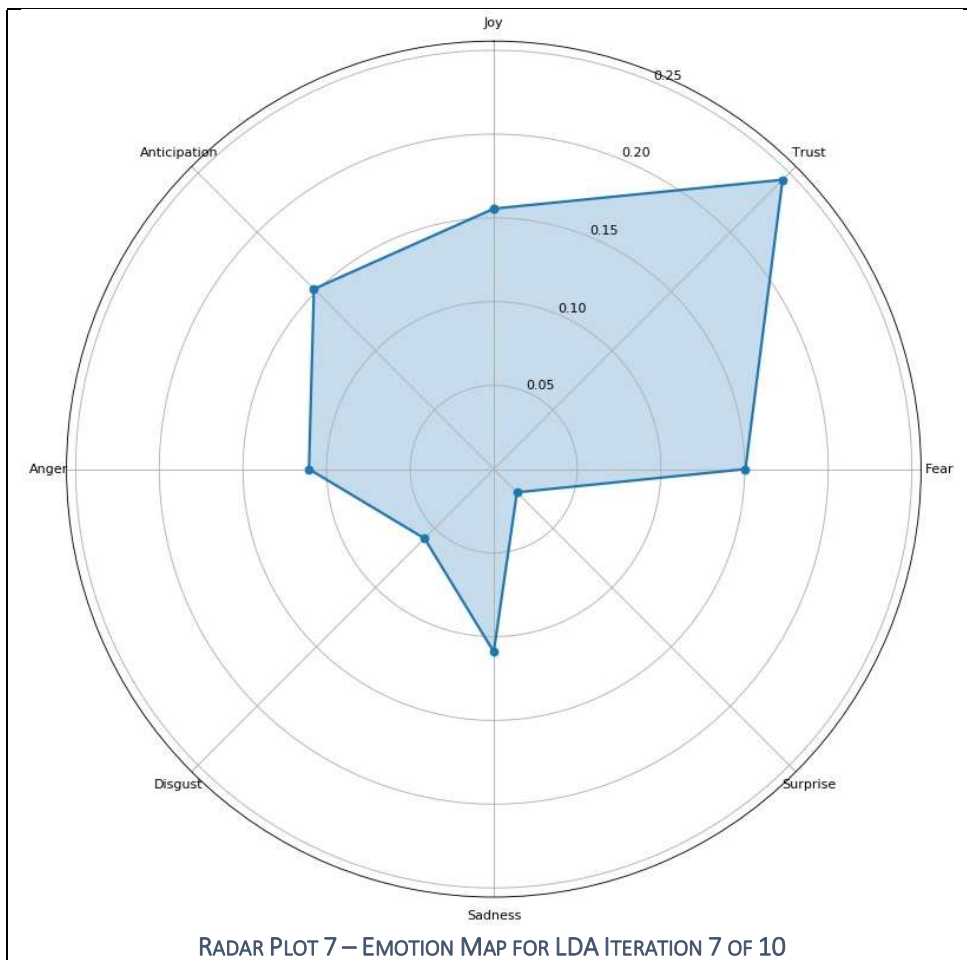
14.77^η επανάληψη της LDA από 10

TABLE 14-7. ΠΙΝΑΚΑΣ ΠΡΟΣΔΙΟΡΙΖΟΜΕΝΩΝ ΛΕΞΕΩΝ ΑΝΑ ΘΕΜΑ ΑΠΟ ΤΗΝ LDA (ΕΠΑΝΑΛΗΨΗ: 7^Η ΑΠΟ 10)

	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05	Topic # 06	Topic # 07	Topic # 08	Topic # 09	Topic # 10	Topic # 11	Topic # 12
0	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah
1	muslim	people	muslim	muslim	state	state	muslim	muslim	muslim	islamic	muslim	muslim
2	state	state	east	state	people	people	islamic	people	islamic	muslim	people	state
3	people	islamic	people	islamic	muslim	muslim	state	islamic	people	state	state	people
4	islam	god	state	people	make	islam	people	state	state	people	islamic	islamic
5	religion	muslim	make	make	islamic	islamic	religion	religion	islam	islam	make	religion
6	prophet	islam	islamic	god	god	give	give	make	know	make	islam	take
7	give	east	sure	islam	know	war	make	god	make	east	time	east
8	come	religion	time	messenger	come	time	islam	east	religion	god	religion	born
9	messenger	kill	prophet	east	time	religion	know	time	take	religion	kill	fight
10	islamic	make	son	take	islam	come	take	islam	war	born	god	make
11	know	take	religion	know	born	god	jihad	messenger	time	good	report	time
12	make	messenger	born	son	east	messenger	east	know	jihad	take	take	day
13	jihad	born	god	born	religion	know	soldier	soldier	give	jihad	messenger	soldier
14	go	know	messenger	jihad	give	make	kill	kill	soldier	know	give	know
15	take	time	good	religion	messenger	crusader	time	take	god	time	know	islam
16	soldier	give	islam	kill	kill	report	born	give	day	prophet	jihad	prophet
17	even	prophet	among	come	fight	jihad	go	report	go	fight	brother	messenger
18	kill	fight	soldier	time	land	take	god	come	born	day	indeed	sure
19	god	go	kill	sure	crusader	brother	messenger	born	prophet	sure	east	war
20	mujahidin	become	come	good	prophet	day	prophet	good	come	soldier	war	come
21	fight	good	take	day	good	kill	war	day	report	messenger	even	land
22	become	jihad	report	war	day	even	come	sure	east	land	khilafah	brother
23	born	report	day	soldier	jihad	enemy	see	jihad	brother	brother	come	give
24	day	group	give	report	take	fight	brother	prophet	crusader	even	prophet	god
25	report	soldier	plus	give	brother	khilafah	become	brother	become	kill	born	crusader
26	woman	war	land	land	lord	prophet	son	war	fight	come	soldier	many
27	khilafah	way	fight	among	among	group	report	even	khilafah	war	become	order
28	land	come	jihad	woman	call	become	land	group	among	report	day	even
29	time	sure	know	mujahidin	woman	land	good	go	son	give	fight	kill

Πλήθος μοναδικών λέξεων: 48

Λίστα μοναδικών λέξεων σε τυχαία σειρά: islamic, god, indeed, allah, mujahidin, religion, group, war, call, muslim, among, born, prophet, messenger, east, way, make, khilafah, give, come, people, order, fight, become, jihad, see, take, islam, report, lord, enemy, kill, woman, state, even, day, soldier, crusader, son, time, brother, plus, sure, go, good, know, land, many



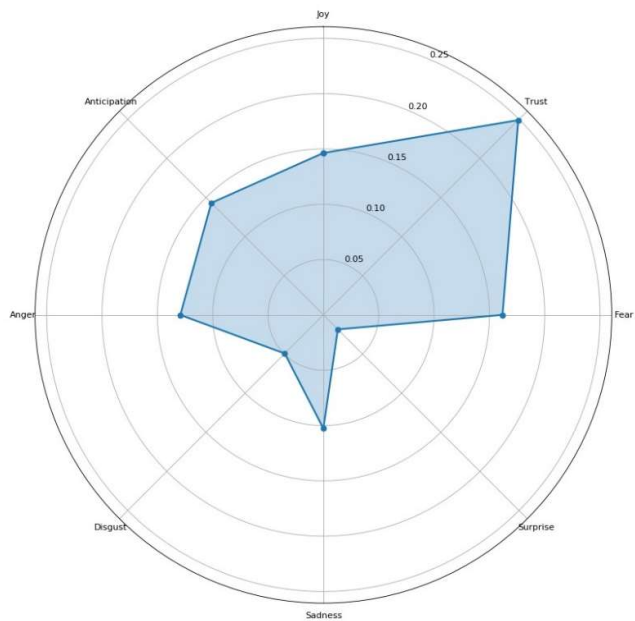
14.88ⁿ επανάληψη της LDA από 10

TABLE 14-8. ΠΙΝΑΚΑΣ ΠΡΟΣΔΙΟΡΙΖΟΜΕΝΩΝ ΛΕΞΕΩΝ ΑΝΑ ΘΕΜΑ ΑΠΟ ΤΗΝ LDA (ΕΠΑΝΑΛΗΨΗ: 8^H ΑΠΟ 10)

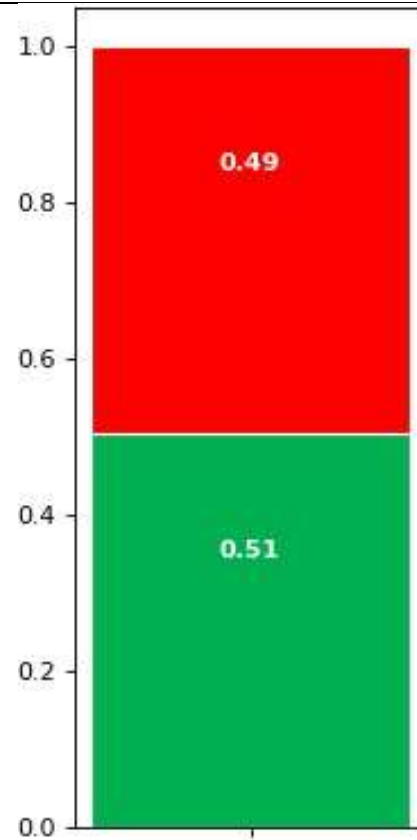
	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05	Topic # 06	Topic # 07	Topic # 08	Topic # 09	Topic # 10	Topic # 11	Topic # 12
0	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah
1	muslim	people	muslim	state	muslim	people	people	state	muslim	muslim	east	state
2	state	islamic	state	muslim	state	muslim	state	people	state	people	muslim	muslim
3	people	muslim	islamic	people	people	state	islamic	muslim	islamic	state	islamic	islamic
4	islamic	islam	people	religion	east	islamic	muslim	islamic	people	islamic	people	people
5	islam	religion	make	make	islamic	islam	make	islam	islam	make	born	east
6	messenger	state	religion	god	religion	make	come	religion	make	know	state	islam
7	make	take	time	islam	islam	religion	islam	take	god	god	sure	religion
8	god	kill	take	islamic	born	god	know	time	time	religion	time	time
9	born	make	come	give	make	soldier	war	prophet	know	messenger	make	born
10	give	know	jihad	born	brother	know	prophet	make	messenger	come	know	take
11	kill	god	islam	east	give	time	messenger	report	give	kill	religion	sure
12	good	time	prophet	time	day	give	day	fight	kill	east	god	give
13	religion	come	messenger	take	time	jihad	religion	god	religion	islam	son	make
14	brother	east	fight	come	soldier	prophet	kill	go	take	jihad	among	know
15	east	messenger	god	jihad	messenger	take	god	born	war	take	take	god
16	take	report	know	messenger	prophet	brother	soldier	war	become	give	fight	messenger
17	soldier	prophet	give	order	kill	east	fight	land	east	soldier	islam	fight
18	time	day	soldier	even	god	war	enemy	messenger	born	time	kill	come
19	report	even	kill	land	know	good	land	day	good	day	give	war
20	go	brother	war	know	sure	messenger	good	know	jihad	war	report	jihad
21	know	indeed	east	prophet	come	report	crusader	even	among	crusader	good	soldier
22	jihad	among	report	good	take	kill	mujahidin	kill	land	prophet	war	report
23	come	give	day	kill	jihad	born	brother	come	brother	report	day	kill
24	crusader	good	call	report	even	even	khilafah	become	prophet	son	messenger	day
25	woman	born	crusader	way	fight	day	take	give	day	fight	prophet	good
26	indeed	khilafah	even	day	report	order	jihad	khilafah	soldier	born	jihad	go
27	mujahidin	jihad	go	woman	battle	crusader	give	soldier	group	become	fact	even
28	fight	son	order	enemy	enemy	among	indeed	enemy	come	without	go	son
29	prophet	become	imam	place	war	see	east	crusader	crusader	order	soldier	brother

Πλήθος μοναδικών λέξεων: 50

Λίστα μοναδικών λέξεων σε τυχαία σειρά:islamic, god, indeed, imam, allah, mujahidin, religion, war, call, group, muslim, among, born, battle, prophet, messenger, east, way, make, khilafah, give, come, people, order, fight, become, jihad, see, take, islam, report, enemy, kill, woman, state, even, day, soldier, crusader, son, time, brother, place, without, fact, sure, good, go, know, land



RADAR PLOT 8 – EMOTION MAP FOR LDA ITERATION 8 OF 10



BAR PLOT 8 – POLARITY SENTIMENT FOR LDA ITERATION 8 OF 10

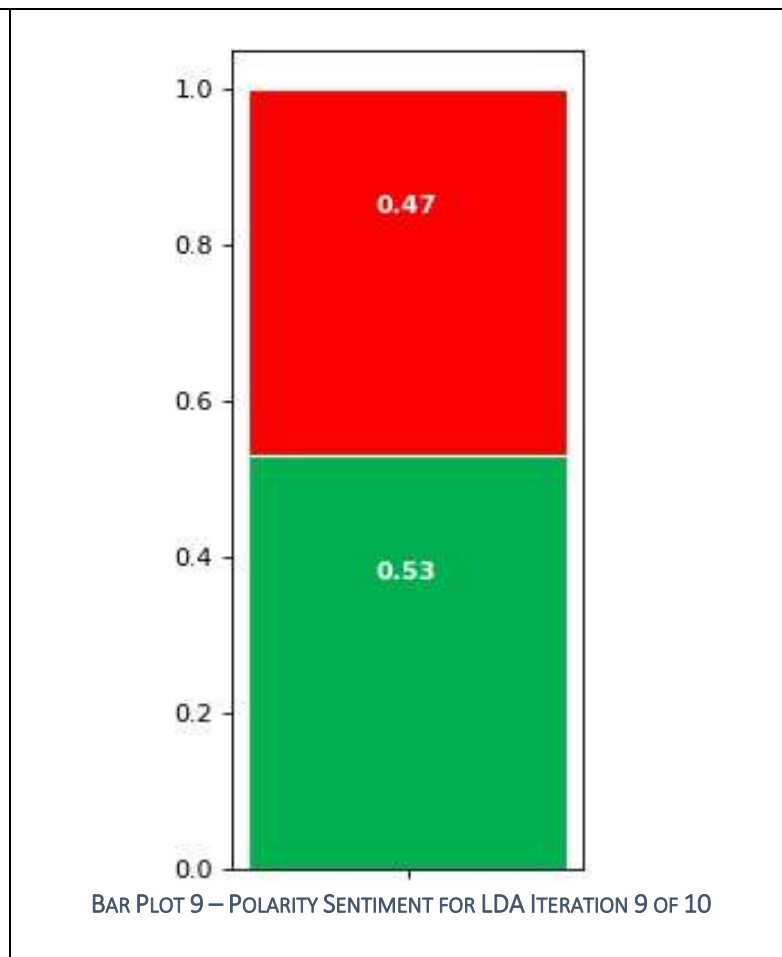
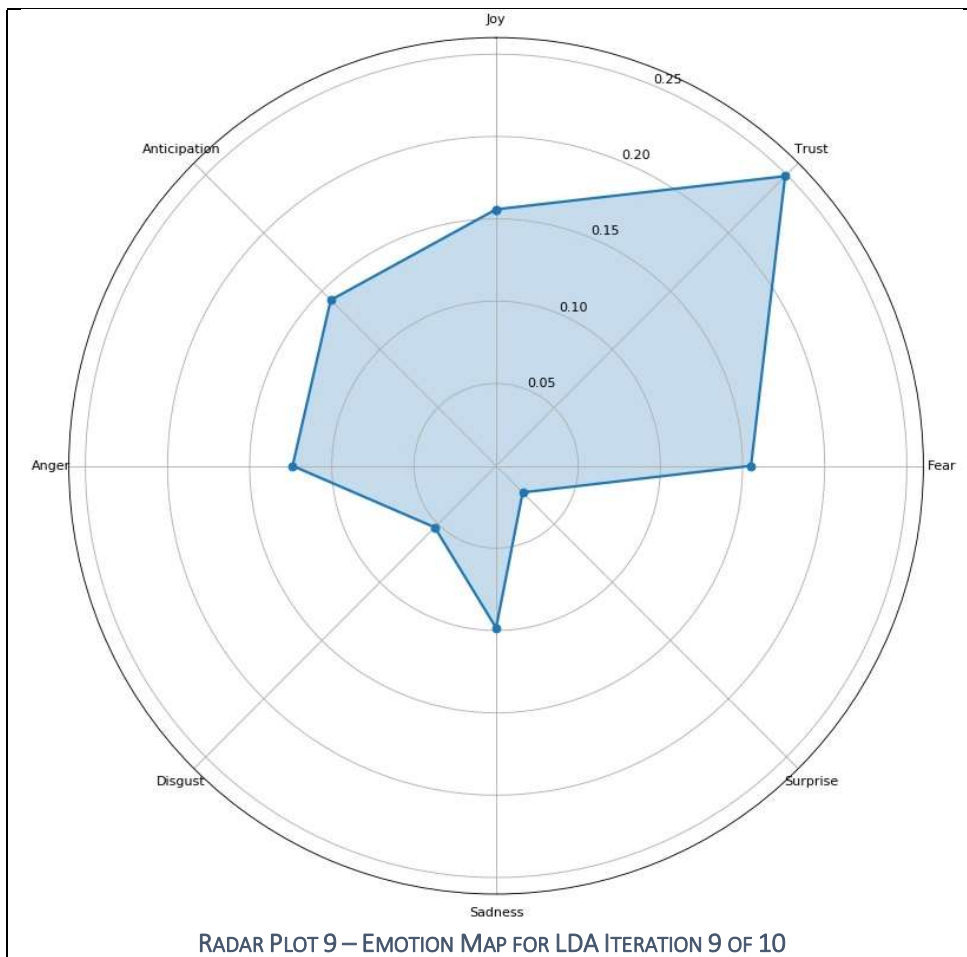
14.99^η επανάληψη της LDA από 10

TABLE 14-9. ΠΙΝΑΚΑΣ ΠΡΟΣΔΙΟΡΙΖΟΜΕΝΩΝ ΛΕΞΕΩΝ ΑΝΑ ΘΕΜΑ ΑΠΟ ΤΗΝ LDA (ΕΠΑΝΑΛΗΨΗ: 9^η ΑΠΟ 10)

	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05	Topic # 06	Topic # 07	Topic # 08	Topic # 09	Topic # 10	Topic # 11	Topic # 12
0	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah
1	people	muslim	state	state	muslim	muslim	muslim	muslim	state	muslim	people	people
2	muslim	people	muslim	muslim	people	state	islamic	state	muslim	people	muslim	state
3	islamic	state	islamic	people	islamic	people	people	people	islamic	state	east	muslim
4	east	religion	make	islamic	state	islamic	state	make	islam	islamic	islamic	islamic
5	state	islamic	people	give	religion	islam	east	islamic	people	islam	state	islam
6	religion	islam	god	islam	islam	make	religion	time	time	sure	kill	religion
7	make	make	give	know	take	give	god	prophet	make	make	make	make
8	know	god	religion	god	time	messenger	born	messenger	messenger	know	born	come
9	born	come	time	religion	make	jihad	islam	religion	prophet	time	sure	time
10	islam	born	islam	messenger	god	time	take	take	religion	god	know	give
11	take	know	come	east	give	god	kill	soldier	war	religion	islam	messenger
12	soldier	give	take	take	prophet	take	make	east	know	jihad	jihad	among
13	come	son	east	prophet	come	know	report	islam	kill	east	war	jihad
14	fight	east	know	make	messenger	good	messenger	know	fight	take	god	god
15	day	report	prophet	jihad	know	prophet	brother	day	brother	messenger	time	prophet
16	messenger	good	kill	good	fight	day	jihad	give	even	day	take	take
17	time	fight	day	become	jihad	brother	good	war	come	born	soldier	enemy
18	kill	time	soldier	brother	report	kill	know	god	day	kill	good	war
19	good	soldier	messenger	war	east	war	fight	come	jihad	good	messenger	even
20	sure	messenger	war	born	born	indeed	give	indeed	god	even	give	crusader
21	prophet	land	jihad	order	kill	crusader	war	crusader	soldier	come	day	know
22	go	take	born	soldier	good	born	khilafah	kill	land	give	among	fight
23	crusader	kill	way	kill	land	soldier	soldier	brother	become	prophet	religion	kill
24	jihad	day	among	day	go	religion	time	become	many	soldier	report	become
25	among	lord	son	time	soldier	even	land	fight	order	crusader	brother	soldier
26	brother	go	khilafah	land	war	battle	among	order	take	many	son	indeed
27	war	way	even	son	brother	many	day	khilafah	go	lord	even	born
28	group	woman	place	report	enemy	become	crusader	land	report	see	place	day
29	give	cent	go	fight	khilafah	come	city	way	towards	become	honor	land

Πλήθος μοναδικών λέξεων: 51

Λίστα μοναδικών λέξεων σε τυχαία σειρά: cent, islamic, god, indeed, allah, group, religion, war, muslim, among, born, battle, prophet, messenger, east, way, make, khilafah, come, give, people, order, fight, become, jihad, honor, see, take, islam, report, lord, enemy, kill, woman, state, even, day, towards, soldier, crusader, son, time, brother, place, city, sure, good, go, know, land, many



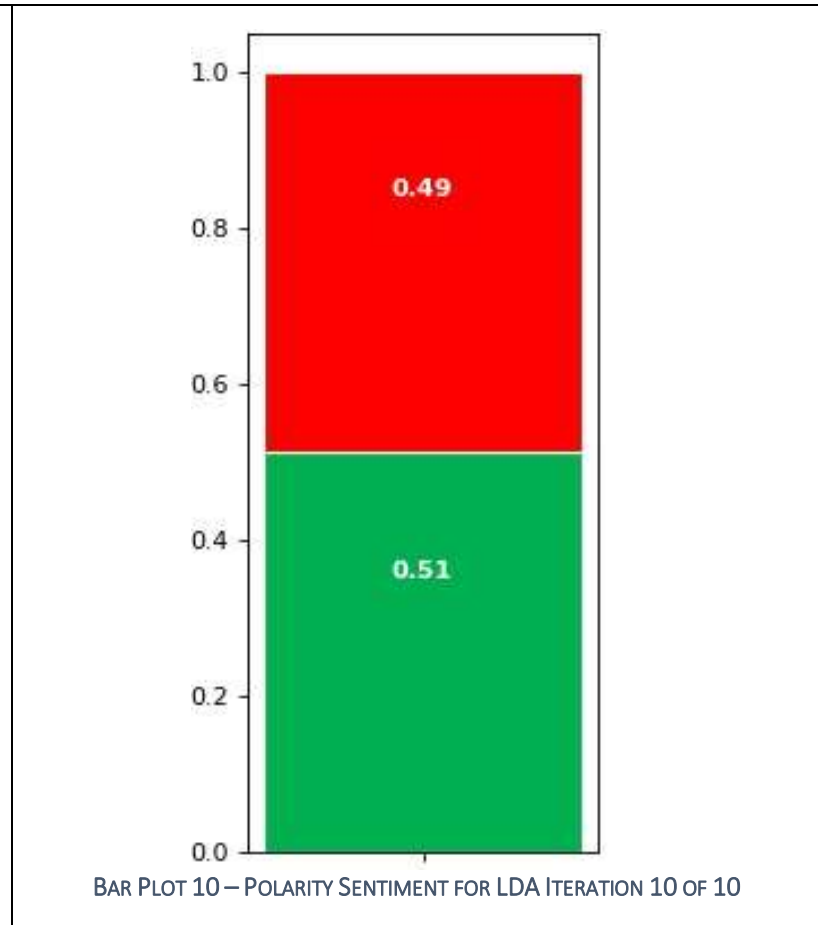
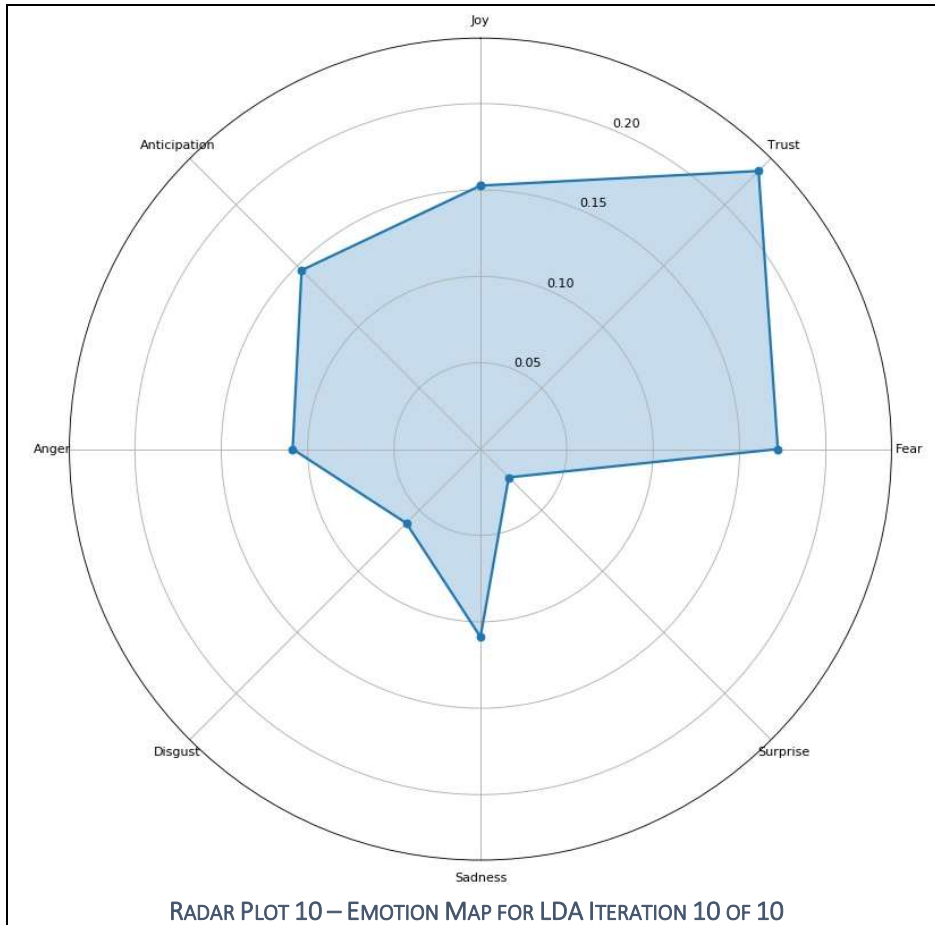
14.10 10^η επανάληψη της LDA από 10

TABLE 14-10. ΠΙΝΑΚΑΣ ΠΡΟΣΔΙΟΡΙΖΟΜΕΝΩΝ ΛΕΞΕΩΝ ΑΝΑ ΘΕΜΑ ΑΠΟ ΤΗΝ LDA (ΕΠΑΝΑΛΗΨΗ: 10^η ΑΠΟ 10)

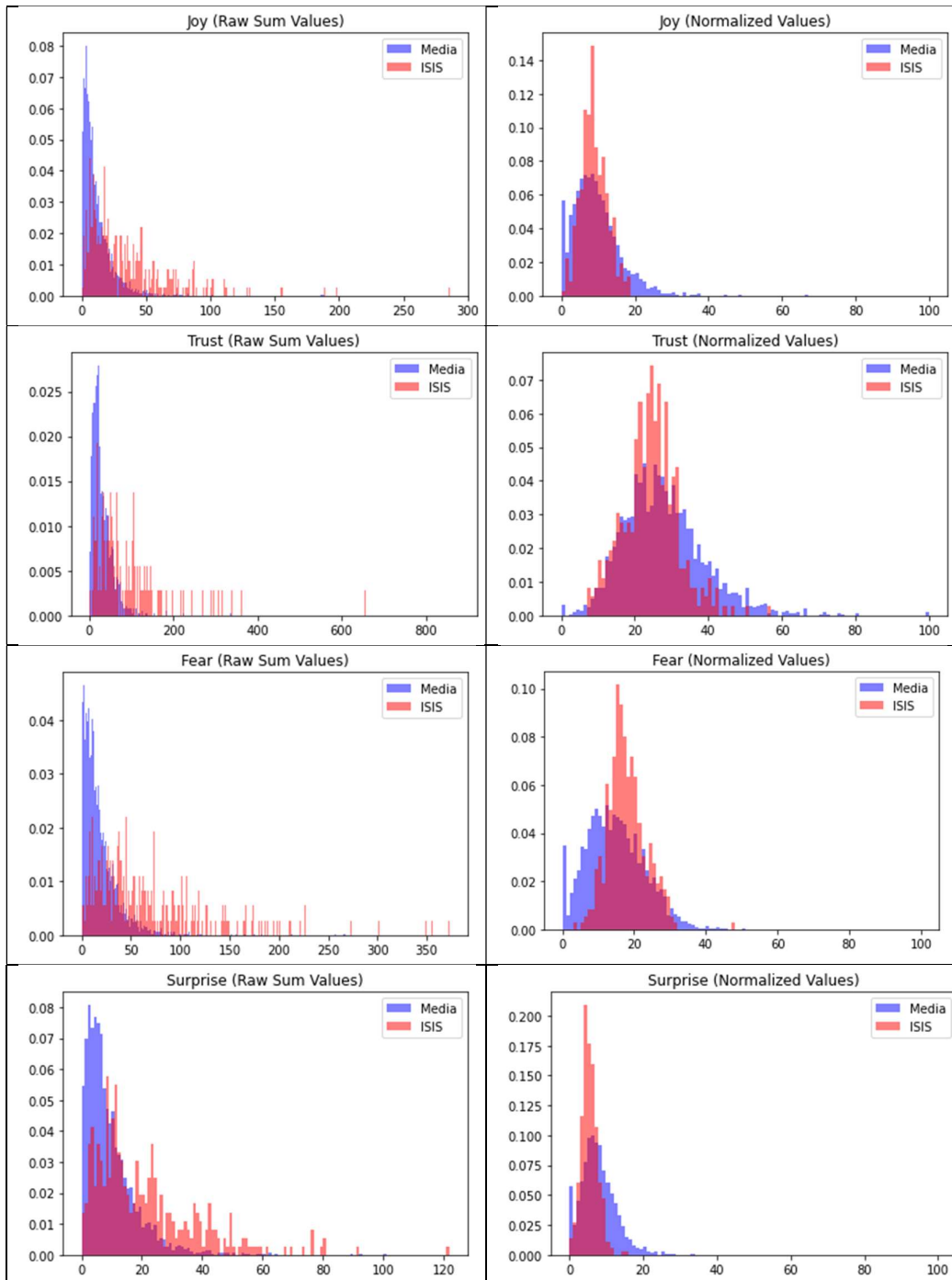
	Topic # 01	Topic # 02	Topic # 03	Topic # 04	Topic # 05	Topic # 06	Topic # 07	Topic # 08	Topic # 09	Topic # 10	Topic # 11	Topic # 12
0	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah	allah
1	state	muslim	muslim	people	muslim	people	islam	state	muslim	state	muslim	islamic
2	muslim	people	state	muslim	people	muslim	state	muslim	state	islamic	people	muslim
3	people	islamic	people	state	state	state	people	people	people	muslim	state	people
4	islamic	east	east	islam	islamic	islamic	islamic	islamic	islamic	people	islamic	religion
5	religion	state	islamic	islamic	religion	islam	muslim	make	time	make	make	state
6	take	religion	born	east	born	religion	come	islam	soldier	god	messenger	time
7	make	time	make	give	take	make	know	messenger	take	islam	religion	god
8	know	know	god	god	make	know	jihad	religion	know	religion	east	make
9	islam	born	islam	messenger	islam	east	give	time	god	give	kill	islam
10	messenger	make	take	take	war	god	religion	god	make	born	good	give
11	fight	islam	time	religion	east	give	make	give	messenger	east	report	good
12	prophet	come	give	come	time	kill	day	prophet	islam	know	know	even
13	land	kill	religion	jihad	come	born	khilafah	kill	east	soldier	come	prophet
14	war	messenger	sure	know	kill	time	kill	report	born	time	day	sure
15	even	fight	day	prophet	land	fight	god	day	religion	kill	born	know
16	kill	take	fight	time	know	jihad	crusader	jihad	prophet	jihad	islam	brother
17	come	war	soldier	fight	jihad	report	take	know	day	prophet	take	messenger
18	jihad	god	come	make	messenger	son	soldier	come	give	take	jihad	born
19	soldier	jihad	good	go	prophet	day	time	land	jihad	messenger	prophet	soldier
20	enemy	land	know	report	good	come	born	take	become	war	brother	kill
21	report	indeed	son	kill	son	take	mujahidin	war	war	call	war	jihad
22	brother	crusader	report	day	brother	prophet	among	east	kill	indeed	fight	come
23	time	among	kill	sure	god	soldier	good	fight	crusader	even	become	take
24	crusader	son	fear	good	give	brother	messenger	soldier	enemy	come	god	call
25	give	good	call	land	towards	khilafah	prophet	go	mujahidin	group	soldier	east
26	day	give	woman	crusader	even	war	brother	brother	many	fight	many	fight
27	khilafah	even	among	born	crusader	enemy	even	born	brother	good	among	go
28	order	go	prophet	war	day	indeed	woman	see	come	day	sure	report
29	become	towards	plus	son	go	order	place	order	like	lord	time	claim

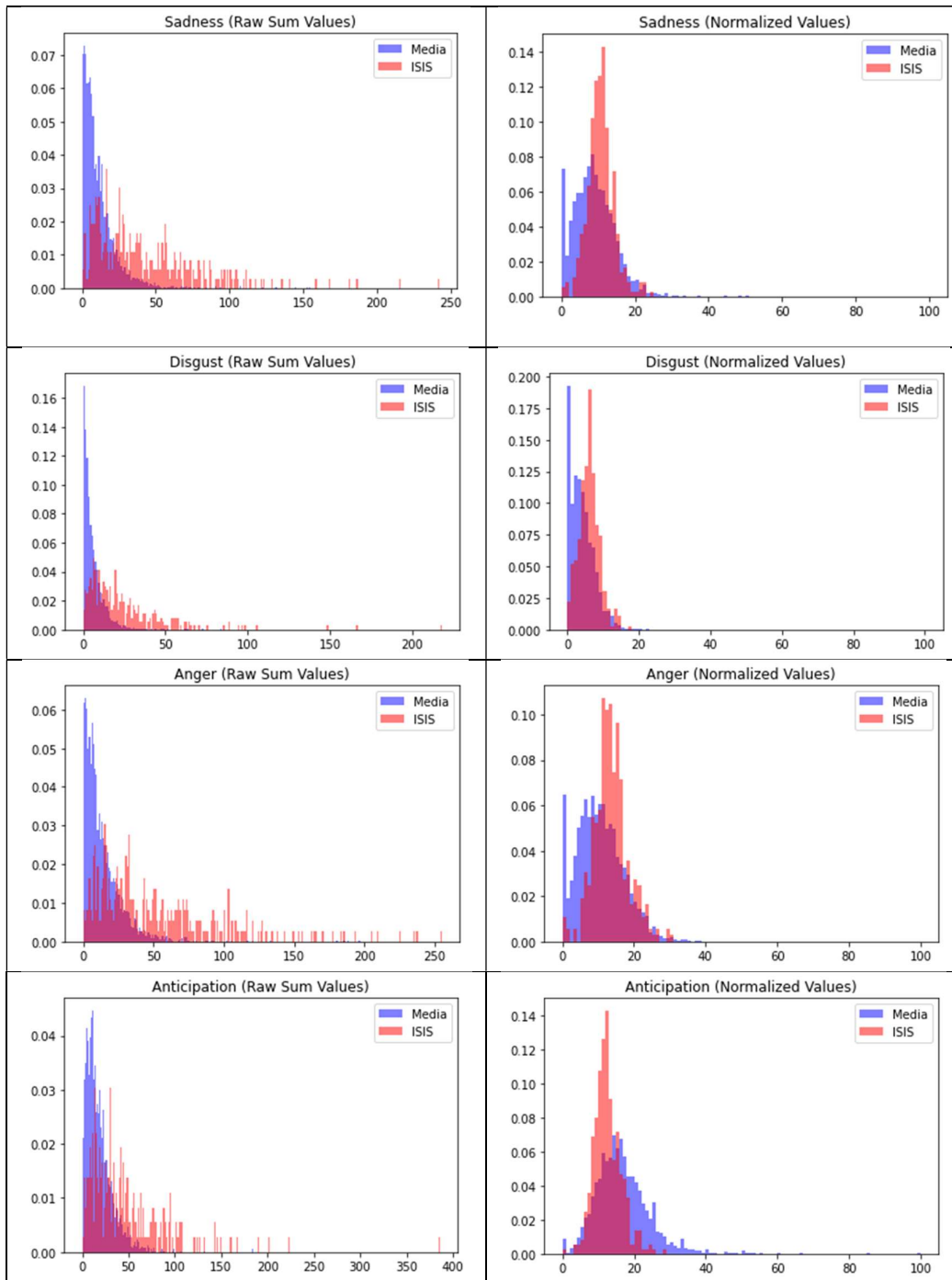
Πλήθος μοναδικών λέξεων: 52

Λίστα μοναδικών λέξεων σε τυχαία σειρά: islamic, god, indeed, allah, call, religion, war, mujahidin, group, muslim, among, born, prophet, messenger, east, make, khilafah, come, give, people, order, fight, become, jihad, see, take, islam, report, lord, enemy, kill, woman, state, even, day, towards, soldier, fear, crusader, son, time, brother, place, claim, plus, sure, like, good, go, know, land, many

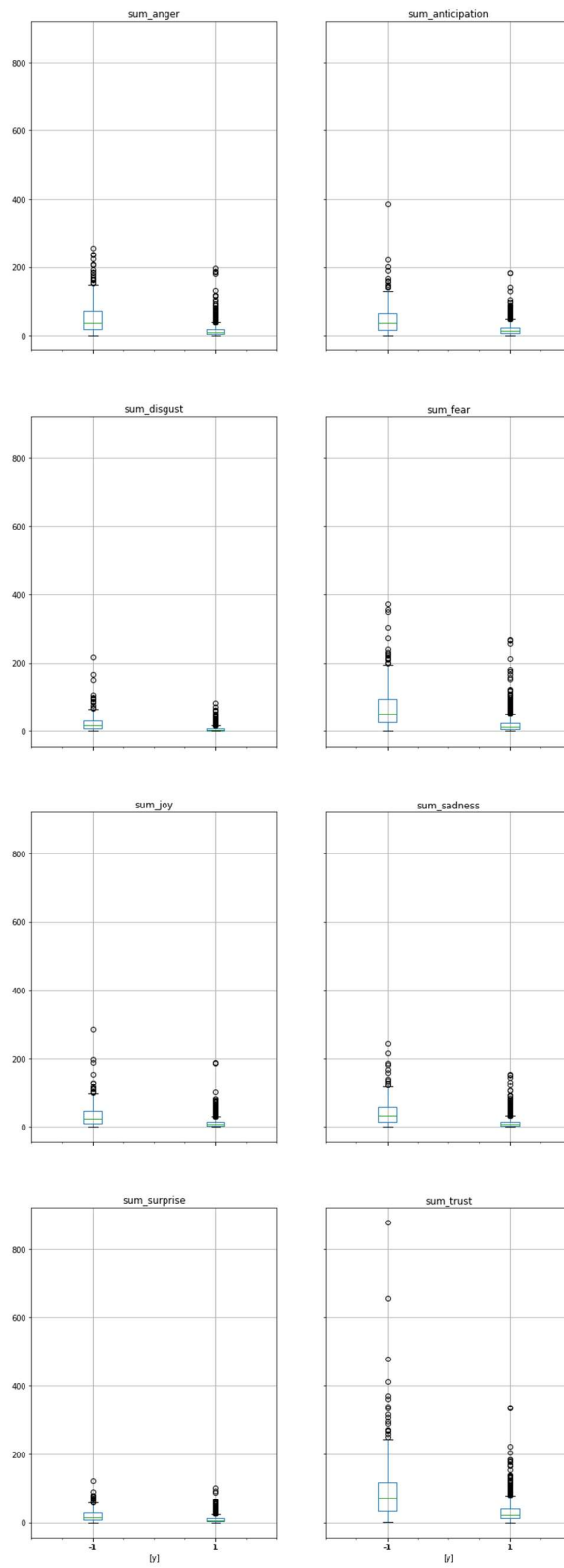


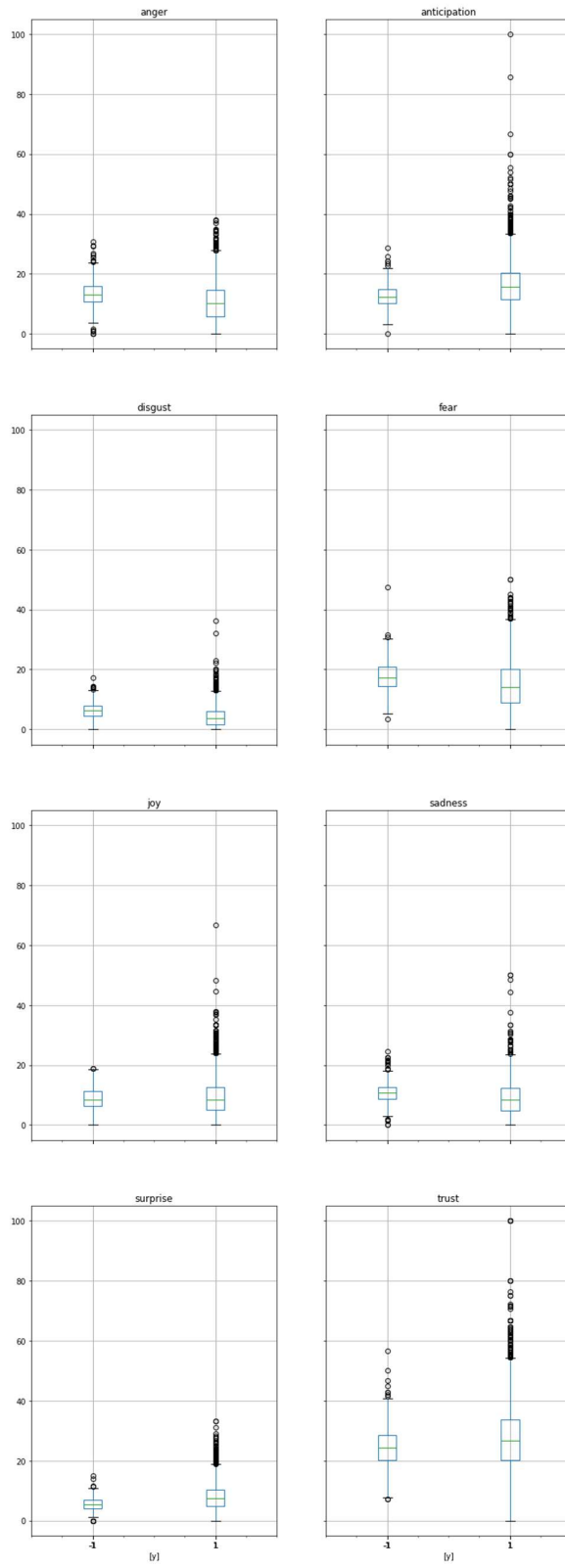
15 Παράρτημα Η – Σχετικά Ιστογράμματα Συναισθημάτων (ISIS / Media)





16 Παράρτημα Θ – Θηκογράμματα Συναισθημάτων





17 Παράρτημα Ι – Ιστογράμματα Πλήθους Λέξεων

17.1abcnews.go.com

Min words: 75

Low 5% Quantile words: 127.30000000000001

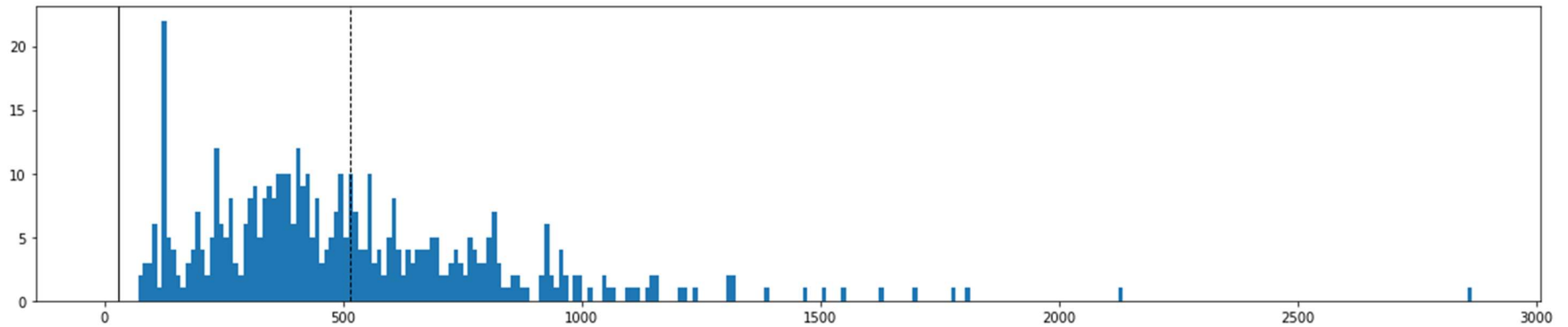
Low Quantile words: 305.25

Mean words: 515.5717299578059

Median words: 444.0

High Quantile words: 671.0

Max words: 2866



17.2 www.cnn.com

Min words: 1

Low 5% Quantile words: 175.0

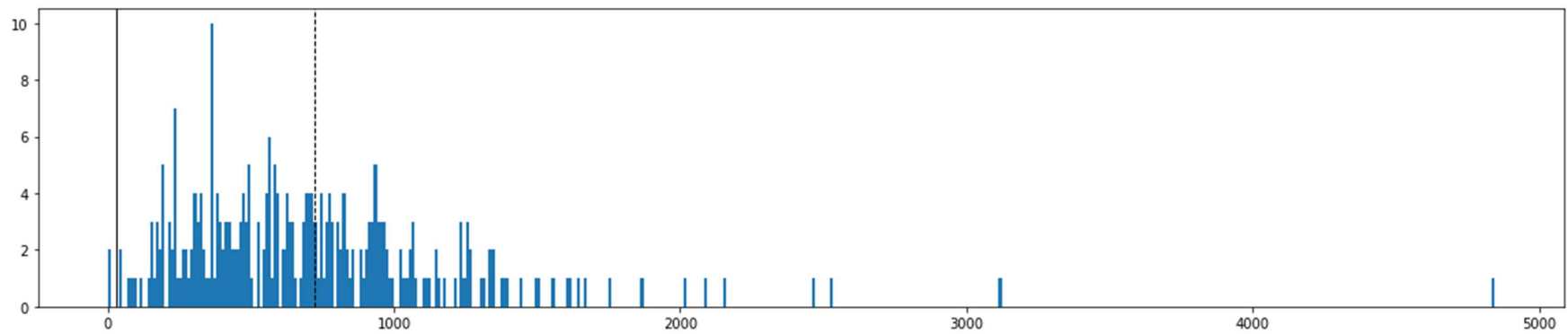
Low Quantile words: 373.5

Mean words: 720.9057971014493

Median words: 630.5

High Quantile words: 933.25

Max words: 4844



17.3www.huffingtonpost.com

Min words: 1

Low 5% Quantile words: 71.75

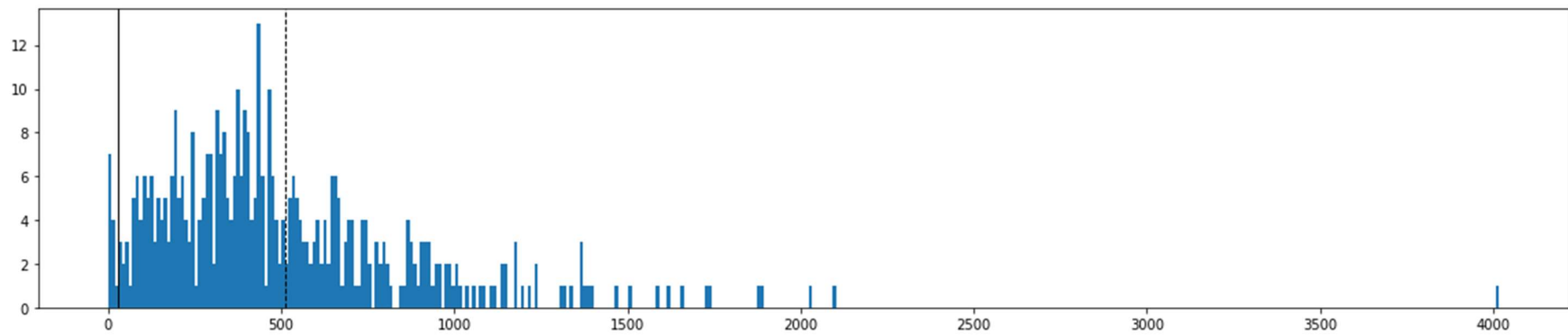
Low Quantile words: 242.75

Mean words: 511.8738532110092

Median words: 421.5

High Quantile words: 668.5

Max words: 4017



17.4www.bbc.co.uk

Min words: 1

Low 5% Quantile words: 34.099999999999994

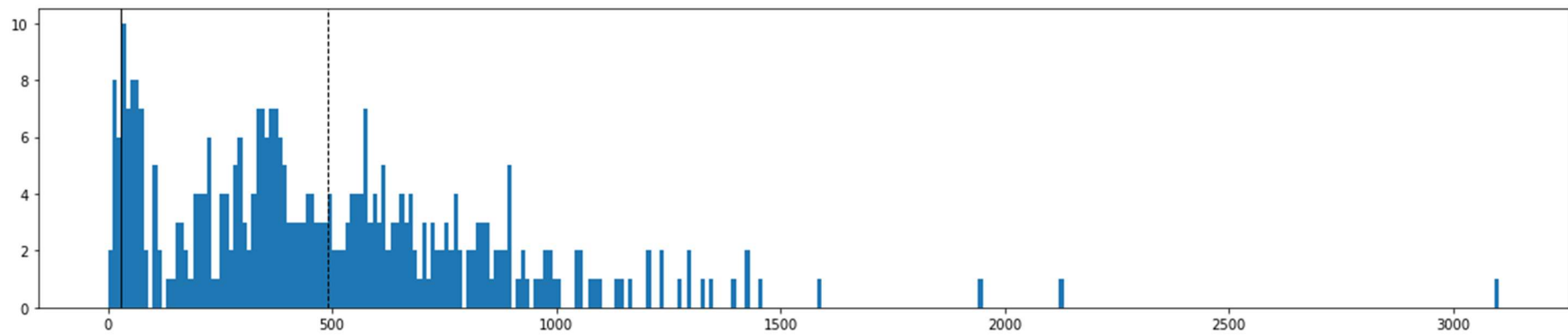
Low Quantile words: 221.5

Mean words: 491.3267605633803

Median words: 421.0

High Quantile words: 678.5

Max words: 3102



17.5tass.com

Min words: 1

Low 5% Quantile words: 80.2

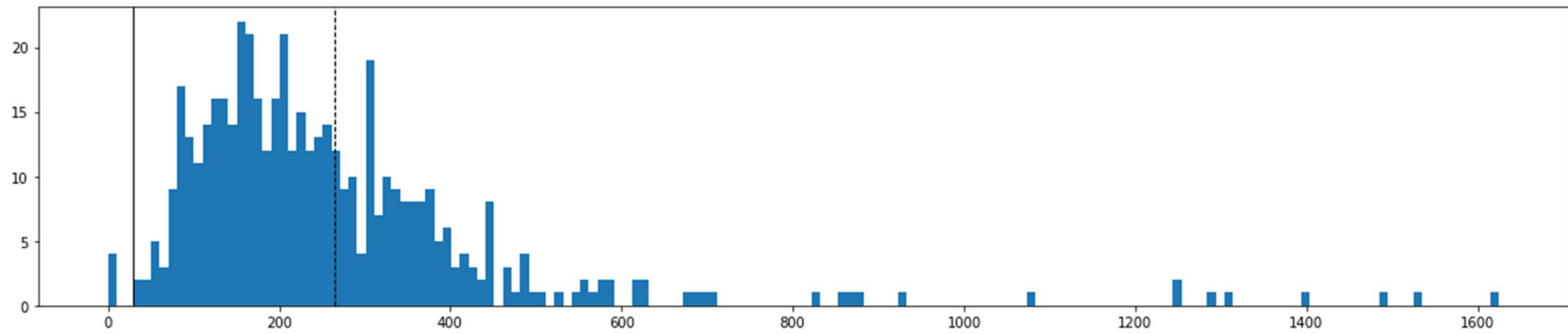
Low Quantile words: 147.0

Mean words: 264.86185567010307

Median words: 216.0

High Quantile words: 326.0

Max words: 1625



17.6www.dw.com

Min words: 1

Low 5% Quantile words: 186.5

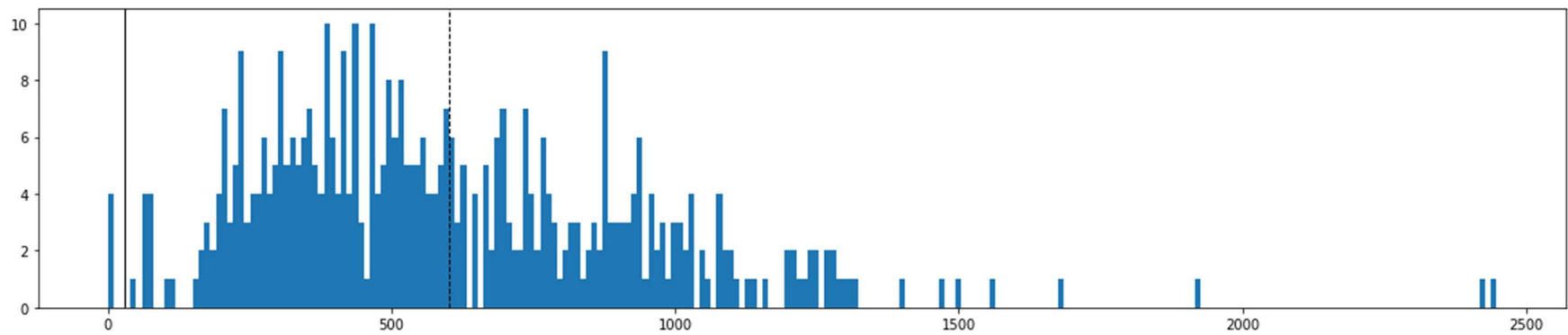
Low Quantile words: 352.0

Mean words: 600.9036697247707

Median words: 532.0

High Quantile words: 811.0

Max words: 2447



17.7www.aljazeera.com

Min words: 1

Low 5% Quantile words: 98.9

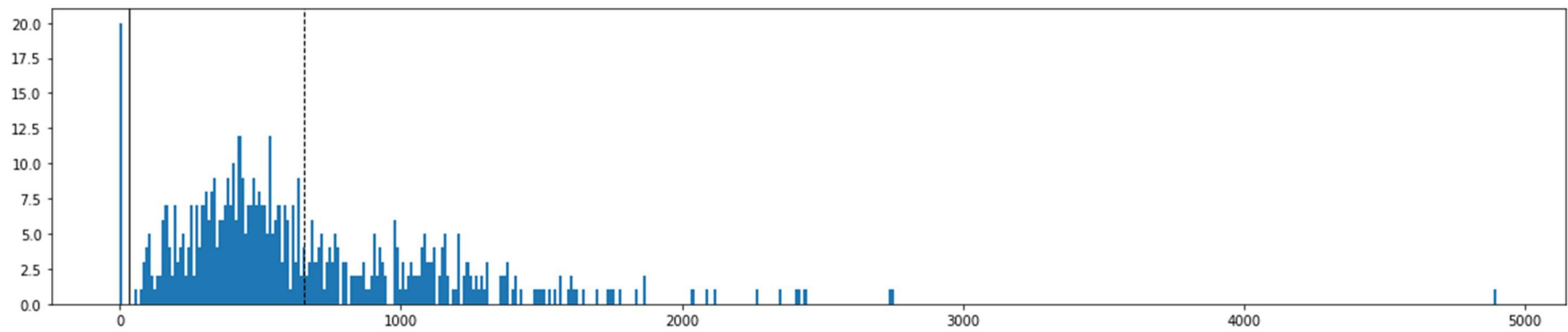
Low Quantile words: 336.5

Mean words: 655.2933810375671

Median words: 529.0

High Quantile words: 907.0

Max words: 4897



17.8europe.chinadaily.com.cn

Min words: 57

Low 5% Quantile words: 214.75

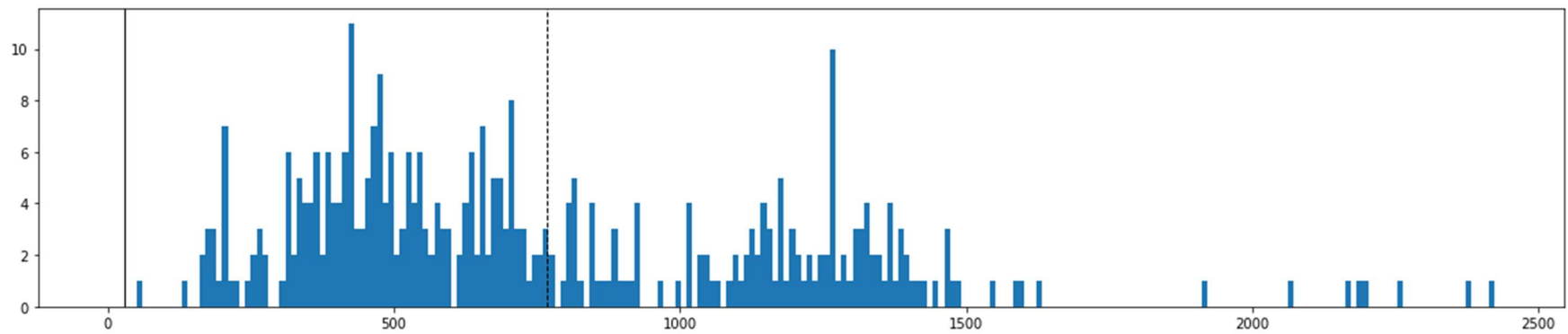
Low Quantile words: 431.75

Mean words: 767.9555555555555

Median words: 659.0

High Quantile words: 1125.75

Max words: 2423



17.9www.rte.ie

Min words: 1

Low 5% Quantile words: 76.1

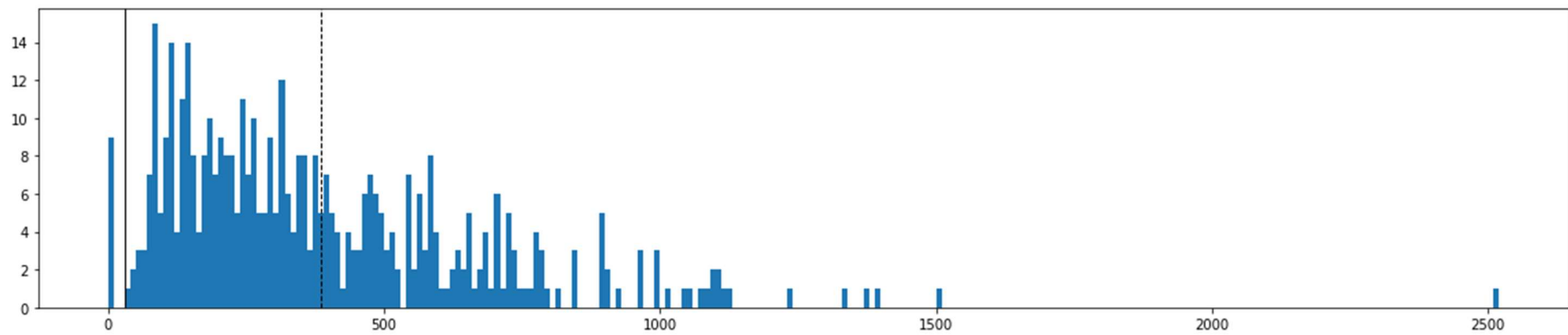
Low Quantile words: 173.5

Mean words: 385.882618510158

Median words: 314.0

High Quantile words: 542.5

Max words: 2520



17.10AI Hayat Media Center

Min words: 67

Low 5% Quantile words: 247.4

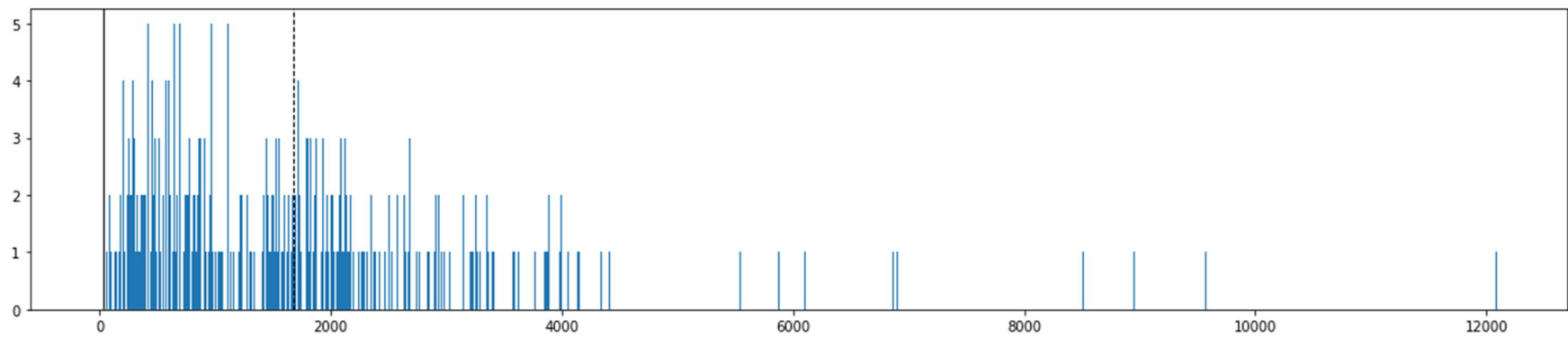
Low Quantile words: 642.5

Mean words: 1682.34435261708

Median words: 1441.0

High Quantile words: 2157.5

Max words: 12090



18 Παράρτημα Κ – Δείκτες Επίδοσης Μοντέλων ανά Αναδίπλωση (Αθροιστικά Συναισθήματα)

18.1 Mahalanobis – Εκπαίδευση

model	accuracy	precision	recall	tp	fp	tn	fn
mahalanobis_0_train	0.9757	1.0	0.9757	(0)	0.0243	0.9757	(0)
mahalanobis_1_train	0.9764	1.0	0.9764	(0)	0.0236	0.9764	(0)
mahalanobis_2_train	0.9734	1.0	0.9734	(0)	0.0266	0.9734	(0)
mahalanobis_3_train	0.9749	1.0	0.9749	(0)	0.0251	0.9749	(0)
mahalanobis_4_train	0.9757	1.0	0.9757	(0)	0.0243	0.9757	(0)
mahalanobis_5_train	0.9776	1.0	0.9776	(0)	0.0224	0.9776	(0)
mahalanobis_6_train	0.9749	1.0	0.9749	(0)	0.0251	0.9749	(0)
mahalanobis_7_train	0.9760	1.0	0.9760	(0)	0.0240	0.9760	(0)
mahalanobis_8_train	0.9753	1.0	0.9753	(0)	0.0247	0.9753	(0)
mahalanobis_9_train	0.9760	1.0	0.9760	(0)	0.0240	0.9760	(0)

18.2 Mahalanobis – Τεστ

model	accuracy	precision	recall	tp	fp	tn	fn
mahalanobis_0_ test	0.8302	0.8301	0.975 2	0.38 02	0.024 8	0.975 2	0.619 8
mahalanobis_1_ test	0.8268	0.8284	0.972 5	0.37 47	0.027 5	0.972 5	0.625 3
mahalanobis_2_ test	0.8356	0.8306	0.983 1	0.37 74	0.016 9	0.983 1	0.622 6
mahalanobis_3_ test	0.8336	0.8297	0.981 4	0.37 47	0.018 6	0.981 4	0.625 3
mahalanobis_4_ test	0.8289	0.8273	0.977 8	0.36 64	0.022 2	0.977 8	0.633 6
mahalanobis_5_ test	0.8282	0.8267	0.977 8	0.36 36	0.022 2	0.977 8	0.636 4
mahalanobis_6_ test	0.8302	0.8281	0.978 7	0.36 91	0.021 3	0.978 7	0.630 9
mahalanobis_7_ test	0.8302	0.8296	0.976 0	0.37 74	0.024 0	0.976 0	0.622 6
mahalanobis_8_ test	0.8295	0.8275	0.978 7	0.36 64	0.021 3	0.978 7	0.633 6
mahalanobis_9_ test	0.8268	0.8284	0.972 5	0.37 47	0.027 5	0.972 5	0.625 3

18.3 ocSVM – Εκπαίδευση

model	accuracy	precision	recall	tp	fp	tn	fn
ocSVM_0_train	0.6776	1.0	0.6776	(0)	0.3224	0.6776	(0)
ocSVM_1_train	0.6300	1.0	0.6300	(0)	0.3700	0.6300	(0)
ocSVM_2_train	0.5924	1.0	0.5924	(0)	0.4076	0.5924	(0)
ocSVM_3_train	0.6030	1.0	0.6030	(0)	0.3970	0.6030	(0)
ocSVM_4_train	0.5989	1.0	0.5989	(0)	0.4011	0.5989	(0)
ocSVM_5_train	0.5141	1.0	0.5141	(0)	0.4859	0.5141	(0)
ocSVM_6_train	0.5376	1.0	0.5376	(0)	0.4624	0.5376	(0)
ocSVM_7_train	0.6802	1.0	0.6802	(0)	0.3198	0.6802	(0)
ocSVM_8_train	0.4631	1.0	0.4631	(0)	0.5369	0.4631	(0)
ocSVM_9_train	0.5779	1.0	0.5779	(0)	0.4221	0.5779	(0)

18.4ocSVM – Τεστ

model	accuracy	precision	recall	tp	fp	tn	fn
ocSVM_0_test	0.2886	0.9855	0.0603	0.9972	0.9397	0.0603	0.0028
ocSVM_1_test	0.2906	0.9730	0.0639	0.9945	0.9361	0.0639	0.0055
ocSVM_2_test	0.2893	0.9857	0.0612	0.9972	0.9388	0.0612	0.0028
ocSVM_3_test	0.2839	0.9688	0.0550	0.9945	0.9450	0.0550	0.0055
ocSVM_4_test	0.2913	0.9610	0.0657	0.9917	0.9343	0.0657	0.0083
ocSVM_5_test	0.3047	0.9789	0.0825	0.9945	0.9175	0.0825	0.0055
ocSVM_6_test	0.2906	0.9730	0.0639	0.9945	0.9361	0.0639	0.0055
ocSVM_7_test	0.2879	0.9714	0.0603	0.9945	0.9397	0.0603	0.0055
ocSVM_8_test	0.2913	0.9863	0.0639	0.9972	0.9361	0.0639	0.0028
ocSVM_9_test	0.2866	0.9706	0.0586	0.9945	0.9414	0.0586	0.0055

18.5 Isolation Forest – Εκπαίδευση

model	accuracy	precision	recall	tp	fp	tn	fn
forest_0_train	0.9996	1.0	0.9996	(0)	0.0004	0.9996	(0)
forest_1_train	0.9996	1.0	0.9996	(0)	0.0004	0.9996	(0)
forest_2_train	0.9996	1.0	0.9996	(0)	0.0004	0.9996	(0)
forest_3_train	0.9996	1.0	0.9996	(0)	0.0004	0.9996	(0)
forest_4_train	1.0000	1.0	1.0000	(0)	0.0000	1.0000	(0)
forest_5_train	1.0000	1.0	1.0000	(0)	0.0000	1.0000	(0)
forest_6_train	1.0000	1.0	1.0000	(0)	0.0000	1.0000	(0)
forest_7_train	1.0000	1.0	1.0000	(0)	0.0000	1.0000	(0)
forest_8_train	0.9996	1.0	0.9996	(0)	0.0004	0.9996	(0)
forest_9_train	1.0000	1.0	1.0000	(0)	0.0000	1.0000	(0)

18.6 Isolation Forest – Τεστ

model	accuracy	precision	recall	tp	fp	tn	fn
forest_0_test	0.7758	0.7721	0.9982	0.0854	0.0018	0.9982	0.9146
forest_1_test	0.7611	0.7599	1.0000	0.0193	0.0000	1.0000	0.9807
forest_2_test	0.7597	0.7589	1.0000	0.0138	0.0000	1.0000	0.9862
forest_3_test	0.7611	0.7603	0.9991	0.022	0.0009	0.9991	0.978
forest_4_test	0.7597	0.7589	1.0000	0.0138	0.0000	1.0000	0.9862
forest_5_test	0.7611	0.7599	1.0000	0.0193	0.0000	1.0000	0.9807
forest_6_test	0.7617	0.7605	1.0000	0.022	0.0000	1.0000	0.978
forest_7_test	0.7617	0.7605	1.0000	0.022	0.0000	1.0000	0.978
forest_8_test	0.7604	0.7594	1.0000	0.0165	0.0000	1.0000	0.9835
forest_9_test	0.7597	0.7589	1.0000	0.0138	0.0000	1.0000	0.9862

18.7 Συγκεντρωτικά Στατιστικά

mahalanobis train accuracy 0.97559 precision 1.00000 recall 0.97559 tp fp 0.02441 tn 0.97559 fn	mahalanobis test accuracy 0.83000 precision 0.82864 recall 0.97737 tp 0.37246 fp 0.02263 tn 0.97737 fn 0.62754
ocSVM train accuracy 0.58748 precision 1.00000 recall 0.58748 tp fp 0.41252 tn 0.58748 fn	ocSVM test accuracy 0.29048 precision 0.97542 recall 0.06353 tp 0.99503 fp 0.93647 tn 0.06353 fn 0.00497
forest train accuracy 0.9998 precision 1.0000 recall 0.9998 tp fp 0.0002 tn 0.9998 fn	forest test accuracy 0.76220 precision 0.76093 recall 0.99973 tp 0.02479 fp 0.00027 tn 0.99973 fn 0.97521

19 Παράρτημα Λ – Δείκτες Επίδοσης Μοντέλων ανά Αναδίπλωση (Κανονικοποιημένα Συναισθήματα)

19.1 Mahalanobis – Εκπαίδευση

model	accuracy	precision	recall	tp	fp	tn	fn
mahalanobis_0_train_scaled	0.9932	1.0	0.9932	(0)	0.0068	0.9932	(0)
mahalanobis_1_train_scaled	0.9901	1.0	0.9901	(0)	0.0099	0.9901	(0)
mahalanobis_2_train_scaled	0.9920	1.0	0.9920	(0)	0.0080	0.9920	(0)
mahalanobis_3_train_scaled	0.9932	1.0	0.9932	(0)	0.0068	0.9932	(0)
mahalanobis_4_train_scaled	0.9935	1.0	0.9935	(0)	0.0065	0.9935	(0)
mahalanobis_5_train_scaled	0.9913	1.0	0.9913	(0)	0.0087	0.9913	(0)
mahalanobis_6_train_scaled	0.9913	1.0	0.9913	(0)	0.0087	0.9913	(0)
mahalanobis_7_train_scaled	0.9745	1.0	0.9745	(0)	0.0255	0.9745	(0)
mahalanobis_8_train_scaled	0.9920	1.0	0.9920	(0)	0.0080	0.9920	(0)
mahalanobis_9_train_scaled	0.9935	1.0	0.9935	(0)	0.0065	0.9935	(0)

19.2 Mahalanobis – Τεστ

model	accuracy	precision	recall	tp	fp	tn	fn
mahalanobis_0_test_scaled	0.7463	0.7539	0.9867	0	0.0133	0.9867	1
mahalanobis_1_test_scaled	0.7470	0.7541	0.9876	0	0.0124	0.9876	1
mahalanobis_2_test_scaled	0.7497	0.7547	0.9911	0	0.0089	0.9911	1
mahalanobis_3_test_scaled	0.7510	0.7551	0.9929	0	0.0071	0.9929	1
mahalanobis_4_test_scaled	0.7477	0.7542	0.9885	0	0.0115	0.9885	1
mahalanobis_5_test_scaled	0.7523	0.7554	0.9947	0	0.0053	0.9947	1
mahalanobis_6_test_scaled	0.7483	0.7544	0.9894	0	0.0106	0.9894	1
mahalanobis_7_test_scaled	0.7329	0.7516	0.9663	0.008	0.0337	0.9663	0.9917
mahalanobis_8_test_scaled	0.7503	0.7549	0.9920	0	0.0080	0.9920	1
mahalanobis_9_test_scaled	0.7477	0.7542	0.9885	0	0.0115	0.9885	1

19.3 οcSVM – Εκπαίδευση

model	accuracy	precision	recall	tp	fp	tn	fn
ocSVM_0_train_scaled	0.5639	1.0	0.5639	(0)	0.4361	0.5639	(0)
ocSVM_1_train_scaled	0.5764	1.0	0.5764	(0)	0.4236	0.5764	(0)
ocSVM_2_train_scaled	0.5878	1.0	0.5878	(0)	0.4122	0.5878	(0)
ocSVM_3_train_scaled	0.5985	1.0	0.5985	(0)	0.4015	0.5985	(0)
ocSVM_4_train_scaled	0.5791	1.0	0.5791	(0)	0.4209	0.5791	(0)
ocSVM_5_train_scaled	0.5711	1.0	0.5711	(0)	0.4289	0.5711	(0)
ocSVM_6_train_scaled	0.6137	1.0	0.6137	(0)	0.3863	0.6137	(0)
ocSVM_7_train_scaled	0.5719	1.0	0.5719	(0)	0.4281	0.5719	(0)
ocSVM_8_train_scaled	0.5806	1.0	0.5806	(0)	0.4194	0.5806	(0)
ocSVM_9_train_scaled	0.5837	1.0	0.5837	(0)	0.4163	0.5837	(0)

19.4 ocSVM – Τεστ

model	accuracy	precision	recall	tp	fp	tn	fn
ocSVM_0_test_scaled	0.2470	0.7273	0.0071	0.9917	0.9929	0.0071	0.0083
ocSVM_1_test_scaled	0.2463	0.6429	0.0080	0.9862	0.9920	0.0080	0.0138
ocSVM_2_test_scaled	0.2463	0.7000	0.0062	0.9917	0.9938	0.0062	0.0083
ocSVM_3_test_scaled	0.2456	0.6364	0.0062	0.989	0.9938	0.0062	0.011
ocSVM_4_test_scaled	0.2477	0.7500	0.0080	0.9917	0.9920	0.0080	0.0083
ocSVM_5_test_scaled	0.2423	0.3750	0.0027	0.9862	0.9973	0.0027	0.0138
ocSVM_6_test_scaled	0.2443	0.5385	0.0062	0.9835	0.9938	0.0062	0.0165
ocSVM_7_test_scaled	0.2430	0.4000	0.0018	0.9917	0.9982	0.0018	0.0083
ocSVM_8_test_scaled	0.2477	0.7143	0.0089	0.989	0.9911	0.0089	0.011
ocSVM_9_test_scaled	0.2450	0.5833	0.0062	0.9862	0.9938	0.0062	0.0138

19.5 Isolation Forest – Εκπαίδευση

model	accuracy	precision	recall	tp	fp	tn	fn
forest_0_train_scaled	0.9996	1.0	0.9996	(0)	0.0004	0.9996	(0)
forest_1_train_scaled	0.9996	1.0	0.9996	(0)	0.0004	0.9996	(0)
forest_2_train_scaled	0.9996	1.0	0.9996	(0)	0.0004	0.9996	(0)
forest_3_train_scaled	0.9996	1.0	0.9996	(0)	0.0004	0.9996	(0)
forest_4_train_scaled	0.9996	1.0	0.9996	(0)	0.0004	0.9996	(0)
forest_5_train_scaled	0.9996	1.0	0.9996	(0)	0.0004	0.9996	(0)
forest_6_train_scaled	0.9996	1.0	0.9996	(0)	0.0004	0.9996	(0)
forest_7_train_scaled	0.9996	1.0	0.9996	(0)	0.0004	0.9996	(0)
forest_8_train_scaled	0.9996	1.0	0.9996	(0)	0.0004	0.9996	(0)
forest_9_train_scaled	0.9996	1.0	0.9996	(0)	0.0004	0.9996	(0)

19.6 Isolation Forest – Τεστ

model	accuracy	precision	recall	tp	fp	tn	fn
forest_0_test_scaled	0.7564	0.7564	1.0000	0	0.0000	1.0000	1
forest_1_test_scaled	0.7564	0.7564	1.0000	0	0.0000	1.0000	1
forest_2_test_scaled	0.7564	0.7564	1.0000	0	0.0000	1.0000	1
forest_3_test_scaled	0.7564	0.7564	1.0000	0	0.0000	1.0000	1
forest_4_test_scaled	0.7544	0.7559	0.9973	0	0.0027	0.9973	1
forest_5_test_scaled	0.7564	0.7564	1.0000	0	0.0000	1.0000	1
forest_6_test_scaled	0.7557	0.7562	0.9991	0	0.0009	0.9991	1
forest_7_test_scaled	0.7564	0.7564	1.0000	0	0.0000	1.0000	1
forest_8_test_scaled	0.7564	0.7564	1.0000	0	0.0000	1.0000	1
forest_9_test_scaled	0.7564	0.7564	1.0000	0	0.0000	1.0000	1

19.7 Συγκεντρωτικά Στατιστικά

mahalanobis train accuracy 0.99046 precision 1.00000 recall 0.99046 tp fp 0.00954 tn 0.99046 fn	mahalanobis test accuracy 0.74732 precision 0.75425 recall 0.98777 tp 0.00083 fp 0.01223 tn 0.98777 fn 0.99917
ocSVM train accuracy 0.58267 precision 1.00000 recall 0.58267 tp fp 0.41733 tn 0.58267 fn	ocSVM test accuracy 0.24552 precision 0.60677 recall 0.00613 tp 0.98869 fp 0.99387 tn 0.00613 fn 0.01131
forest train accuracy 0.9996 precision 1.0000 recall 0.9996 tp fp 0.0004 tn 0.9996 fn	forest test accuracy 0.75613 precision 0.75633 recall 0.99964 tp 0.00000 fp 0.00036 tn 0.99964 fn 1.00000